Theory of multibin tests: Definition and existence of extraneous tests

Elihu Lubkin

Physics Department, The University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201
(Received 9 January 1974)

Dirac's sets of commuting observables, in the guise of lists of mutually orthogonal projections which add to the unit matrix, are readily extended to a convex completion. But, furthermore, there exist lists of nonnegative Hermitian matrices which sum to the unit matrix which do not even belong to this convex completion. It is shown that these "extraneous" lists appear as tests in ordinary quantum-mechanical experiments. This circumstance leads to simpler rules for injecting measurement theory into the social sciences than might otherwise be proposed. Various relationships between lists of orthogonal projections and more general tests are given. The problem of devising rules of inference by direct computation is very briefly engaged.

1. INTRODUCTION

The epistemology of the physics of the past half century is crudely conveyed by the formula Tr PA for the expectation value of an "observable" A in the state P. The notion of "observable" was a convenience for incorporating classical machinery into the new physics, yet it is desirable to clarify the epistemology as follows: Associated to each method of preparation designated as a state-producing method there is a *state* matrix P, and associated to each further continuation of procedures which could lead to b mutually exclusive and exhaustive results there is a list $A=(A_1,\ldots,A_b)$ of matrices, such that Tr PA_k is the probability that state P processed by *b-test* A lead to outcome k.

Technical difficulties will be minimized throughout by using $n\times n$ matrices of complex numbers, n finite. Such matrices will be said to be of *size n*. The test is of *size n* and *bin number b*. A few more terms and notions are most concisely introduced in brief review of standard material:

The primordial state of quantum mechanics is the *pure state*, a one-dimensional Hermitian projection. Landau,[1] von Neumann,[2] and Weyl[3] observed that the feasible procedure of what I shall call *rouletting* states, namely producing any of several states by random choice with preassigned probabilities, extends the pure states to their convex completion, the body of *states*, nonnegative Hermitian matrices of trace 1. Indeed, unitary transformation of any state matrix to diagonal form displays it as a convex combination of (orthogonal) one-dimensional projections. The extreme states are the pure states.

#1. *Definition*: The primordial tests of quantum mechanics are the lists $(E_1,\ldots,E_b)$ of mutually orthogonal Hermitian projections which sum to the unit matrix I. I call these *sharp* tests.

A lingering commitment of scientific thought to an objective universe of states has led to little interest in generalizing[4] the sharp tests. Yet a laboratory assistant who inserts different b-channel experiments between the state and a common array of b output bins or counters according to a predetermined schedule of probabilities is of course rouletting tests. The b-tests $A^k=(A_1^k,\ldots,A_b^k)$ and rouletting probabilities $p_k\geqq 0$, $\Sigma p_k=1$, lead to the rouletted b-test $A=(A_1,\ldots,A_b)$ with $A_i=\Sigma p_k A_i^k$. Indeed, the matrices A_i so defined do generate the rouletted probabilities of outcomes stemming from any state if they are put together with the state matrix P into the Tr PA_i formula, but no other matrices A_i' will do provided enough states P are available, since Tr $P(A_i-A_i')=0$ for all one-dimensional projections P already implies that $A_i-A_i'=0$. Even if the original tests A^k are sharp, the rouletted test A is sharp only if the A^k coincide (#25).

Therefore b-tests not themselves sharp are readily produced from sharp b-tests by such an annoying assistant, or else by imperfect sorting into the bins, a more likely defect. But are there yet other kinds of tests?

#2. *Definition of b-plex*: #8 below shows that the only sensible b-tests have each A_i nonnegative Hermitian, and $\Sigma_{i=1}^b A_i=I$. A b-tuple of matrices so restricted will be called a *b-plex*.

These restrictions are similar to those for a state matrix, there being a clause for nonnegativity and a clause for normalization.

#3. *Definition*: *Convex combination* of b-plexes A^k with coefficients p_k will be related to the usual convex combination of matrices by requiring commutativity of convex combination with selection of the ith coordinate matrix, i.e., $(\Sigma p_k A^k)_i=\Sigma p_k A_i^k$.

Convex combination of b-plexes corresponds to the physical notion of rouletting the corresponding tests.

The b-plexes of $n\times n$ matrices obviously form a closed convex body for each b, n.

Were the situation as simple for tests as for states, the following would hold: Convex completion of the sharp b-plexes would produce all the b-plexes, all sharp b-plexes would be extreme, and all extreme b-plexes would be sharp. Indeed, all sharp b-plexes are extreme (#25), but the other statements are false for $b\geqslant 3$.

#4. *Definitions*: The convex completion of the sharp b-plexes will be called the *undersharp* body (undersharp b-body in size n). Those b-plexes which are not undersharp are *extraneous*. If the matrices A_i of b-plex $(A_1,\ldots,A_b)$ commute, the b-plex is *commutative*.

We will see that commutative b-plexes are undersharp (#44). Since a 2-plex is necessarily commutative there is no extraneous 2-plex. Extraneous 3-plexes and 4-plexes of 2×2 matrices will be constructed, however (Sec. 5). Even though corresponding physical tests cannot be achieved by any rouletting of sharp tests, it is shown in Sec. 7 that physical tests nevertheless exist corresponding to arbitrary b-plexes—provided that any unitary transformation is regarded as physically achievable by a motion. In a subsequent paper a magnetic spin system is described wherein the unitary transformations and other constructions of Sec. 7 are indeed physically achievable. It would therefore be unadvisable to restrict the following summary of quantum matrix epistemology, for example, by imposing undersharpness on the tests.

#5. *The matrix format* (*MF*): If state procedure i together with test procedure j yield probabilities p_{ijk} for the several outcomes $k=1,\ldots,b_j$ of the test, then nonnegative Hermitian matrices P_i and A_{jk} are to be sought fulfilling

$$\text{Tr } P_i=1,\quad \sum_k A_{jk}=I,$$
$$\text{Tr } P_iA_{jk}=p_{ijk}.$$

A fit of MF could inject the methods of modern physics into empirical domains where quantum interference effects might be large but unsuspected. Especially if n is small, the clarity achieved by having few parameters could make further prior theoretical assumptions unnecessary. Schrödinger equation-like drift of "isolated states," for example, could be detected by comparing state preparations which differ only in respect to a variable "coasting" time, without any prior knowledge of a Hamiltonian, of existence of a Hamiltonian, or surety as to what constitutes "isolation."

I have presented elsewhere[5] the guess that this sort of analysis might prove interesting for data taken in "p_{ijk}" form in the study of visual perception. Several groups of workers in quantum logic have similar hopes of finding quantum interference effects in psychology or elsewhere in the social sciences.

So as to avoid misunderstanding, note that I am using ordinary density matrices for initial states, but avoid them so meticulously for "final states" as to even avoid using the term "final state." Nevertheless the test matrix A_{jk} associated with bin k of test j is nonnegative Hermitian, hence differs from a "final state density matrix" only in its normalization. (Tr A_{jk} may be anything from 0 to the format size n.) In an empirical domain where n is known and n^2-1 real-linearly independent state matrices are physically available, each new test matrix A_{jk} may be fixed in terms of its n^2-1 probabilities against these states; it is not mandatory to talk about the whole test $A=(A_{j1},\ldots,A_{jb})$. Then "unnormalized final state density matrices" will do.

The plan of the sequel is as follows. First, #8 records the elementary argument for not looking at anything more general than b-plexes (within the main context of this paper, namely, the Tr PA formula for probabilities from complex matrices). Then the study of extreme points among the b-plexes is related to linear independence of spaces $X(A_i)$ of 2-sided nonnegative variation or of spaces $Y(A_i)$ of 2-sided "spindle" variation. The nonnegative variation version gives a lemma, #23, that $(A_1,\ldots,A_b)$ extreme requires all A_i to be singular (except when $b-1$ of them vanish), which later helps determine all extreme b-plexes of 2×2 matrices. The spindle-variation version establishes the extremity of the sharp b-plexes almost immediately.

The possibility that summing some of the A_i together produce a sharp b'-plex introduces a notion of segmentation, unsegmentability, and a reduction of any b-plex into its unsegmentable parts; the opposite process of assembling segments in particular can be used to generate more extraneous examples.

A different sort of extraction of a sharp structure from a b-plex yields a "reduction" into a sharp b-plex and a b-plex of matrices none of which has 1 for an eigenvalue, in the manner of a direct sum.

The existence of extraneous b-plexes is established in Sec. 5 by finding all extraneous extreme b-plexes of 2×2 matrices without zero matrices; for these $b=3$ or 4.

It is proved in Sec. 6 that commutative b-plexes are undersharp, whence 2-plexes, also called *questions*, are necessarily undersharp. This result together with the common emphasis upon sharp and mixed[4] questions is held to account for a lack of attention to the phenomenon of extraneity.

The issue of whether extraneous b-plexes can occur as physical tests is at this point very interesting, because such b-plexes indeed exist, but cannot of course be reached as physical tests by rouletting the usual sharp tests. Section 7 is a construction not involving convexity arguments which settles the issue by showing how *any* b-plex may occur as a physical test, provided that any unitary transformation may be eigeneered as a motion. The construction is similar to Landau tracing[1] for state matrices, which as is well known leads from pure states to mixtures and so simulates rouletting, yet for tests is more powerful than rouletting in that extraneous tests appear, not only mixtures of sharp tests.

Two kinds of sufficiency for undersharpness are given in Sec. 9, the first depending upon dominance of one bin over all others, the second upon special hypotheses involving close relationship of a test to questions by attaining the b bins through $b-1$ successive bifurcations.

An ideally simple fit (Sec. 10) of MF to "data" involving lots of orthogonality opens the final and most important topic, the praxis of fitting MF. Improvement of computability by elimination of invariance (Sec. 11) leads to a conjectural alternative "Sparse Format" (SF).

#6. *Notation*: V or V_n will denote the Hilbert space of complex n-tuples on which the $n\times n$ matrices may act.

$\lambda(M)$ will denote the space of eigenvectors of matrix M belonging to λ. Thus, $\lambda(M)=\{0\}$ unless λ is an eigenvalue of M, and $0(M)$ is the dernel of M. If M is normal,

$\Sigma_{\lambda\neq 0}\lambda(M)$ is the orthogonal complement $(0(M))_\perp$ of $0(M)$. Orthogonal projection on this *space of M* will be denoted E_M.

I or I_n denotes the $n\times n$ unit matrix. Thus, $E_I = V$. $\mathrm{diag}(a_1, a_2, \ldots)$ is obvious notation for a diagonal matrix. The elements of matrix A_k are a_{kij}.

$A\cdot B \equiv \mathrm{Tr}\, A^\dagger B$ defines a positive definite Hermitian inner product on the $n\times n$ complex matrices, real for A, B Hermitian. If $A\cdot B = 0$, A and B will be said to be *trace-orthogonal*. The accompanying norm $\|A\| \equiv (A\cdot A)^{1/2}$ is the positive square root of the sum of the absolute squares of A's matrix elements. Thus $\|I\| = n^{1/2}$.

A matrix with nonnegative real eigenvalues is *nonnegative*. If furthermore it is nonsingular it is *positive*. $A \geq B$ means matrix $A - B$ is nonnegative.

#7. *Topology*: Topological statements relating to various sets of matrices will always refer to the norm topology on the relevant set. Topological statements about sets of b-tuples of matrices refer to the topology induced on subsets by the Cartesian product of the aforesaid norm topology for each of the b coordinates, equivalently, to a norm topology wherein the norm of a b-tuple of matrices is the positive square root of the sum of the absolute squares of all the matrix elements.

Inasmuch as all matrices of interest will be nonnegative $n\times n$ Hermitian matrices $\leq I$, the matrices all lie within the closed ball of radius $n^{1/2}$, which is compact. A b-tuple of such matrices has b-tuple norm $\leq (bn)^{1/2}$; all b-tuples of interest thus also lie in a compact ball. This comment should make the compactness of various sets (b-plexes, states, the spindle, the undersharp body, commutative b-plexes) obvious upon definition.

2. PHYSICAL BASIS FOR THE B-PLEX CONDITIONS

Part 1 of the following theorem is standard material, worked out for the sake of comparison with part 2.

#8. *The b-plex theorem*: If probabilities p_k, $p_k \geq 0$, $\Sigma_k p_k = 1$, are to be computed from the rule $p_k = \mathrm{Tr}\, PA_k$ and state matrices P and b-test bin matrices $A_1, \ldots, A_b$, then

1. If any sharp n-plex represents some test, then $P = P^\dagger$, $P \geq 0$, and $\mathrm{Tr}\, P = 1$.

2. If any 1-dimensional projection represents some state, then $(A_1, \ldots, A_b)$ is a b-plex.

#9. *Proof of* 1: Matrix A_1 may be any 1-dimensional projection $|x\rangle\langle x|$, hence $\mathrm{Tr}\, P|x\rangle\langle x| = \langle x|P|x\rangle$ is a probability, therefore nonnegative. Hence $P \geq 0$. Also $\langle x|P - P^\dagger|x\rangle = \langle x|P|x\rangle - \langle x|P|x\rangle^* = 0$, so $P = P^\dagger$. $\Sigma p_k = 1 = \Sigma \mathrm{Tr}\, PA_k = \mathrm{Tr}\, P\Sigma A_k = \mathrm{Tr}\, P$ if any b-plex A occurs as a test, and *sharp* n-plexes are given to occur.

The hypothesis of part 1 is stated with unnecessary strength in order to refer to an analog of Dirac's complete sets of commuting observables.

Proof of 2: $p_k = \mathrm{Tr}\, |x\rangle\langle x|A_k = \langle x|A_k|x\rangle \geq 0$, all x, shows both $A_k \geq 0$ and $A_k = A_k^\dagger$. $\Sigma p_k = 1$ yields $\langle x|\Sigma A_k - I|x\rangle = 0$ for all x, whence $\Sigma A_k = I$. QED

3. BASIC RESULTS ON THE EXTREME POINTS OF THE BODY OF B-PLEXES

#10. *Definition of spindle*: The intersection $\{A|A = A^\dagger, A \geq 0\} \cap \{B|B = B^\dagger, B = I - A, A \geq 0\}$ of the nonnegative cone with the I-shifted nonpositive cone will be called the *spindle*, or spindle (n): n being the size, as always.

#11. *Remarks*: A 2-plex A is defined by its first coordinate A_1, which is an element of the spindle. Conversely any element A_1 of the spindle paired with $I - A_1$ forms a 2-plex. This correspondence is obviously a convexity isomorphism.

The spindle also appears as the set of all possible matrices which may occur in a fixed place, e.g., place 1, of a b-plex.

#12. *Definition*: For A Hermitian nonnegative,

$\{X|X = X^\dagger$ and $\exists\, \lambda > 0$ such that

$-\lambda \leq u \leq \lambda$ implies $A + uX \geq 0\}$

will be called the *space of 2-sided nonnegative variation of* A, and denoted $X(A)$. The characterization $X(A) = \{E_A M E_A | M = M^\dagger\}$ of #17 shows $X(A)$ to be a real linear space.

#13. *Definition*: For A Hermitian nonnegative,

$\{Y|Y = Y^\dagger$ and $\exists\, \lambda > 0$ such that

$-\lambda \leq u \leq \lambda$ implies $A + uY \in$ spindle$\}$

will be called the *space of 2-sided spindle variation of* A.

#14. *Remark*: $Y(A) = X(A) \cap X(I - A)$ shows $Y(A)$ to be a real linear space and is otherwise useful.

#15. *Lemma for off diagonal zeros*: If a nonnegative Hermitian matrix M has a 0 in a diagonal place, it has 0's in the corresponding row and column.

#16. *Proof*: The result is a direct consequence of nonnegativity of the 2×2 minors $m_{ii}m_{jj} - m_{ij}m_{ji} = 0 - |m_{ij}|^2$ with $m_{ii} = 0$. QED

#17. *Theorem*: If A is a nonnegative Hermitian matrix, $X(A)$ is the set of Hermitian matrices of form $E_A M E_A$, i.e., which act in $E_A V$. If A is diagonalized, say $A = \mathrm{diag}(a_1, \ldots, a_k, 0, \ldots, 0)$ with $a_i > 0$ for $i \leq k$, $a_i = 0$ for $i > k$, then $X \in X(A)$ if and only if X is Hermitian with 0 elements outside the upper-left $k\times k$ bloc.

#18. *Proof*: Since $A + \lambda X$, $A - \lambda X$ must both be nonnegative for $\lambda > 0$ small, it is clear why X must vanish in the lower right-hand bloc. The "lemma for off-diagonal zeros" applied to $A + \lambda X$ shows $A + \lambda X$, hence X itself, to have zeros everywhere else outside the upper left-hand bloc. Continuity in λ of the upper left minor determinants of $A + \lambda X$ and positivity of A considered as a matrix in the upper left bloc shows that X may not be further delimited. The form $E_A M E_A$ generalizes this to A not necessarily diagonal, inasmuch as it specializes correctly when A is diagonal. QED

#19. *Corollary*: $\dim X(A) = (\mathrm{rank}\, A)^2$, for a non-

negative Hermitian matrix A.

#20. *Corollary*: The following are equivalent for nonnegative Hermitian A:

A is of rank 1;

A is a positive multiple of a 1-dimensional projection;

$a_{ij}=x_i^*x_j$ or $A=|x\rangle\langle x|$ with $x\neq 0$ not necessarily normalized;

A is proportional to a pure state;

$\dim X(A)=1$;

$X(A)=\{rA\,|\,r \text{ real}\}$.

#21. *Theorem*: The following are equivalent conditions for b-plex $A=(A_1,\ldots,A_b)$:

A is an extreme b-plex;

the spaces $X(A_1),\ldots,X(A_b)$ are linearly independent;

the spaces $Y(A_1),\ldots,Y(A_b)$ are linearly independent.

#22. *Proof*: Failure of extremity depends on the ability to express A as a midpoint of varied b-plexes, $A\pm X$, $X\neq 0$. That $A\pm X$ be b-plexes requires only that the $A_k\pm X_k$ be nonnegative (hence also $A_k+uX_k\geq 0$ for $-1\leq u\leq 1$), *and* that $\Sigma_k A_k\pm X_k=I$. Since $\Sigma_k A_k=I$, this reduces to $\Sigma_k X_k=0$, a relation of linear dependence among the b elements $X_1,\ldots,X_b$ of the spaces of 2-sided nonnegative variation.

The alternative statement in terms of the $Y_k(A)$ follows by observing that whenever an expression as a midpoint is achieved, each X_k involved belongs to $Y(A_k)$, not only to $X(A_k)$. This is so because each matrix of a b-plex belongs to the spindle. Hence $A_k+uX_k\geq 0$ above may be replaced by $A_k+uX_k\in$ spindle. QED

#23. *Singularity lemma*: If b-plex A is extreme, either each A_k is singular or else b-1 of them are 0, the remaining one being I.

#24. *Proof*: Suppose A is extreme and A_1 for example is nonsingular. Then #17 shows $X(A_1)$ to be all of V. Linear independence of the $X(A_k)$ (#21) then requires $X(A_k)=\{0\}$ for $k\neq 1$, whence $A_k=0$ for $k\neq 1$, again by #17. QED

#25. *Theorem*: Sharp b-plexes are extreme.

#26. *Proof*: If E is a Hermitian projection, $X(E)=\{X\,|\,X=EME,\ M=M^\dagger\}$ and $X(I-E)=\{X\,|\,X=(I-E)M(I-E),\ M=M^\dagger\}$. Hence $Y(E)=X(E)\cap X(I-E)=\{0\}$. Since the coordinates A_k of a sharp b-plex are projections, the $Y(A_k)$ are all $\{0\}$ and hence are linearly independent. QED

4. NOTIONS OF REDUCIBILITY

Two notions of a separation of sharp structure out of a b-plex will now be given: segmentation and 1-elimination. Applied to an extreme b-plex they will lead to extraneous extreme b'-plexes of smaller size and bin number. This general knowledge is interesting in itself, and will slightly streamline the subsequent discussion of extreme b-plexes of 2×2 matrices.

A. Segmentation

#27. *Definition of fusion*: A b'-plex A' obtained from a b-plex A by adding together some of the matrices, so that $b'<b$, will be said to have been obtained by *fusion*. The corresponding bins will be said to be *fused*. Fusion corresponds empirically to gathering the counts in the b bins into b' subtotals. Also, A is said to be *finer* than A'.

#28. *Definition of segmentation*: If b-plex A fuses to a sharp b'-plex E, E is a *segmentation* of A. If a segmentation of bin number ≥ 2 exists, A is *segmentable*, if not, A is *unsegmentable*.

#29. *Definition of* (b,E)*-plex*: A list $A=(A_1,\ldots,A_b)$ of b nonnegative Hermitian matrices which sums to matrix E will be called a (b,E)-plex. This notation seems useful only for E a projection; therefore the notation will imply that E is a projection. Thus a b-plex is a (b,I)-plex. The rank of E will be said to be the *rank* of a (b,E)-plex.

#30. *Segmention theorem*: Every b-plex A has a unique finest segmentation $E=(E_1,\ldots,E_s)$, finer than each other segmentation of A. The b_k matrices A_i of A which fuse to projection E_k form a *segment* of A which (listed in any order) is an unsegmentable (b_k,E_k)-plex. Each original matrix A_i is of form $E_kA_iE_k$. E_k is the *segment projection* of A_i. If the $(E_1,\ldots,E_s)$ are simultaneously diagonalized with the 1 eigenvalues of each E_k conveniently together, then each $A_i\in\{A_1,\ldots,A_b\}$ has nonzero elements only in the diagonal square bloc corresponding to the 1's of its segment projection. If A is a convex combination $A=\Sigma p_r A^r$, then each A^r has E for a segmentation, and the combination is attained by performing the corresponding combination within each segment. A is extreme if and only if each segment is extreme among the appropriate (b_k,E_k)-plexes.

A sharp b-plex segments into its single projections.

The study of extreme b-plexes may be considered to be reduced to the study of unsegmentable extreme b-plexes, by the segmentation theorem.

#31. *Proof*: More generally, the reduction to bloc form and the reduction of convex combination applies to any segmentation, in particular to a segmentation into two projections, E, F, $E+F=I$: by renumbering matrices, $A_1+\ldots+A_{b_1}=E$, $A_{b_1+1}+\ldots+A_b=F$. $A_1\leq E$ with E in diagonal form shows A_1 to have diagonal zeros where E does, and therefore to be in bloc form $A_1=EA_1E$ by the "lemma on offdiagonal zeros."

If $A=\Sigma p_r A^r$, then $A_1=\Sigma p_r A_1^r$; hence $A_1^r\leq A_1\leq E$ and $A_1^r=EA_1^rE$ as above: not only A_1 but each A_1^r is zero outside the E-bloc.

It remains only to show that there is a finest segmentation. The essence of this is obviously that if (E,F), (G,H) are segmentations, then all four projections E, F, G, H commute and define a finer segmenta-

tion (EG, EH, FG, FH). Indeed, each matrix A_k of the original b-plex $(A_1, \ldots, A_b)$ is either of form EA_kE or FA_kF, and is also either of form GA_kG or HA_kH. There are altoghether four types of A_k: if $A_k = EA_kE = GA_kG$ it is of type (E, G), etc. Sum all A_k of each type, to get four subtotal matrices A_{EG}, A_{EH}, A_{FG}, A_{FH}. Since the sum of all the A_k with $A_k = EA_kE$ is E itself, $A_{EG} + A_{EH} = E$. Similarly, $A_{FG} + A_{FH} = F$, $A_{EG} + A_{FG} = G$, $A_{EH} + A_{FH} = H$.

Since A_{EG} is a sum of terms of form EA_kE, $EA_{EG}E = A_{EG}$. Similarly, $GA_{EG}G = A_{EG}$. Since $EF = FE = 0$, $FA_{EG} = FEA_{EG}E = 0$ and also $A_{EG}F = 0$: Each A_{PQ} is unaffected by P, Q multiplication but is annihilated by the other two projections. Hence E multiplication left and right of $A_{EG} + A_{FG} = G$ yields $A_{EG} = EG$ and $A_{EG} = GE$. Thus, E and G commute. Similarly for other pairs among E, F, G, H. It follows that (EG, EH, FG, FH) is a 4-plex of mutually orthogonal projections. In the guise of A_{EG}, A_{EH}, A_{FG}, A_{FH} we already know that A fuses to the 4-plex in question. (More precisely the fusion may be to a 3-plex or a 2-plex if one or two of EG, EH, FG, FH vanishes.)

B. 1-Elimination

#32. *Definition*: If $1(M) = \emptyset$, M is *1-free*. $(A_1, \ldots, A_b)$ is *1-free* if each A_k is 1-free.

#33. *Theorem*: All the matrices of a b-plex of $n \times n$ matrices may be simultaneously unitarily transformed to bloc form, with an upper left $s \times s$ bloc sharp structure which is diagonal (1's and 0's on the diagonal), a lower right $(n-s) \times (n-s)$ square bloc of matrices which form a b-plex of 1-free $(n-s) \times (n-s)$ matrices, and 0's connecting the blocs.

#34. *Proof*: If one of the matrices of a b-plex has eigenvalue 1, say A_1, go to a representation where A_1 is diagonal with 1 at upper left. All the other matrices A_k must have 0 in the upper left place, hence zeros elsewhere in the first row and column by the "lemma on offdiagonal zeros." Consequently, $A_1 = I - \Sigma_{k \neq 1} A_k$ also has zeros off the diagonal in row 1 and in column 1. Iterate this extraction of 1's until the residual $(n-s) \times (n-s)$ matrices have no more 1 eigenvalues. QED

If such a form is a convex combination, it is so by virtue of a corresponding decomposition in its 1-free bloc, inasmuch as the sharp bloc is extreme and hence is merely repeated in each term of the decomposition. Thus:

#35. *Theorem*: An extreme b-plex is unitarily equivalent to a direct sum of a sharp bloc and a 1-free extreme bloc in the sense of #33, and conversely each such direct sum is extreme.

The following is easy to verify. It unites both of the above notions of reduction.

#36. *Theorem*: The segment projections E_k of the finest segmentation of a b-plex are obtained from the segment projections e_k of the finest segmentation of its 1-free part by fusing the associated sharp parts.

Each zero matrix in the b-plex of 1-free parts is in particular a segment, the corresponding segment of the whole b-plex then being the corresponding single sharp-part projection filled out with zeros. The 1-free part may not consist entirely of zero matrices, but there may be no 1-free part, in which case the b-plex is sharp and so segments into its singleton projections anyhow: an empty 1-free part may be regarded as segmenting into singletons for the sake of #36.

#37. *Definition*: An unsegmentable extreme 1-free b-plex is *primitive*.

A primitive b-plex is necessarily extraneous: otherwise it would be undersharp and extreme, hence sharp, hence not 1-free.

The finest segmentation of any extreme b-plex comprises any number of 1-bin zero matrices, and segments free of zero matrices: Zero matrices are always allowed because $X(0) = \{0\}$.

5. EXISTENCE OF EXTREME EXTRANEOUS B-PLEXES AND DESCRIPTION OF ALL EXTREME B-PLEXES OF 2×2 MATRICES

#38. *Theorem*: The extreme b-plexes of 2×2 matrices are as follows: (I), (P, Q) sharp with rank P = rank $Q = 1$, primitive 3-plexes, primitive 4-plexes, and the b-plexes obtained from these by listing zero matrices. Each coordinate of a primitive 3-plex or 4-plex is of form $|\mathbf{a}|I + \mathbf{a} \cdot \sigma$ in terms of a real 3-vector $\mathbf{a}$ and the Pauli matrices. Primitive 3-plexes correspond to $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ which form a triangle of nonzero area and perimeter 1, primitive 4-plexes to $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, $\mathbf{a}_4$ which form a quadrilateral of perimeter 1 which does not lie in a plane.

#39. *Proof*: For $n = 2$, segmentability of an extreme b-plex after the zero matrices are omitted implies a splitting into blocs of size 1, hence sharpness. Thus an extraneous extreme b-plex of 2×2 matrices segments into zeros and only one unsegmentable segment. It is easy to see by 1-elimination that this nontrivial segment must be primitive.

Each 2×2 Hermitian matrix M is conveniently represented in terms of a real scalar part m_0 and a real vector part $\mathbf{m}$ by expansion in the Pauli matrices, $M = m_0 I + \mathbf{m} \cdot \sigma$. By #23, each $A_1, \ldots, A_b$ is singular; hence $m_0 = |\mathbf{m}|$. Also $|\mathbf{m}| > 0$ as zero matrices have been cast out. Thus each A_k is a positive multiple of a 1-dimensional projection, with a multiplier < 1, being 1-free. By #20, $X(A_k) = \{rA_k | r \text{ real}\}$. The condition of linear independence of $X(A_1), \ldots, X(A_b)$ (see #21) becomes linear independence of the 4-vectors $(|\mathbf{a}_k|, \mathbf{a}_k)$. Extraneous b-plexes are impossible for $b < 3$, linear independence is impossible for $b > 4$; hence the only possible cases are for 3 and 4 bins. The b-plex condition is here $\Sigma_k |\mathbf{a}_k| = 1$, $\Sigma \mathbf{a}_k = 0$. A 3-plex, 4-plex of such rank-1 matrices is represented in terms of vector parts by a closed triangle, quadrilateral of unit perimeter. If the triangle encloses nonzero area, it is easy to see that the $|\mathbf{a}_k|$ scalar parts lead to linear independence of the three 4-vectors and hence to extremity for the 3-plex; if not, not. For $b = 4$, the quadrilateral must

be nonplanar, in which case the $|\mathbf{a}_k|$ components produce nonzero 4-volume, hence extremity.

Since there is no room in size 2 for two primitive segments or for 1-eigenvalue structure together with a primitive segment, the exhaustive listing of #38 follows.

QED

#40. *Remarks*: Each of the four types of extreme point listed in #38 comes in several connected components, one for each placement of extra zero matrices. The real dimensionalities of these four types are obviously 0, 2, 5, 8, respectively. The 5-dimensional types built on a primitive 3-plex of course are available only for $b \geq 3$; the 8-dimensional types built on a primitive 4-plex only for $b \geq 4$. Thus, for $b \geq 3$, there are many more extraneous extreme points than sharp ones (i.e., $5 > 2$).

As the triangle and quadrilateral figures for the extraneous extreme 3-, 4-plexes have unit perimeter with no side vanishing, 1-freeness is geometrically evident. Similarly noncommutativity is evident as noncollinearity of the vector parts.

As one side of a 4-plex quadrilateral tends to 0 with the perimeter kept 1 and the resulting triangle nondegenerate, we find the general primitive 3-plex with one zero matrix to be a limit of primitive 4-plexes. Similarly the general sharp 2-plex of orthogonal 1-dimensional projections together with one zero matrix is a limit of primitive 3-plexes, the "bilateral" of perimeter 1.

6. COMMUTATIVITY AND UNDERSHARPNESS

It is possible to comment on why the phenomenon of extraneity has escaped attention heretofore economically as an aside, in the development of other material relating commutativity and undersharpness.

#41. *Theorem*: For extreme b-plexes, commutativity is equivalent to sharpness.

#42. *Proof*: Sharpness obviously implies commutativity.

To show the converse, simultaneously diagonalize the b matrices $A_k = \mathrm{diag}(a_{k1}, \ldots, a_{kn})$ of the commutative b-plex A. If any of these diagonal elements a_{kj} is neither 0 nor 1, then $\Sigma_{k=1}^{b} a_{kj} = 1$ shows that there is another a_{ij}, $i \neq k$, also properly between 0 and 1. Replacing a_{kj}, a_{ij} by $a_{kj} \pm \epsilon$, $a_{ij} \mp \epsilon$ represents A as a midpoint of b-plexes, provided $0 < |\epsilon| \leq a_{kj}$ and $|\epsilon| \leq a_{ij}$.

Therefore, all the elements are 0 or 1. $\Sigma_{k=1}^{b} a_{kj} = 1$ shows that there is only one 1 to a bin, i.e., the projections are mutually orthogonal. QED

#43. *Corollary*: All extreme 2-plexes are sharp. All 2-plexes are undersharp. Sharpness and extremity are equivalent for $b = 2$.

Thus extraneity is not present for $b = 2$. It has been common to replace a b-bin test by its various fusions into questions[4] or 2-bin tests (A, B), or even to consider only the sharp 2-tests (E, F), equivalent to projections E or subspaces $1(E)$. The association of $1(E)$ with "yes" and $0(E)$ with "no" introduces a closer linguistic parallel to conventional logic than a direct study of the b-tests without fusion, yet misses the phenomenon of extraneity. The argument that extraneous b-plexes can actually occur in a physical test is deferred until after another commutativity-undersharpness corollary:

#44. *Corollary*: Commutative b-plexes are undersharp; the closed convex hull of the commutative b-plexes is the undersharp b-body; there is no point extreme in the closed convex hull of the commutative b-plexes which fails to be extreme among all b-plexes.

#54 or #56 makes it obvious that there are noncommutative undersharp b-plexes for $b \geq 3$.

#45. *Remark*: The spindle is compact, whence the set of b-plexes is compact; similarly for the set of commutative b-plexes, which is obviously closed. The real dimension of the b-plex body is $(b-1)n^2$, finite. An extreme point of the closed convex hull of a compact set C in a finite-dimensional space itself belongs to C.[6,7]

#46. *Proof*: An extreme point P of the closed convex hull H of the commutative b-plexes is itself commutative. It has been shown in #42 how to express a nonsharp commutative b-plex as a midpoint between distinct commutative b-plexes A and B. Were P not itself sharp, P would therefore be such a midpoint. This would contradict P's extremity in H, a body which contains the segment AB. Hence P is sharp: The extreme points of H are all sharp. Since all sharp b-plexes are commutative and extreme, they are conversely all extreme points of H. QED

7. THE EXISTENCE OF EXPERIMENTS INVOLVING TESTS CORRESPONDING TO ANY B-PLEX

#47. *Theorem*: For any given b-plex of $n \times n$ matrices, there exists a physical b-bin test whose b-plex is the given b-plex.

#48. *Discussion*: "Physical test" must be formulated in terms of metaphysical conventions of physical—mathematical correspondence in order to develop the assertion into a clear-cut theorem, but the mathematization is half the story. Therefore both the mathematization and the verification are lumped in #51 below. The clear-cut theorem is:

#49. *Technical theorem*: Let U be a unitary matrix acting on $V_n \otimes V_b$, ρ a state on V_n, e a sharp b-test on V_b, I the unit matrix on V_n, and a the b-plex on V_n defined by

$$\forall \rho,\ \forall k,\ \mathrm{Tr}\ U(\rho \otimes e_1)U^{\dagger}(I \otimes e_k) = \mathrm{Tr}\ \rho a_k:$$

this map from $bn \times bn$ unitaries to b-plexes of $n \times n$ matrices is surjective.

#50. *Metaphysical requirements*: A sharp b-test must exist as a possible test procedure upon a system, the *probe*, which is describable in terms of $b \times b$ matrices. The probe must be in the corresponding eigenstate of the sharp b-test immediately after the test

is effected. It must be possible to assemble the probe in one of these eigenstates e_1 together with an arbitrary state of a *system of interest*, describable by an arbitrary $n\times n$ matrix ρ. The joint development of the system-probe complex thereafter must eventuate in a unitary transform in the usual way of the initial state $\rho\otimes e_1$. By imposing various "fields" on the system-probe complex, possibly space and time dependent, it must be possible to engineer an arbitrary unitary transform of the complex. It must be possible then to impulsively query the probe with the sharp b-test. Briefly, a Dirac complete set of commuting observables exists for observation of the probe, and any unitary motion of the system—probe complex is possible.

It should be noted that not only the probe but also the system of interest may be engineered to fit in conveniently with these requirements. The assertion is that all b-plexes are attainable in some physical domain of states and tests; not that for any physical n-dimensional state space one can devise so many b-tests as to saturate all b-plexes for each b.

An example using spin-$\frac{1}{2}$ magnets in which the arbitrary unitary transformations and the other details of the construction are indeed physically achievable, is given in the following paper.

#51. *Proof*: Consider a system of interest, the *system*, which is coupled to an instrument by means of immediate contact with a portion of the instrument called the *probe*. The remainder of the instrument contains the registers or bins for recording a result, and will be called the *register*. The state space of system, probe, and register if needed would be taken to be the tensor product of three individual state spaces. However, the role of the register will be merely to effect a sharp test $(E_1,\ldots,E_b)$ upon the system-probe complex at some "final" time. Therefore the probabilities associated with the b bins $k=1,\ldots,b$ are

$$p_k=\mathrm{Tr}\,P'E_k, \tag{1}$$

where P' is the system—probe state at the final time. The matrices P', E_k thus act on system—probe space, with the register not formalized further: the states and probabilities are relative to the register as observer.[8,9]

The system is originally in state ρ, the probe originally in a special state e_1, a 1-dimensional projection, hence the original state P of the system—probe complex is the Kronecker product $P=\rho\otimes e_1$.

The subsequent measurement consists of a unitary motion U of the system—probe complex, producing state $P'=UPU^\dagger$ at the final time, at which time the register enters to make its observation or bin-reduction.

U does so much for us that this final observation need only interact with the probe; thus, $E_k=I\otimes e_k$ in terms of a probe-space sharp test $(e_1,\ldots,e_b)$, and a unit matrix called I acting in the n-dimensional system space. The notation has anticipated the convenience that the probe's initial state e_1 be prearranged by a similar earlier measurement of the probe by the register which has noted and set outcome 1. It is also convenient to have each e_k a 1-dimensional projection; thus probe space is b-dimensional Hilbert space.

Hence

$$p_k=\mathrm{Tr}\;U(\rho\otimes e_1)U^\dagger(I\otimes e_k). \tag{2}$$

It remains to compute the b-plex system-space matrices a_k such that also

$$p_k=\mathrm{Tr}\;\rho\;a_k, \tag{3}$$

by summing over the probe-space indices. In obvious index notation,

$$p_k=u_{ab,cd}\,\rho_{ce}\,e_{1df}\,u^*_{gh,ef}\,\delta_{ga}\,e_{khb} \tag{4}$$

$$=\rho_{ce}\,a_{kec}. \tag{5}$$

So that (5) be true for arbitrary states ρ, the matrix elements of a_k must be

$$a_{kec}=u_{ab,cd}\;e_{1df}\,u^*_{gh,ef}\,\delta_{ga}\,e_{khb}. \tag{6}$$

The choice of basis $e_{khb}=\delta_{kh}\delta_{kb}$ leads to

$$a_{kec}=\sum_g\;u_{gk,c1}\,u^*_{gk,e1}; \tag{7}$$

the summation convention being dropped. Index g labels a state-space basis, hence assumes n values. Hence matrix a_k is a sum of n matrices, the gth one having general e,c matrix element of form ${}_gx_e\,{}_gx_c$ in terms of n vectors ${}_gx$. If the $u_{gk,e1}$ are provisionally regarded as free complex numbers, not entries from a unitary matrix, the various vectors x are arbitrary. But $x^*_e x_c=\langle b_e|x^*\rangle\langle x^*|b_c\rangle$ is the (e,c) element of a general nonnegative multiple of a 1-dimensional projection. A general nonnegative Hermitian matrix may always be represented as a sum of n such, inasmuch as it is even a sum of n nonnegative multiples of mutually orthogonal diagonalizing 1-dimensional projections. Hence aside from the provisionally dropped constraint of unitarity, expression (7) would show matrices a_k to be free nonnegative matrices. (9) below shows that unitarity of U requires $a=(a_1,\ldots,a_b)$ to be a b-plex; the problem is to see that *any* b-plex can be achieved under the constraint of unitarity.

We only wish to show $(a_1,\ldots,a_b)$ to be a free b-plex, not an entirely free b-tuple of nonnegative matrices. Thus,

$$\sum_k\;a_{kec}=\delta_{ec} \tag{8}$$

is imposed. (8) is translated by (7) into

$$\sum_{gk}\;u_{gk,c1}\,u^*_{gk,e1}=\delta_{ec}=\delta_{ec}\,\delta_{11} \tag{9}$$

on the u's. Equation (9) can be expressed as orthonormality of the columns labeled $(1,1),\ldots,(n,1)$ of matrix U. Any n columns of complex numbers which meet (9) can have $nb-n$ further columns of complex numbers adjoined such that all nb columns are orthonormal, by means of the well-known Schmidt process; of course the U so achieved is indeed unitary. QED

The other columns of U produce b-1 other b-plexes corresponding to replacing the initial register setting e_1 by $e_2,\ldots,e_b$. Thus (9) generalizes to a notion of b associated b-plexes, which however is not examined further.

8. TESTS AS OBSERVABLES

If an experimental test is formally a convex combination of other tests, its bin outputs are indistinguishable from those of a test concocted by rouletting those other tests, whether it is undersharp or extraneous. Such a test may therefore be felt to be dirty, in providing a confusion of information from the other tests which had better be kept separate. A laboratory assistant who mixes tests with a roulette wheel is of course not recommended. This conceptual rejection of mixity is particularly moot if the other, less mixed tests are empirically unfeasible. But mixity is not even formally available for resolving a test into sharper components if an *accepted* fit in the matrix format finds the test to be extreme, even if the test is thus also found not to be sharp. An extraneous extreme test represents a maximally resolved observable wherein the grossly separating output bins do not correspond to a sharp resolution of states into mutually orthogonal compartments. We may therefore look upon a sharp test as an unusually fortunate success in seeking classical Aristotelian sorting, not just freedom from mixity. If there are enough physical tests, it pays to use sharp tests for labeling states. But a paucity of tests owing perhaps to emphasis on tests of a certain type could interfere with this.

If states and tests are unknown except as procedures which can be coupled to yield inaccurately known probabilities, the attempt to choose a special form for some of the matrices could distort the statistical analysis which would be tried to fit MF. Extraneous and undersharp tests would here necessarily be on an equal footing, with perhaps no test being known to be clearly one or the other prior to effecting a fit.

9. UNDERSHARPNESS THEOREMS

Despite the existence of extraneous tests, it is frequently possible to argue undersharpness.

#52 *Theorem*: Fusion preserves undersharpness.

#53. *Proof*: This follows straightforwardly from the obvious lemma that fusion of two bins preserves sharpness. QED

#54. *Theorem*: If the sum of the maximum eigenvalues $p_1, \ldots, p_{b-1}$ of bin matrices $A_1, \ldots, A_{b-1}$ of b-plex A does not exceed 1, then A is undersharp.

#55. *Proof*: Form the commutative and therefore undersharp b-plexes $A^k = (0, \ldots, p_k^{-1}A_k, 0, \ldots, I - p_k^{-1}A_k)$ with zero matrices except in places k and b, using zero for $p_k^{-1}A_k$ if $p_k = 0$. The original b-plex A is represented by $A = \sum_{k=1}^{b-1} p_k A^k + (1 - \sum_{k=1}^{b-1} p_k)\,(0, \ldots, 0, I)$ as a convex combination of undersharp b-plexes. QED

#56 *Corollary*: The body of undersharp b-plexes of $n \times n$ matrices has the same real dimension, $(b-1)n^2$, as the body of all b-plexes.

#57. *Concatenation model*: The following model for an idealized succession of "measurement stages" produces an undersharp test, it if is analyzed in b-plex language: Output states are defined for each bin at each stage m. The idea of an m-output P being split into $m+1$ outputs EPE, FPF by orthogonal projections E, F can be made grammatical in terms of b-plexes, by considering a sharp b-test $A = (A_1, \ldots, A_{b-1}, A_b)$ to be replaced at the next stage by a $(b+1)$-test, $A' = (A_1, \ldots, A_{b-1}, A'_b, A'_{b+1})$, where only bin b has been split, according to $A'_b = A_b E A_b$, $A'_{b+1} = A_b F A_b$. This can be shown to be undersharp by 1-elimination. If A' is defined for A a *specific* convex combination of sharp tests as the corresponding convex combination, then although the coefficients of convex combination to be used in the next stage are not generally defined, they may be shown to be defined if E is of rank 1. Hence the model is limited to a succession of such (E, F) bifurcations with rank $E = 1$.

#52 shows that lumping data together cannot engender extraneity. #54 shows that if there are b-1 bins of interest and one last "beam dump" which collects the lion's share of events (as defined), one hasn't got extraneity.

10. REPERTORIES

Considerations relating to a fit of ideally simple data to MF are developed in order to illustrate the possibility of a fit. #66 provides an idealized lower-bound argument for the format size.

#58. *Lemma*: For A, B nonnegative Hermitian matrices, $A \cdot B \geq 0$. If furthermore $A \cdot B = 0$, then $AB = BA = 0$.

#59. *Proof*: Diagonalize A. $A \cdot B \equiv \mathrm{Tr}\, AB = \sum a_k b_{kk}$ is a sum of nonnegative numbers, whence $A \cdot B \geq 0$.

If $A \cdot B = 0$, $a_k = 0$ or $b_{kk} = 0$ for each k. Renumber the positive a_k so that they lie together on the upper left part of the diagonal. B is zero in these places. The "lemma for offdiagonal zeros" shows B to be zero elsewhere, except in the lower right bloc. Therefore, $AB = 0$. $BA = 0$ merely repeats the conclusion, so as to emphasize that A and B commute. QED

#60. *Trace characterization of sharpness*: A b-plex is sharp if and only if its matrices are mutually trace-orthogonal.

#61. *Proof*: Let the b-plex be $(A_1, \ldots, A_b)$. $A_i A_j = 0$ for $i \neq j$ follows from #58. Therefore, $A_i \sum_{j \neq i} A_j = 0$, hence $A_i(I - A_i) = 0$. Thus $A_i = A_i^2$ is a projection. $A_i A_j = 0$ for $i \neq j$ shows the projections to be mutually orthogonal. QED

#62. *Definition*: A family of b states $P_1, \ldots, P_b$ and one b-test $(A_1, \ldots, A_b)$ such that $P_i \cdot A_j = \delta_{ij}$ will be called a *b-repertory*. An n-repertory will be called a *repertory*.

#63. *Remark*: Since state-test traces are empirically attainable as probabilities, the notion of b-repertory is empirical. #66 therefore exemplifies a manner of gaining an empirical lower bound for n.

#64. *Repertory theorem*: If $P_1, \ldots, P_n$, $(A_1, \ldots, A_n)$ constitute a repertory, then $P_i = A_i$. Thus these are mutually orthogonal 1-dimensional projections.

#65. *Proof*: If $i \neq j$, $P_i \cdot A_j = 0$. By #58, $P_i A_j = 0 = A_j P_i$. Hence $\sum_k P_i A_k = P_i A_i$. But also $\sum_k P_i A_k = P_i I = P_i$. Hence $P_i A_i = P_i = A_i P_i$. Bring A_i and P_i to simultaneously diagonal form $\mathrm{diag}(a_1, \ldots, a_n)$, $\mathrm{diag}(\rho_1, \ldots, \rho_n)$. $P_i A_i = P_i$ reads $a_k \rho_k = \rho_k$, $k = 1, \ldots, n$. Therefore either $a_k = 1$ or $\rho_k = 0$, each k. Therefore $\mathrm{Tr}\, A_i \geq \mathrm{rank}\, P_i \geq 1$, and also $\mathrm{Tr}\, A_i > 1$ unless $a_k = 1$ for precisely one k, in which case $P_i = A_i$ is a 1-dimensional projection. But $\sum_i \mathrm{Tr}\, A_i = \mathrm{Tr}\, I = n = \sum_i 1$ shows that none of the inequalities may hold. QED

#66. *Theorem*: If there is a b-repertory of $n \times n$ matrices, $b \leq n$.

#67. *Proof*: Suppose there is a case with $n < b$. Enlarge the P matrices to $b \times b$ matrices P' by writing extra rows and columns of zeros; enlarge the A matrices to $b \times b$ matrices A' by writing extra diagonal ones and offdiagonal zeros in A'_1, extra zeros generally in $A'_2, \ldots, A'_b$. The new matrices constitute a repertory of $b \times b$ matrices with $P'_1 \neq A'_1$ in violation of #64. QED

11. THE PROBLEM OF COMPUTING MATRICES FROM PROBABILITY DATA

The problem posed by MF is not of course solvable in terms of specific matrix elements, owing to unitary and antiunitary *conjugation* invariance. If enough dot products $P \cdot A$ are given, the dot products of each matrix with the matrices of a sharp n-plex will be known, as exemplified in #64. This will at least fix the solution modulo conjugation.

Since most of MF is given in terms of dot products, it may seem that the invariance is a broader $O(n^2)$ orthogonal invariance in the n^2-dimensional real Euclidean space of Hermitian matrices. The invariance is however cut down to conjugations by the requirement that the P, A be nonnegative, and by the fact that $I \cdot I = n$ does not specify I adequately. Indeed, the positive cone is not orthogonally invariant for $n \geq 3$. The subgroup of $O(n^2)$ which leaves I fixed as a point and which leaves the positive cone fixed as a set is of course precisely the conjugations; this is a restatement of a theorem of Wigner.[10,11]

The direct computation of unknown dot products (prediction of probabilities) from known dot products (data) is conceivable, free of the ambiguity of conjugation, provided that one can usefully state the following condition: The figure delimited by the dot products must be $O(n^2)$-congruent to a subset of the nonnegative cone by a congruence which carries the vector representing I into the positive cone axis. The axis is that subset of the positive cone which remains *pointwise* fixed under all congruences which leave the positive cone setwise fixed. I have been unable to do this explicitly without resorting to extra dot products with a basis, which restores the aspect of conjugation-variant elements in lack of definition of the basis, modulo conjugation. The formal removal of the matrices of a basis by an existence quantifier is of course not computationally effective.

The use of matrices becomes an advantage if very many dot products appear between fewer matrices, with n small. The conjugation invariance does not get any worse as more states and tests become included, so long as the "dynamical system" is the same, i.e., so long as n need not be enlarged. It is indeed the calculation of various properties from few simply determined states and observables which convinces us of the soundness of quantum mechanics. This is to say that MF may be good enough.

The following is *conjecture* on how nevertheless to avoid it.

All dot products between vectors in the computation are required to be nonnegative as a general precaution (#58) and I is not defined beyond $I \cdot I = n$. This will hopefully make the shape staked out by many vectors roughly congruent with the shape of the positive cone, and will identify I by "self-centering" on the axis of this shape, provided that there exists a unique and much overdetermined solution correctly subjected to MF in size n. It is conjectured that sufficient overdetermination gives enough dot products to freeze everything rigidly in the geometry of $O(n^2)$ congruence with only a rule against obtuse angles to replace positivity so as to keep the vectors from fanning out. In other words, if such a very good solution *of the unexamined matrix format* exists, then the following format will find it:

#68. *Sparse format* (*SF*): List all the state and test "matrices" relevant to the experiment consecutively: $(M_1, \ldots, M_{s+b_1+\ldots+b_t}) = (P_1, \ldots, P_s, A_{11}, \ldots, A_{1b_1}, \ldots, A_{t1}, \ldots, A_{tb_t})$, regarding these as symbols, not as matrices, not even as vectors. The symbols $M_R \cdot M_S$ which occur below are to be regarded as known or unknown numbers X_{RS} with the following restrictions, stated where convenient in terms of the old symbols:

$$X_{SR} = X_{RS}, \quad \text{and} \quad X_{RS} \geq 0, \quad \text{each } R, S,$$

$$\text{rank of matrix } (RS \to X_{RS}) \leq n^2,$$

$$P_i \cdot A_{jk} = p_{ijk} \quad \text{for measured probabilities},$$

$$\sum_k A_{jk} \cdot M_s \text{ has a value independent of } j \text{ called } I \cdot M_s,$$

$$\sum_{kk'} A_{jk} \cdot A_{j'k'} = n, \quad \text{each } j, j',$$

and

$$I \cdot P_i = 1.$$

Values of $P_i \cdot A_{jk}$ computed from this for unmeasured probabilities constitute the "predictions."

#69. *Discussion*: It is easy to select vectors from the cone of all vectors at angles $\leq 45°$ from an axis, whose mutual dot products satisfy $X_{RS} \geq 0$, but which form a shape not congruent to a subset of the nonnegative cone—when $n \geq 3$. I is of course laid off along the axis. This exemplifies a solution of SF "spurious" in not corresponding to a solution of MF. The true nonnegative cone is more complicated. But if some of the data themselves fill in the shape of the true nonnegative cone, then nonobtuseness would hold the rest to the nonnegative shape. Unfortunately, it is hard to prove that any data will freeze in SF, even if the same data freezes, say, by virtue of #64, in MF.

SF itself may be taken as a generalization of quantum mechanics (can it make sense if n is not an integer

but n^2 is?) provided that SF may freeze itself, before all the p_{ijk} are given. (Freezing of the unphysical $P \cdot P$ and $A \cdot A$ dot products is unclear even when all p_{ijk} are given.) If freezing to shapes which won't fit into the nonnegative cone is possible, it is pertinent to ask whether actual data support ordinary quantum mechanics by SF-freezing to shapes which do fit into the nonnegative cone without being told to do so in advance through MF.

Although no formal invariance blocks SF from freezing on definite values X_{RS}, both SF and MF might be approached in practice by varying parameters so as to reduce the numerical discrepancies between $P_i \cdot A_{jk}$ format expressions and p_{ijk} data. Such a computational approach incorporating deus-ex-machina introduction of numbers to break symmetry deadlocks (e.g., choice of i or $-i$ to first step away from the reals in MF), would proceed to successively improved solutions without any particular difficulty owing to existence of conjugation-equivalent solutions. In practice, therefore, the conjugation reason for seeking to replace MF by SF could be moot.

Indeed, direct inference from measured probabilities to predicted probabilities is in itself unlikely to reveal much. MF, SF are more likely to be useful with the help of exhaustive data to search for low-n fits, for the purpose of seeking the familiar matrices of simple quantum-mechanical physics in unfamiliar contexts.

ACKNOWLEDGMENTS

The detection of mathematical errors by my wife Thelma and some transfer of her involvement with economics are acknowledged. The interest of Professor David Finkelstein and of Professor C.A. Hooker has been a stimulus to writing.

[1] L.D. Landau, Z. Physik 45, 430 (1927).
[2] J. von Neumann, Göttinger Nachrichten, 1927.
[3] H. Weyl, Z. Physik 46 (1927).
[4] R. Giles, J. Math. Phys. 11, 2139 (1970), where the author discusses mixed 2-tests.
[5] "Statistics Conference," May 1972 at the University of Western Ontario, London, Ontario. A proceedings volume edited probably by A.C. Hooker and W.L. Harper is planned.
[6] G. Köthe, *Topological Vector Spaces* (Springer, New York, 1969), Vol. I, p. 332.
[7] N. Bourbaki, *Éléments de mathématique XV, Livre V, Espaces vectorielles topologiques, Act. Sci. et Ind. No.* 1189 (Hermann, Paris, 1953), Chap. I, p. 85, Prop. 4.
[8] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. Beyer (Princeton U.P., 1955).
[9] H. Everett, III, Rev. Mod. Phys. 29, 454 (1957).
[10] V. Bargmann, J. Math. Phys. 5, 862 (1964).
[11] E.P. Wigner, *Group Theory* (Academic, New York, 1959), pp. 233—36.

A physical system which can be forced to execute an arbitrary unitary transformation, and its use to perform arbitrary tests

Elihu Lubkin

Physics Department, The University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

(Received 31 January 1974)

It is shown that a system of N spin-(1/2) magnets can be constructed so freely manipulable that an experimenter can impose any unitary transformation upon its state space, N any positive integer. This is then used with $N \geq b-1+2\log_2 n$ to construct arbitrary states over an n-dimensional Hilbert space and to perform arbitrary b-tests upon these states.

1. INTRODUCTION

In a previous paper (MB) on multibin tests,[1] it was shown that b-bin tests of all sorts can be achieved in the ordinary atomic physics of a system associated with an n-dimensional Hilbert space, n finite, provided that an experimenter could force a probe-system complex to undergo motion $P \to UPU^\dagger$ with a unitary conjugation which may be selected freely by the experimenter.

It is shown here how this may actually be done for a specific class of systems. The constructibility of general unitary motions is the essence of this, and may be of more general interest than its use in MB. Hence the free-unitary construction is first presented separately, the details of application to MB being done at the end, in Sec. 4.

2. ACHIEVABILITY

A physical system is given together with certain basic motions. The states of the system are described by density matrices P, and the basic motions by unitary conjugations, $P \to UPU^\dagger$. These will correspond in the usual way to basic Hamiltonians H: $U=\exp(-iHt)$, where time t may be chosen free nonnegative by the experimenter. By varying external fields and other external parameters, the experimenter may also choose various Hamiltonians. By succession of these operations, the semigroup of *achievable* unitary transformations of all products of basic operations U may be attained by the experimenter. If all elements of the 1-parameter semigroup $\{\exp(-iHt)\,|\,t \geq 0\}$ are achievable, then H will be said to be *achievable*, too, even if H is not a basic Hamiltonian. In recognition of the nonempirical character of sufficiently small discrepancies, a limit of achievable motions or of Hamiltonians will also be said to be *achievable*. The U will be of determinant 1, the H of trace 0, without loss of generality.

The sequel is limited to systems with finite-dimensional Hilbert spaces. This immediately turns the closure of our semigroup into a group. This is most conveniently formulated for the ensuing computations in infinitesimal terms, in #1 and #3:

#1. *Inversion lemma*: If H is achievable, then $-H$ is achievable.

#2. *Proof*: It suffices to show that if U is achievable, then U^{-1} is a limit of achievable matrices. The positive powers U, U^2, U^3, $\cdots$ are all achievable by repetition. Compactness of the set of unitary matrices implies that either U is of finite order r, so that $U^{r-1}=U^{-1}$ is directly achievable, or else that there is a point of accumulation and also a Cauchy subsequence U^{p_1}, U^{p_2}, $\cdots$ of increasing powers near it. Hence for any $\epsilon>0$, there are integers r, s, $r>s>0$, with $\|U^{p_r}-U^{p_s}\|<\epsilon$. If W is unitary, $\|WA\|=(\mathrm{Tr}A^\dagger W^\dagger WA)^{1/2}=\|A\|$. Hence, also $\|U^{p_r-p_s-1}-U^{-1}\|<\epsilon$. Since p_r-p_s-1 is a nonnegative integer, $U^{p_r-p_s-1}$ is achievable. QED

#3. *Theorem*: If a set of Hermitean matrices is achievable, then its Lie algebra is also achievable.

#4. *Sketch of proof*: To achieve a combination of $H_1, \ldots, H_m$ with small nonnegative coefficients $p_1,\ldots,p_m$, effect H_1 motion for a time proportional to p_1, then H_2 motion for a time proportional to p_2, etc. To first order, the resulting motion, $\exp(-iH_m p_m)\cdots\times\exp(-iH_1 p_1)$ approximates the desired $\exp(-i\Sigma p_k H_k)$. In order to avoid the second-order commutators for large p_k, p_k may be replaced by p_k/ν for large ν, with the whole process "compounded" ν times.

The inversion lemma immediately extends this to arbitrary real linear combinations.

If A and B are achievable, then a motion generated by the Lie bracket $i(AB-BA)$ for a short time t^2 is approximated by performing the A motion for time t, the B motion for time t, the $-A$ motion (approximately if necessary) for time t, then the $-B$ motion (approximately) for time t; the result is $\exp[-it^2 i(AB-BA)]$ to order t^2. (The approximations for $-A$, $-B$ must be good to order t^2.) Larger times may be attained by compounding. QED

3. THE SPIN SYSTEM

#5. *Notation*: The tensor products $I\otimes\cdots\otimes\sigma_i\otimes\cdots\otimes I$ with Pauli matrix σ_i (i.e., either σ_1, σ_2, or σ_3) as the kth factor and 2×2 unit matrices I as the other $N-1$ tensor factors will be denoted σ_{ki}. A tensor product of l Pauli matrices and $N-l$ unit matrices will be said to be *an elementary matrix of length l*. A linear combination of such objects of a fixed length l will also be said to be of length l. $\Sigma_{i=1}^3 \sigma_{ki}\sigma_{mi}=\boldsymbol{\sigma}_k\cdot\boldsymbol{\sigma}_m$.

#6. *Theorem*: The real Lie algebra of Hermitian matrices with bracket operation $[A,B]=i(AB-BA)$ generated by the σ_{ki} (all k) and the $\boldsymbol{\sigma}_k\cdot\boldsymbol{\sigma}_m$ (all k,m) is the set of all traceless Hermitian $2^N\times2^N$ matrices.

#7. *Proof*: The elementary matrices of length ≥ 1 are Hermitian, traceless, and mutually trace-orthogonal. There are 4^N-1 such elementary matrices. Hence they constitute a real vector-space basis for the (4^N-1)-real-dimensional space of all traceless Hermitian $2^N\times2^N$ matrices.

It therefore suffices to show that the given generators generate all the elementary matrices of length ≥ 1.

Those of length 1 are given, together with special combinations of length 2. The general objects of length 2 can be obtained by computing a few commutators: $\frac{1}{2}[\sigma_1\otimes I, \sigma_1\otimes\sigma_1+\sigma_2\otimes\sigma_2+\sigma_3\otimes\sigma_3]=\sigma_2\otimes\sigma_3-\sigma_3\otimes\sigma_2$. Then $-\frac{1}{2}[\sigma_2\otimes I, \sigma_3\otimes\sigma_2-\sigma_2\otimes\sigma_3]=\sigma_1\otimes\sigma_2$. Finally $\frac{1}{2}[I\otimes\sigma_3, \sigma_1\otimes\sigma_2]=\sigma_1\otimes\sigma_1$. Thus all $\sigma_i\otimes\sigma_j$ are generated. If $N-2$ extra tensor factors I are carried in all ways, all the elementary matrices of length 2 are obtained.

$-\frac{1}{2}[\sigma_i\otimes\sigma_2\otimes I, I\otimes\sigma_3\otimes M]=\sigma_i\otimes\sigma_1\otimes M$ shows how to increase the length of M by 2, essentially arbitrarily. Since all objects of lengths 1 and 2 are already available, this length-increasing operation inductively reaches to the remaining elementary matrices of all lengths ≥ 1. QED

#8. *The physical spin system*: The kinematical situation of #5 obviously suits a system of N spin-$\frac{1}{2}$ subsystems, or "spins." Let these have magnetic moments. Construct these so well mutually isolated that a magnetic field can be applied to any single spin in an arbitrary direction, while the other $N-1$ spins are at 0 field. Hence the σ_{ki} are achievable (and the $-\sigma_{ki}$ even without using #1). Furthermore, let the spins be independently movable so that any two spins (the kth and mth) can be juxtaposed so closely as to develop a nonzero Heisenberg interaction Hamiltonian $\lambda\boldsymbol{\sigma}_k\cdot\boldsymbol{\sigma}_m$, $\lambda\neq 0$, for any chosen length of time, then again separated: while one pair is so interacting, the $N-2$ other spins do not move. Now the $\boldsymbol{\sigma}_k\cdot\boldsymbol{\sigma}_m$ (or the $-\boldsymbol{\sigma}_k\cdot\boldsymbol{\sigma}_m$ if $\lambda<0$) are also achievable, whence from #6 all motions are achievable.

4. APPLICATION TO THE CONSTRUCTION OF MULTIBIN TESTS

The N spins will be used to construct both the state space B_2 (an n-dimensional Hilbert space V_n) and the probe space D (a V_b), needed in MB. First a $V_{n^2}=B$ is constructed instead of a V_n for the luxury of obtaining arbitrary mixed states[2] without outside rouletting. In order to make the measurement of the probe simply feasible by means of Stern–Gerlach experiments, $b-1$ other spins, enough in principle for a $V_{2^{b-1}}$, will be used to construct D. Thus,

$$2^{N-(b-1)}\geq n^2,$$

i.e., $N\geq b-1+2\log_2 n$ spins are needed for the present construction.

Let the whole thing first be set with all N spins down, say, by using a thermal contact with a sufficiently cold reservoir and an overall magnetic field. Call this state Vac_N. Then contact with the reservoir is broken, and the reservoir is not used any further.

Achievability of arbitrary unitary motions is so convenient that the arbitrary states as well as the arbitrary tests are constructed by using motions. Vac_N is a tensor product of states down for each spin; in particular, $\mathrm{Vac}_N=\mathrm{Vac}_{N-b+1}\otimes\mathrm{Vac}_{b-1}$, with $\mathrm{Vac}_{N-b+1}\in A$, $\mathrm{Vac}_{b-1}\in C$, A and C being the $V_{2^{N-b+1}}$ and the $V_{2^{b-1}}$ corresponding to a partition of the spins into $N-b+1$ spins and $b-1$ spins. Choose an n^2-dimensional sub-Hilbert space B of A such that $\mathrm{Vac}_{N-b+1}\in B$, and choose B_1 and B_2 so that $B=B_1\otimes B_2$, where both B_1 and B_2 are V_n's. Thus the whole V_{2^N} is of form $()B_1\oplus B_2)\oplus R)\oplus C$, where R is the orthogonal complement of B in A.

In order to produce an arbitrary pure state $|x\rangle\langle x|$ with $x\in B$ without other effect, perform a unitary motion which rotates Vac_{N-b+1} into x and leaves Vac_{b-1} unchanged. Landau-tracing $|x\rangle\langle x|$ over B_1 produces an arbitrary mixed state[2] ρ over B_2 appropriate for computing probabilities against the subsequent test, since the test will not involve B_1. [Indeed, to get general $\rho=\Sigma_{i=1}^n w_i|x_i\rangle\langle x_i|$, $w_i\geq 0$, $\Sigma w_i=1$, $x_i\in B_2$, $\langle x_i|x_j\rangle=\delta_{ij}$, arrange to produce the B state $\Sigma_{i=1}^n w_i^{1/2} b_i\otimes x_i$, where $(b_1,\ldots,b_n)$ is any convenient orthonormal basis of B_1.]

C is a $V_{2^{b-1}}$, a system of $b-1$ spin-$\frac{1}{2}$'s. $\mathrm{Vac}_{b-1}\in C$ has all these down, hence has total angular momentum $j=\frac{1}{2}(b-1)$ and "z" component $j_z=-j$. The $(b=2j+1)$-dimensional subspace corresponding to this j for more general j_z will be the probe space, D. Thus, $C=D\oplus R'$, with R' another residual space.

The general unitary motion of system and probe together required in MB will now be imposed upon the space $B_2\otimes D$, with no motion in B_1, R, or R'.

The final measurement of the probe required in MB will seek its j_z value. Since this value is the sum of separate $\frac{1}{2}\sigma_3$ values of the $b-1$ spins which constitute the C system, it can be measured by performing separate spin-$\frac{1}{2}$ Stern–Gerlach experiments: The construction has the spins so foreign to each other for other purposes as to make a simpler, direct measurement of the j_z of system C or D possibly awkward.

The arguments given, including the processes of generation of Lie algebra elements, are all constructive, so that experimental procedure for demonstrating arbitrary tests, including the extraneous tests of MB, has been given. The fact that such general tests can in principle be effected is, however, of more interest than a laboratory demonstration, hence no care has been taken to seek *easy* laboratory procedures. The moral is already made in MB, that a rule in any way restricting b-tests of $n\times n$ matrices to any set more limited than b-tuples of nonnegative Hermitian matrices which sum to the unit matrix, could not be tolerated in conventional quantum mechanics. This provides guidance in extracting rules of quantum epistemology from atomic physics, for possible application elsewhere.

[1] E. Lubkin, J. Math. Phys. 14, xxx (1974), hereafter designated MB. Terminology of MB will be used freely.

[2] Ability to produce an arbitrary *pure* state in B_2 would have sufficed to compute the test matrices of the subsequent test. If arbitrary mixed states are nevertheless desired they can instead be obtained by rouletting pure states. Thus the device of constructing and using B_1 is not strictly needed for the purpose of establishing the existence of arbitrary tests. The device is adapted from J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by R. Beyer (Princeton U.P., Princeton, 1955).

Eigenvalue problem for Lagrangian systems. VII*

E. M. Barston

Department of Mathematics, University of Illinois at Chicago Circle, Chicago, Illinois 60680
(Received 9 November 1973)

Minimax principles for a subset of the real eigenvalues of the quadratic eigenvalue problem $(\omega^2 A + \omega B + C)\xi = 0$ are presented, where A, B, and C are formally self-adjoint operators mapping a dense subspace Δ of a complex Hilbert space E into E, and $A > 0$. These results are applied to the problems of the small oscillations about equilibrium of a vertically stratified, viscous, heterogeneous incompressible fluid in a gravitational field and the oscillations of a rotating thin annular disk, and it is shown that the minimax principles characterize infinitely many eigenvalues of these systems.

I. INTRODUCTION

The linear analysis of the oscillations and/or stability of many conservative dynamical systems about states of steady motion (gyroscopic Lagrangian systems) and certain nonconservative dynamical systems about states of equilibrium leads to the quadratic eigenvalue problem

$$(\omega^2 A + \omega B + C)\xi = 0,$$

where A, B and C are formally self-adjoint linear operators mapping a dense subspace Δ of a complex Hilbert space E into E, the eigenvalue ω is a real or complex number, and ξ (an eigenvector corresponding to the eigenvalue ω) is a nonzero element of Δ.[1-11] A fairly large body of literature dealing with various aspects of this problem and certain generalizations has recently developed, with the major part devoted to the establishment of the completeness properties of the eigenvectors and generalized eigenvectors in certain special cases. An extensive list of references can be found in Ref. 12. Relatively little attention has been paid to the important question of stability.[2-5,8-11,13,14]

Several authors[15-20] have considered the subclass of "overdamped" systems, i.e., those systems where A, B, and C satisfy the additional restriction (called the overdamping condition) $(\eta, B\eta)^2 - 4(\eta, A\eta)(\eta, C\eta) > 0$ for all nonzero $\eta \in \Delta$. In particular, Duffin[15] obtained minimax principles for the eigenvalues of finite-dimensional overdamped systems, and Turner[18] has obtained corresponding results for a special class of infinite-dimensional overdamped systems.

In this paper, with suitable restrictions placed on the operators A, B, and C so as to insure the existence of a point spectrum, we obtain minimax principles for a subset of the real eigenvalues of nonoverdamped systems. Our results generalize previous results for overdamped systems and contain them as a special case. An extension of this type is highly desirable in view of the fact that the vast majority of problems arising from the analysis of physical systems do not satisfy the overdamping condition. A different set of minimax principles for finite-dimensional nonoverdamped systems has been given by the author in Ref. 21.

Two applications of these results are discussed in Sec. IV, after the development of the minimax principles in Secs. II and III. The first application concerns the well-known problem of small oscillations about equilibrium of a vertically stratified, viscous, heterogeneous incompressible fluid in a gravitational field. The second concerns the oscillations of a rotating thin annular disk. We show, in particular, that infinitely many of the eigenvalues of these systems are given by the minimax principles developed herein.

II. THE MINIMAX PRINCIPLE

We adopt the following hypothesis, denoted as (H1), throughout this paper: Let E be a complex Hilbert space, Δ an infinite-dimensional subspace of E, and let A, B, and C be linear formally self-adjoint operators from Δ into E [an operator F is formally self-adjoint on Δ provided that for all η, $\zeta \in \Delta$, $(\eta, F\zeta) = (F\eta, \zeta)$]. Further restrictions on A, B, and C will be imposed as we proceed.

We define $C_\omega \equiv \omega^2 A + \omega B + C$, $-\infty < \omega < \infty$. For each real ω, C_ω maps Δ into E and is formally self-adjoint on Δ. Let S_k denote the set of all k-dimensional subspaces of Δ, $k = 1, 2, 3, \cdots$. For each positive integer k and every real ω, we define the extended real-valued function

$$\Lambda_k(\omega) \equiv \sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, C_\omega \xi)}{(\xi, \xi)} \tag{1}$$

(it is to be understood that the infemum is over all nonzero $\xi \in V$). Clearly, $\Lambda_k(\omega) \geq \Lambda_{k+1}(\omega)$, $k = 1, 2, 3, \cdots$. If A, B, and C are bounded above on Δ by a, b_+, and c, respectively $[(\xi, A\xi) \leq a\|\xi\|^2$ for all $\xi \in \Delta$, etc., where a, b_+ and c are finite], then for $0 \leq \omega < \infty$ $\Lambda_k(\omega)$ is finite and is bounded above by $\omega^2 a + \omega b_+ + c$, while if A and C are bounded above and B is bounded below by b_- on Δ $[(\xi, B\xi) \geq b_-\|\xi\|^2$ for some finite b_- and all $\xi \in \Delta]$, then, for $-\infty < \omega \leq 0$, $\Lambda_k(\omega)$ is finite and is bounded above by $\omega^2 a + \omega b_- + c$. Let $\Phi \equiv \{\phi_1, \phi_2, \ldots, \phi_k\}$ denote any set of k orthonormal vectors from Δ. Then $V \equiv \text{span}\Phi \in S_k$, and $\inf_{\xi \in V}[(\xi, C_\omega \xi)/(\xi, \xi)]$ equals the least eigenvalue $\lambda_k^\Phi(\omega)$ of the $k \times k$ Hermitian matrix whose ijth element is $(\phi_i, C_\omega \phi_j)$, $i, j = 1, \cdots, k$. Hence

$$\lambda_k^\Phi(\omega) \leq \Lambda_k(\omega) \tag{2}$$

for all positive integers k and all real ω. In particular, if C is bounded above by c and A and B are bounded on Δ, then $\Lambda_k(\omega)$ is finite for all real ω and we have

$$\lambda_k^\Phi(\omega) \leq \Lambda_k(\omega) \leq \omega^2\|A\| + |\omega|\|B\| + c. \tag{3}$$

Furthermore, $\Lambda_k(\omega)$ is continuous, as we now show.

Lemma 1: Let C be bounded above and A and B be bounded on Δ. Then for each fixed positive integer k, $\Lambda_k(\omega)$ is continuous in ω, $-\infty < \omega < \infty$.

Proof: Let ω and δ be real.

$$\Lambda_k(\omega \pm \delta) = \sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, C_{\omega \pm \delta}\xi)}{(\xi, \xi)}$$

$$= \sup_{V\in S_k} \inf_{\xi\in V} \left(\frac{(\xi, C_\omega \xi)}{(\xi,\xi)} + (\delta^2 \pm 2\omega\delta) \frac{(\xi, A\xi)}{(\xi,\xi)} \pm \delta \frac{(\xi, B\xi)}{(\xi,\xi)} \right)$$

$$\leq \sup_{V\in S_k} \inf_{\xi\in V} \left(\frac{(\xi, C_\omega \xi)}{(\xi,\xi)} + (\delta^2 + 2|\omega||\delta|)\|A\| + |\delta|\,\|B\| \right),$$

so that

$$\Lambda_k(\omega \pm \delta) \leq \Lambda_k(\omega) + |\delta|[(2|\omega| + |\delta|)\|A\| + \|B\|]. \quad (4)$$

Replace ω by $\omega \mp \delta$ in Eq. (4) to obtain

$$\Lambda_k(\omega) \leq \Lambda_k(\omega \mp \delta) + |\delta|[(2|\omega| + 3|\delta|)\|A\| + \|B\|]. \quad (5)$$

Equations (4) and (5) imply

$$|\Lambda_k(\omega+\delta) - \Lambda_k(\omega)| \leq |\delta|[(2|\omega| + 3|\delta|)\|A\| + \|B\|],$$

which proves the continuity of $\Lambda_k(\omega)$.

We use this result together with some additional assumptions on A, B, and C to establish the existence of zeros of $\Lambda_k(\omega)$.

Theorem 1: Let C be bounded above and A and B be bounded on Δ. Suppose that for some real Ω there exists a subspace Δ_+ of Δ with $M \equiv \dim \Delta_+ < \infty$ such that $\sup_{\xi\in\Delta_-}[(\xi, C_\Omega\xi)/(\xi,\xi)] < 0$, where $\Delta_- \equiv \Delta_+^\perp \cap \Delta$. Suppose further that $A > 0$ on some subspace Σ of Δ with $N \equiv \dim \Sigma > M$. Then for each positive integer k satisfying $M+1 \leq k \leq N$, there exists $\omega_k^+ \in (\Omega, \infty)$ and $\omega_k^- \in (-\infty, \Omega)$ such that $\omega_{k+1}^+ \geq \omega_k^+$, $\omega_{k+1}^- \leq \omega_k^-$, and $\Lambda_k(\omega_k^\pm) = 0$.

Proof. Let $\Phi = \{\phi_j\}_1^N$ be an orthonormal set of elements from Σ. Let Σ_k denote the span of $\{\phi_1, \cdots, \phi_k\}$. Then $A > 0$ on Σ implies that $\inf_{\xi\in\Sigma_k}[(\xi, A\xi)/(\xi,\xi)] > 0$, $1 \leq k \leq N$.

Since

$$\lambda_k^\Phi(\omega) \equiv \inf_{\xi\in\Sigma_k} \frac{(\xi, C_\omega\xi)}{(\xi,\xi)} = \inf_{\xi\in\Sigma_k} \left(\omega^2 \frac{(\xi, A\xi)}{(\xi,\xi)} + \omega \frac{(\xi, B\xi)}{(\xi,\xi)} + \frac{(\xi, C\xi)}{(\xi,\xi)} \right)$$

$$\geq \omega^2 \inf_{\xi\in\Sigma_k} \frac{(\xi, A\xi)}{(\xi,\xi)} - |\omega|\,\|B\| + \inf_{\xi\in\Sigma_k} \frac{(\xi, C\xi)}{(\xi,\xi)},$$

it follows that $\lambda_k^\Phi(\omega) \to \infty$ as $|\omega| \to \infty$ for $1 \leq k \leq N$, and so Eq. (2) implies that $\Lambda_k(\omega) \to \infty$ as $|\omega| \to \infty$ for $1 \leq k \leq N$. Let $k \geq M+1$ and $V \in S_k$. Then there exists a nonzero $\eta \in V \cap \Delta_-$, so that

$$\inf_{\xi\in V} \frac{(\xi, C_\Omega\xi)}{(\xi,\xi)} \leq \frac{(\eta, C_\Omega\eta)}{(\eta,\eta)} \leq \sup_{\xi\in\Delta_-} \frac{(\xi, C_\Omega\xi)}{(\xi,\xi)}.$$

Thus for $k \geq M+1$, $\Lambda_k(\Omega) \leq \sup_{\xi\in\Delta}[(\xi, C_\Omega\xi)/(\xi,\xi)] < 0$. The existence of zeros $\omega_k^\pm$ of Λ_k in $(-\infty, \Omega)$ and (Ω, ∞) now follows from the continuity of $\Lambda_k(\omega)$ (Lemma 1) and the fact that $\Lambda_k(\omega) \to \infty$ as $|\omega| \to \infty$ for $1 \leq k \leq N$. Since $\Lambda_k(\omega) \geq \Lambda_{k+1}(\omega)$ for all real ω and all positive integers k, it follows that $\Lambda_{k+1}(\omega_k^\pm) \leq 0$ and therefore that $\Lambda_{k+1}(\omega)$ has a zero in $(-\infty, \omega_k^-]$ and in $[\omega_k^+, \infty)$, $M+1 \leq k \leq N-1$. Hence we may always select the $\omega_k^\pm$ so that $\omega_{k+1}^+ \geq \omega_k^+$ and $\omega_{k+1}^- \leq \omega_k^-$.

We proceed to obtain a minimax principle for certain of the zeros established in Theorem 1; however, we shall adopt a somewhat modified hypothesis:

(H2) For some real Ω, there exists a subspace Δ_+ of Δ with $M \equiv \dim \Delta_+ < \infty$ such that $\sup_{\xi\in\Delta}[(\xi, C_\Omega\xi)/(\xi,\xi)] \leq 0$, where $\Delta_- \equiv \Delta_+^\perp \cap \Delta$.

(H3) $A > 0$ on Δ. Let $D \equiv \{\xi \mid \xi \in \Delta,\ d(\xi) \geq 0\}$, where

$$d(\xi) \equiv (\xi, B\xi)^2 - 4(\xi, A\xi)(\xi, C\xi) = (\xi, B_\Omega\xi)^2 - 4(\xi, A\xi)(\xi, C_\Omega\xi),$$

where $B_\Omega \equiv 2\Omega A + B$. Clearly $\Delta_- \subset D$. For all nonzero $\xi \in D$, we define

$$Q_\pm(\xi) \equiv \frac{-(\xi, B\xi) \pm [d(\xi)]^{1/2}}{2(\xi, A\xi)}. \quad (6)$$

(H4) $\Gamma_+ \equiv \inf_{\xi\in D} Q_+(\xi) > -\infty$, $\Gamma_- \equiv \sup_{\xi\in D} Q_-(\xi) < \infty$. Let $k \geq M+1$. Then for any $V \in S_k$, $V \cap \Delta_-$ contains a nonzero element so that $V \cap D$ is nontrivial. Hence the real numbers

$$\Omega_k^+ \equiv \inf_{V\in S_k} \sup_{\xi\in V\cap D} Q_+(\xi), \quad k \geq M+1, \quad (7)$$

and

$$\Omega_k^- \equiv \sup_{V\in S_k} \inf_{\xi\in V\cap D} Q_-(\xi), \quad k \geq M+1, \quad (8)$$

are well-defined and we have

$$\Gamma_+ \leq \Omega_{M+1}^+ \leq \Omega_{M+2}^+ \leq \cdots \leq \Omega_k^+ \leq \Omega_{k+1}^+ \leq \cdots, \quad (9)$$

$$\Gamma_- \geq \Omega_{M+1}^- \geq \Omega_{M+2}^- \geq \cdots \geq \Omega_k^- \geq \Omega_{k+1}^- \geq \cdots. \quad (10)$$

With the aid of the following lemma, we obtain sufficient conditions that $\Lambda_k(\Omega_k^\pm) = 0$.

Lemma 2. Let (H1) and (H3) hold, let λ be a real number, $\epsilon > 0$, $2\omega A + B$ be uniformly bounded above on Δ by γ for all $\omega \in (\lambda, \lambda+\epsilon)$, and let V be a subspace of Δ such that for every nonzero $\xi \in V \cap D$, $Q_+(\xi) < \lambda + \epsilon$. Then for all nonzero $\xi \in V$, $(\xi, C_\lambda\xi)/(\xi,\xi) \geq -\epsilon|\gamma|$.

Proof: Let $\xi \in V$, $\xi \neq 0$. If $\xi \notin D$, then $(\xi, C_\omega\xi)/(\xi,\xi) > 0$ for all real ω. Suppose $\xi \in D$. If $Q_+(\xi) \leq \lambda$ or $Q_-(\xi) \geq \lambda$, then $(\xi, C_\lambda\xi)/(\xi,\xi) \geq 0$. If $Q_-(\xi) < \lambda < Q_+(\xi)$, then by the mean-value theorem

$$f_\xi[Q_+(\xi)] - f_\xi(\lambda) = f_\xi'(x)[Q_+(\xi) - \lambda]$$

for some $x \in (\lambda, Q_+(\xi))$, where $f_\xi(\omega) \equiv (\xi, C_\omega\xi)/(\xi,\xi)$, $-\infty < \omega < \infty$. Now $f_\xi[Q_+(\xi)] = 0$, $0 < Q_+(\xi) - \lambda < \epsilon$, and $f_\xi'(x) = (\xi, [2xA + B]\xi)/(\xi,\xi) \leq |\gamma|$. Therefore, $-f_\xi(\lambda) \leq |\gamma|\epsilon$, i.e., $(\xi, C_\lambda\xi)/(\xi,\xi) \geq -\epsilon|\gamma|$. Hence, in any case, $(\xi, C_\lambda\xi)/(\xi,\xi) \geq -\epsilon|\gamma|$.

Let J_+ denote the set of all positive integers n for which there exists an $\epsilon(n) > 0$ such that $2\omega A + B$ is uniformly bounded above on Δ for all $\omega \in (\Omega_n^+, \Omega_n^+ + \epsilon)$. Similarly, we define J_- to be the set of all n for which an $\epsilon(n) > 0$ exists such that $2\omega A + B$ is uniformly bounded below on Δ for all $\omega \in (\Omega_n^- - \epsilon, \Omega_n^-)$.

Theorem 2: Let (H1)—(H4) hold. Then if $\Omega_l^+ \geq \Gamma_-$ for some $l \geq M+1$, $\Lambda_k(\Omega_k^+) = 0$ for all $k \geq l$, $k \in J_+$.

Proof: Since $\Gamma_- \leq \Omega_l^+ \leq \Omega_k^+$ for $k \geq l$, it suffices to show that $\Omega_k^+ \geq \Gamma_-$, $k \in J_+$, implies $\Lambda_k(\Omega_k^+) = 0$. Suppose $\Omega_k^+ \geq \Gamma_-$,

and let $V \in S_k$. Then

$$\Omega_k^+ \leq \sup_{\xi \in V \cap D} Q_+(\xi) = \sup_{\xi \in V \cap D_0} Q_+(\xi),$$

where $D_0 \equiv \{\xi \mid \xi \in D,\ \|\xi\| = 1\}$. Since $\dim V < \infty$, $V \cap D_0$ is compact and $Q_+(\xi)$ is continuous on $V \cap D_0$ and thus assumes its supremum there. Hence there exists a nonzero $\xi \in V \cap D$ such that $Q_+(\xi) \geq \Omega_k^+ \geq \Gamma_- \geq Q_-(\xi)$, which implies that $(\xi, C_{\Omega_k^+}\xi)/(\xi, \xi) \leq 0$; therefore, $\inf_{\xi \in V}[(\xi, C_{\Omega_k^+}\xi)/(\xi, \xi)] \leq 0$. This last result holds for arbitrary $V \in S_k$, and we conclude from Eq. (1) that $\Lambda_k(\Omega_k^+) \leq 0$. Let $k \in J_+$ and $\Omega_k^- \geq \Gamma_-$. Then there exists $\epsilon_0 > 0$ and γ_0 such that $2\omega A + B$ is bounded above on Δ by γ_0 for all $\omega \in (\Omega_k^+, \Omega_k^+ + \epsilon_0)$. Given any positive $\epsilon < \epsilon_0$, there exists $V_\epsilon \in S_k$ such that $\Omega_k^+ \leq \sup_{\xi \in V_\epsilon \cap D} Q_+(\xi) < \Omega_k^+ + \epsilon$. Hence for all nonzero $\xi \in V_\epsilon \cap D$, $Q_+(\xi) < \Omega_k^+ + \epsilon$, and Lemma 2 implies $(\xi, C_{\Omega_k^+}\xi)/(\xi, \xi) \geq -\epsilon|\gamma_0|$ for all nonzero $\xi \in V_\epsilon$. Therefore, $\inf_{\xi \in V_\epsilon}[(\xi, C_{\Omega_k^+}\xi)/(\xi, \xi)] \geq -\epsilon|\gamma_0|$ so that $0 \geq \Lambda_k(\Omega_k^+) = \sup_{V \in S_k} \inf_{\xi \in V}[(\xi, C_{\Omega_k^+}\xi)/(\xi, \xi)] \geq -\epsilon|\gamma_0|$. Since this holds for all sufficiently small $\epsilon > 0$, $\Lambda_k(\Omega_k^+) = 0$.

Theorem 3: Let (H1)–(H4) hold. Then if $\Omega_l^- \leq \Gamma_+$ for some $l \geq M+1$, $\Lambda_k(\Omega_k^-) = 0$ for all $k \geq l$, $k \in J_-$.

Proof: Let $\tilde{B} \equiv -B$. The proof consists of replacing B by $\tilde{B}$ and using Theorem 2. Consider the triple of operators $(A, \tilde{B}, C)$ in place of (A, B, C). Obviously $(A, \tilde{B}, C)$ satisfies (H1) and (H3) since (A, B, C) does. For the triple $(A, \tilde{B}, C)$, C_ω is replaced by $\tilde{C}_\omega \equiv \omega^2 A + \omega\tilde{B} + C = C_{-\omega}$. Since (H2) holds for (A, B, C), it holds for $(A, \tilde{B}, C)$ with the same Δ_+ and Δ_-, but with $-\Omega$ replacing Ω. Corresponding to $\Lambda_k(\omega)$, $d(\xi)$, D, $Q_\pm(\xi)$, $\Gamma_\pm$, and $\Omega_k^\pm$, we have:

$$\tilde{\Lambda}_k(\omega) \equiv \sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, \tilde{C}_\omega \xi)}{(\xi, \xi)} = \Lambda_k(-\omega),$$

$$\tilde{d}(\xi) \equiv (\xi, \tilde{B}\xi)^2 - 4(\xi, A\xi)(\xi, C\xi) = d(\xi),$$

$$\tilde{D} = \{\xi \mid \xi \in \Delta,\ \tilde{d}(\xi) \geq 0\} = D,$$

$$\tilde{Q}_\pm(\xi) \equiv \frac{-(\xi, \tilde{B}\xi) \pm [\tilde{d}(\xi)]^{1/2}}{2(\xi, A\xi)} = -Q_\mp(\xi),$$

$$\tilde{\Gamma}_+ \equiv \inf_{\xi \in \tilde{D}} \tilde{Q}_+(\xi) = \inf_{\xi \in D}[-Q_-(\xi)] = -\Gamma_- > -\infty,$$

$$\tilde{\Gamma}_- \equiv \sup_{\xi \in \tilde{D}} \tilde{Q}_-(\xi) = \sup_{\xi \in D}[-Q_+(\xi)] = -\Gamma_+ < \infty,$$

$$\tilde{\Omega}_k^+ \equiv \inf_{V \in S_k} \sup_{\xi \in V \cap \tilde{D}} \tilde{Q}_+(\xi) = -\sup_{V \in S_k} \inf_{\xi \in V \cap D} Q_-(\xi) = -\Omega_k^-,$$

$$\tilde{\Omega}_k^- \equiv \sup_{V \in S_k} \inf_{\xi \in V \cap \tilde{D}} \tilde{Q}_-(\xi) = -\inf_{V \in S_k} \sup_{V \cap D} Q_+(\xi) = -\Omega_k^+.$$

Finally, we observe that $2\omega A + B$ is uniformly bounded below on Δ for $\omega \in (\Omega_k^- - \epsilon, \Omega_k^-)$ if and only if $2\omega A + \tilde{B}$ is uniformly bounded above on Δ for $\omega \in (\tilde{\Omega}_k^+, \tilde{\Omega}_k^+ + \epsilon)$, so that $\tilde{J}_+ = J_-$. Now $\Omega_l^- \leq \Gamma_+$ implies $\tilde{\Omega}_l^+ \geq \Gamma_-$, and Theorem 2 gives $\tilde{\Lambda}_k(\tilde{\Omega}_k^+) = 0$ for all $k \geq l$, $k \in \tilde{J}_+$. Since $\tilde{\Lambda}_k(\tilde{\Omega}_k^+) = \Lambda_k(-\tilde{\Omega}_k^+) = \Lambda_k(\Omega_k^-)$, the theorem is proved.

III. SUFFICIENT CONDITIONS FOR THE EXISTENCE OF EIGENVALUES

The minimax principles for certain of the zeros of the functions $\Lambda_k(\omega)$ being established, we next consider the problem of whether a zero Ω of $\Lambda_k(\omega)$ is indeed an eigenvalue of the system $C_\omega \xi = 0$, i.e., whether C_Ω has a nontrivial nullspace. Two sufficient conditions for an affirmative answer are given in this section. The first (Theorem 4) requires that at the given zero Ω the operator C_Ω is decomposable into K-P, where K and P are Hermitian, K is compact, and $\inf_E[(\xi, P\xi)/(\xi, \xi)] > 0$. The second (Theorem 5) requires that $C_\Omega = T - L$, where T is Hermitian and L admits a positive, compact, Hermitian inverse. The latter result is particularly useful in applications to physical systems.

Theorem 4: Let A, B, and C be bounded linear Hermitian operators from E into E, and let $\Lambda_k(\Omega) = 0$ (with $\Delta = E$) for some real Ω and all $k \in U$, where U is some subset of the positive integers. Suppose further that $C_\Omega = K - P$ where P is a positive Hermitian operator and K is a compact Hermitian operator, with $\inf_E[(\xi, P\xi)/(\xi, \xi)] > 0$. Then U is a finite set, and for each $k \in U$ there exists a nonzero $\xi_k \in E$ such that $C_\Omega \xi_k = 0$, where $(\xi_k, \xi_l) = 0$ if $k \neq l$, $k, l \in U$.

Proof: Let $k \in U$. Then

$$0 = \Lambda_k(\Omega) = \sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, C_\Omega \xi)}{(\xi, \xi)}$$

$$= \sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, [K-P]\xi)}{(\xi, \xi)}$$

$$= \sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, P\xi)}{(\xi, \xi)} \left(\frac{(\xi, K\xi)}{(\xi, P\xi)} - 1 \right),$$

which implies

$$\sup_{V \in S_k} \inf_{\xi \in V} \frac{(\xi, K\xi)}{(\xi, P\xi)} = 1, \quad k \in U. \tag{11}$$

Now P admits the positive bounded Hermitian square root $P^{1/2}$ with the positive bounded Hermitian inverse $P^{-1/2}$. We set $\zeta = P^{1/2}\xi$ and obtain from Eq. (11)

$$\sup_{V \in S_k} \inf_{\zeta \in V} \frac{(\zeta, \tilde{K}\zeta)}{(\zeta, \zeta)} = 1, \quad k \in U, \tag{12}$$

where $\tilde{K} \equiv P^{-1/2} K P^{-1/2}$ is compact and Hermitian. Clearly (12) implies that $\|\tilde{K}\| > 0$. It follows from well-known theorems on compact Hermitian operators that $\tilde{K}$ admits of a nonempty set of real nonzero eigenvalues $\{\mu_i^+\} \cup \{\mu_j^-\}$ with corresponding orthonormal eigenvectors $\{\phi_i^+\} \cup \{\phi_j^-\}$ possessing the following properties[22]:

$$\mu_1^+ \geq \mu_2^+ \geq \cdots > 0;\ \mu_1^- \leq \mu_2^- \leq \cdots < 0$$

[each of the sets $\{\mu_i^+\}$ and $\{\mu_j^-\}$ may be finite, infinite, or (but not both) empty]; $\lim_{i \to \infty} \mu_i^+ = 0$ (if $\{\mu_i^+\}$ is infinite); $\lim_{j \to \infty} \mu_j^- = 0$ (if $\{\mu_j^-\}$ is infinite); and for all $\zeta \in E$,

$$(\zeta, \tilde{K}\zeta) = \sum_i \mu_i^+ |(\zeta, \phi_i^+)|^2 + \sum_j \mu_j^- |(\zeta, \phi_j^-)|^2. \tag{13}$$

For $k \in U$, $\lambda_k = 1$, and we have $\lambda_l \geq \lambda_k = 1$ for all positive integers $l \leq k$, where $\lambda_l \equiv \sup_{V \in S_l} \inf_{\xi \in V}[(\xi, \tilde{K}\xi)/(\xi, \xi)]$, $l = 1, 2, 3, \cdots$. This implies that $\mu_1^+, \ldots, \mu_k^+$ all exist. Indeed, suppose that $\{\mu_i^+\}$ consists of precisely $n < k$ positive eigenvalues. Then given any $V \in S_k$, there exists a nonzero $\zeta \in V$ such that $(\zeta, \phi_i^+) = 0$ for all the n ϕ_i^+. Therefore, Eq. (13) implies $(\zeta, \tilde{K}\zeta) \leq 0$, so that $\inf_{\xi \in V}[(\xi, \tilde{K}\xi)/(\xi, \xi)] \leq 0$. Since V was arbitrary, we conclude that $\lambda_k = \sup_{V \in S_k} \inf_{\xi \in V}[(\xi, \tilde{K}\xi)/(\xi, \xi)] \leq 0$, a

contradiction. Hence $\mu_1^+, \ldots, \mu_k^+$ exist. Let $V_0 \equiv \text{span}\{\phi_1^+, \ldots, \phi_k^+\}$. Then for $\zeta \in V_0$, $\|\zeta\|^2 = \sum_1^k |(\zeta, \phi_i^+)|^2$ and it follows from Eq. (13) that

$$\inf_{\zeta \in V_0} \frac{(\zeta, \tilde{K}\zeta)}{(\zeta, \zeta)} = \inf_{\zeta \in V_0} \frac{\sum_1^k \mu_i^+ |(\zeta, \phi_i^+)|^2}{\sum_1^k |(\zeta, \phi_i^+)|^2} = \mu_k^+.$$

Hence $1 = \lambda_k \geq \mu_k^+$. On the other hand, let $V \in S_k$. Then there exists a nonzero $\zeta \in V$ such that $(\zeta, \phi_i^+) = 0$ for $i = 1, 2, \ldots, k-1$. Equation (13) yields

$$\frac{(\zeta, \tilde{K}\zeta)}{(\zeta, \zeta)} = \frac{\sum_{i \geq k} \mu_i^+ |(\zeta, \phi_i^+)|^2 + \sum_j \mu_j^- |(\zeta, \phi_j^-)|^2}{\|\zeta\|^2}$$

$$\leq \frac{\sum_{i \geq k} \mu_i^+ |(\zeta, \phi_i^+)|^2}{\|\zeta\|^2}$$

$$\leq \mu_k^+ \frac{\sum_{i \geq k} |(\zeta, \phi_i^+)|^2}{\|\zeta\|^2} \leq \mu_k^+,$$

where the last estimate follows from Bessel's inequality Therefore, $\inf_{\zeta \in V}[(\xi, \tilde{K}\xi)/(\xi, \xi)] \leq \mu_k^+$, and since V was an arbitrary element of S_k, it follows that $\lambda_k \leq \mu_k^+$. Thus $k \in U$ implies $\mu_k^+ = \lambda_k = 1$, $\tilde{K}\phi_k^+ = \phi_k^+$, and therefore $(K-P)\zeta_k = 0$, where $\zeta_k \equiv P^{-1/2}\phi_k^+ \neq 0$. Since $\mu_k^+ = 1$ for $k \in U$, U must be a finite set ($\mu_i^+ = 1$ for infinitely many i contradicts $\lim_{i \to \infty} \mu_i^+ = 0$). Let m denote the number of elements in U. The set of m vectors $T \equiv \{\zeta_k | k \in U\}$ is obviously linearly independent and is a basis for $S \equiv \text{span}\, T$, so that $m = \dim S$. Let $\{\psi_j\}_{j=1}^m$ be an orthonormal basis for S, and for each $k \in U$, set $\xi_k \equiv \psi_j$, where $U = \{k_1, k_2, \ldots, k_m\}$ and j satisfies $k_j = k$. Since the ξ_k are linear combinations of the ζ_k, we have $C_\Omega \xi_k = (K-P)\xi_k = 0$ for all $k \in U$. This completes the proof.

Corollary 1: Let $A = K_1$, $B = K_2 - P_2$, and $C = K_3 - P_3$, where K_1, K_2, and K_3 are compact Hermitian operators from E into E, P_2 and P_3 are bounded Hermitian operators from E into E, and $K_1 > 0$. Let (H2) and (H4) hold, with $\Delta = E$. [Note that if $C \leq 0$, (H2) and (H4) hold with $M = \Omega = 0$, $\Delta_- = E$, and $\Gamma_- \leq 0 \leq \Gamma_+$.] Then the following conclusions hold:

(A) If $\inf_E\{(\xi, [\omega P_2 + P_3]\xi)/(\xi, \xi)\} > 0$ for $\omega > \omega_0$ and if $\Omega_m^+ \geq \Gamma_-$ and $\Omega_m^+ > \omega_0$ for some positive integer m, then for all $k \geq m$ there exists a nonzero $\xi_k^+ \in E$ such that $C_{\Omega_k^+}\xi_k^+ = 0$; furthermore, $(\xi_k^+, \xi_l^+) = 0$ if $\Omega_k^+ = \Omega_l^+$, $k \neq l$.

(B) If $\inf_E\{(\xi, [\omega P_2 + P_3]\xi)/(\xi, \xi)\} > 0$ for $\omega < \omega_0$ and if $\Omega_m^- \leq \Gamma_+$ and $\Omega_m^- < \omega_0$ for some positive integer m, then for all $k \geq m$ there exists a nonzero $\xi_k^- \in E$ such that $C_{\Omega_k^-}\xi_k^- = 0$; furthermore, $(\xi_k^-, \xi_l^-) = 0$ if $\Omega_k^- = \Omega_l^-$, $k \neq l$.

(C) If $\inf_E[(\xi, P_2\xi)/(\xi, \xi)] > 0$, then $\lim_{k \to \infty} \Omega_k^+ = \infty$, and the numbers m and ω_0 required in (A) exist.

(D) If $\sup_E[(\xi, P_2\xi)/(\xi, \xi)] < 0$, then $\lim_{k \to \infty} \Omega_k^- = -\infty$, and the numbers m and ω_0 required in (B) exist.

Proof: (A) By hypothesis, (H1)–(H4) hold with $\Delta = E$. Thus Ω_k^+ is defined by Eq. (7) for all $k \geq M+1$. If $\Omega_m^+ \geq \Gamma_-$, Theorem 2 implies $\Lambda_k(\Omega_k^+) = 0$ for all $k \geq m$. If $\Omega_m^+ > \omega_0$, the stated conclusion then follows from Theorem 4, since $k \geq m$ implies $\Omega_k^+ \geq \Omega_m^+ > \omega_0$ and we have $C_{\Omega_k^+} = K - P$, where $K \equiv (\Omega_k^+)^2 A + \Omega_k^+ K_2 + K_3$ is compact and $P \equiv \Omega_k^+ P_2 + P_3$ satisfies $\inf_E[(\xi, P\xi)/(\xi, \xi)] > 0$ for $k \geq m$. The proof of (B) is similar. To prove (C), let $\alpha \equiv \inf_E[(\xi, P_2\xi)/(\xi, \xi)] > 0$, let N be any given positive number, and let $0 < \epsilon < \alpha/4N$. Since K_2 is compact, there exists a finite-dimensional subspace E_1 of E such that $\sup_{E_1^\perp}[|(\xi, K_2\xi)|/(\xi, \xi)] < \alpha/2$. Since A is compact, there exists a finite-dimensional subspace E_2 of E such that $\sup_{\xi \in E_2^\perp}[(\xi, A\xi)/(\xi, \xi)] < \epsilon$. Let $n_1 \equiv \dim E_1$, $n_2 \equiv \dim E_2$, and set $S \equiv \{\xi \,|\, \xi = \alpha_1\xi_1 + d_2\xi_2 + d_3\xi_3, \xi_1 \in E_1, \xi_2 \in E_2, \xi_3 \in \Delta_+, \alpha_i \in Z, i = 1, 2, 3\}$ [Z is the set of all complex numbers and Δ_+ is defined in (H2)]. Then S is a subspace of E with $\dim S \leq n_1 + n_2 + M$, and $\zeta \in S^\perp$ implies $\zeta \in \Delta_- \cap E_1^\perp \cap E_2^\perp$. Let $k \geq n_1 + n_2 + M + 1$, and $V \in S_k$. Then there exists a nonzero $\zeta \in V$ such that $\zeta \in S^\perp$. Hence $\zeta \in V \cap D$ and

$$Q_+(\zeta) = \frac{-(\zeta, B\zeta) + \sqrt{d(\zeta)}}{2(\zeta, A\zeta)} \geq -\frac{(\zeta, B\zeta)}{2(\zeta, A\zeta)}$$

$$= \frac{(\zeta, P_2\zeta) - (\zeta, K_2\zeta)}{2(\zeta, A\zeta)} \geq \frac{\alpha/2}{2\epsilon} > N,$$

so that $\sup_{\zeta \in V \cap D} Q_+(\xi) > N$. Since this holds for any $V \in S_k$, $\Omega_k^+ \geq N$ for all $k \geq n_1 + n_2 + M + 1$. Finally, we note that ω_0 may be taken to be $\alpha^{-1}\|P_3\|$. Statement (D) follows by applying (C) to $C_{-\omega} = \omega^2 A + \omega(-B) + C$.

We consider next the case where C_Ω contains an operator with a positive, compact, Hermitian inverse.

Lemma 3: Let (H1) hold, let $\Lambda_n(\Omega) = 0$ for some real Ω and positive integer n, and suppose that $C_\Omega = T - L$, where T is a bounded Hermitian operator from E into E and L admits the positive compact Hermitian inverse K such that $KL = I$ on Δ. Then $\sup_{V \in S_n} \inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} = 0$, where $k \equiv K^{1/2}$ and $\tilde{T} \equiv kTk$.

Proof: Let $V \in S_n$. Since $\Lambda_n(\Omega) = 0$, $\inf_{\zeta \in V}[(\xi, C_\Omega\xi)/(\xi, \xi)] \leq 0$, so that given $\epsilon > 0$, there exists $\xi \in V$ with $\|\xi\| = 1$ such that $(\xi, C_\Omega\xi) < \epsilon$. Set $\zeta \equiv kL\xi$. Then $\zeta \neq 0$ and $k\zeta = KL\xi = \xi$. Hence $\epsilon > (\xi, C_\Omega\xi) = (\xi, [T - L]\xi) = (k\zeta, Tk\zeta) - (k\zeta, L\xi) = (\zeta, [\tilde{T} - I]\zeta)$, and $1 = \|\xi\|^2 = \|k\zeta\|^2 = (\zeta, K\zeta) \leq \|K\|(\zeta, \zeta)$, so that $(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta) < \epsilon\|K\|$. Therefore, $\inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} < \epsilon\|K\|$, which implies that $\inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} \leq 0$. Since V was an arbitrary element of S_n, it follows that $\sup_{V \in S_n} \inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} \leq 0$. Given $\epsilon > 0$, there exists $V \in S_n$ such that $\inf_{\zeta \in V}[(\xi, C_\Omega\xi)/(\xi, \xi)] > -\epsilon$, so that $(\xi, C_\Omega\xi)/(\xi, \xi) > -\epsilon$ for all nonzero $\xi \in V$. Let $\zeta \in kL(V)$, $\zeta \neq 0$. Then $\zeta = kL\xi$ for some nonzero $\xi \in V$, $k\zeta = \xi$, and we have $(\zeta, [\tilde{T} - I]\zeta)/(\zeta, K\zeta) = (\xi, C_\Omega\xi)/(\xi, \xi) > -\epsilon$, which implies that $(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta) > -\epsilon\|K\|$. Since this holds for all nonzero $\zeta \in kL(V)$, $\inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} \geq -\epsilon\|K\|$, and therefore $\sup_{V \in S_n} \inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} \geq -\epsilon\|K\|$. This holds for arbitrary $\epsilon > 0$, so that $\sup_{V \in S_n} \inf_{\zeta \in kL(V)}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} \geq 0$, which completes the proof.

Lemma 4: Let the hypothesis of Lemma 3 hold, and suppose that $\overline{kL(\Delta)} = E$. Then $\sup_{V \in \sigma_n} \inf_{\zeta \in V}\{(\zeta, [\tilde{T} - I]\zeta)/(\zeta, \zeta)\} = 0$, where σ_n is the set of all n-dimensional subspaces of E.

Proof: Let $V \in \sigma_n$, $0 < \epsilon < \frac{1}{2}$, $\{\phi_j\}_1^n$ be an orthonormal basis for V, and $\rho = \sup_{x \in E_n} |x|/\|x\|$, where E_n denotes

the complex Euclidean n-dimensional vector space $\{x \mid x=(x_1, x_2, \ldots, x_n)$, x_j complex, $j=1, \ldots, n\}$, $|x| \equiv \Sigma_{j=1}^n |x_j|$, and $\|x\|^2 \equiv \sum_{j=1}^n |x_j|^2$. Since $\overline{kL(\Delta)}=E$, there exists $\psi_j \in kL(\Delta)$ such that $\|\phi_j - \psi_j\| < \epsilon\rho^{-1}$, $j=1, \ldots, n$. The set $\{\psi_j\}_1^n$ is linearly independent. Indeed, suppose $\Sigma_1^n \alpha_j \psi_j = 0$. Then

$$\sum_1^n |\alpha_j|^2 = \left\| \sum_1^n \alpha_j \phi_j \right\|^2 = \left\| \sum_1^n \alpha_j (\phi_j - \psi_j) \right\|^2 \leq \left(\sum_1^n |\alpha_j| \, \|\phi_j - \psi_j\| \right)^2 \leq \epsilon^2 \rho^{-2} \left(\sum_1^n |\alpha_j| \right)^2 \leq \epsilon^2 \sum_1^n |\alpha_j|^2,$$

which implies $\Sigma_1^n |\alpha_j|^2 = 0$. Since $\psi_j \in kL(\Delta)$, there exists a nonzero $\chi_j \in \Delta$ such that $\psi_j = kL\chi_j$, $j=1, \ldots, n$, and the set $\{\chi_j\}_1^n$ is clearly linearly independent. Let $V_\epsilon \equiv \text{span}\{\chi_j\}_1^n$. Then $V_\epsilon \in S_n$. Let $\eta \in kL(V_\epsilon)$, $\eta \neq 0$, and set $\xi = \eta/\|\eta\|$. Then $\xi = \sum_1^n \alpha_j \psi_j$, and we define $\zeta = \Sigma_1^n \alpha_j \phi_j$ and $\gamma^{-1} \equiv \|\zeta\| = (\sum_1^n |\alpha_j|^2)^{1/2}$. Now

$$|1-\gamma^{-1}| = \left| \left\| \sum_1^n \alpha_j \psi_j \right\| - \left\| \sum_1^n \alpha_j \phi_j \right\| \right| \leq \left\| \sum_1^n \alpha_j (\psi_j - \phi_j) \right\| \leq \sum_1^n |\alpha_j| \left\| \psi_j - \phi_j \right\| < \epsilon\rho^{-1} \sum_1^n |\alpha_j| \leq \epsilon\gamma^{-1},$$

so that $|\gamma - 1| < \epsilon < \frac{1}{2}$, and $\gamma^{-1} < 2$. Therefore, we have

$$\|\xi - \gamma\zeta\| = \left\| \sum_1^n \alpha_j (\psi_j - \gamma\phi_j) \right\| \leq \left\| \sum_1^n \alpha_j (\psi_j - \phi_j) \right\| + |1-\gamma| \left\| \sum_1^n \alpha_j \phi_j \right\| \leq \epsilon\rho^{-1} \sum_1^n |\alpha_j| + |1-\gamma|\gamma^{-1} \leq \epsilon\gamma^{-1} + \epsilon\gamma^{-1} < 4\epsilon,$$

so that

$$\left| \frac{(\eta, [\tilde{T}-I]\eta)}{(\eta,\eta)} - \frac{(\zeta, [\tilde{T}-I]\zeta)}{(\zeta,\zeta)} \right| = |(\xi, [\tilde{T}-I]\xi) - (\gamma\zeta, [\tilde{T}-I]\gamma\zeta)| \leq |(\xi, [\tilde{T}-I][\xi-\gamma\zeta])| + |([\xi-\gamma\zeta], [\tilde{T}-I]\gamma\zeta)| \leq 2\|\xi-\gamma\zeta\|\|\tilde{T}-I\| < 8\epsilon\|\tilde{T}-I\|.$$

Thus we have shown that given any $V \in \sigma_n$ and $0 < \epsilon < \frac{1}{2}$, there exists a $V_\epsilon \in S_n$ such that for any nonzero $\eta \in kL(V_\epsilon)$, there exists a nonzero $\zeta \in V$ such that

$$\frac{(\zeta, [\tilde{T}-I]\zeta)}{(\zeta,\zeta)} < \frac{(\eta, [\tilde{T}-I]\eta)}{(\eta,\eta)} + 8\epsilon\|\tilde{T}-I\|.$$

Hence $\inf_{\zeta \in V} \{(\zeta, [\tilde{T}-I]\zeta)/(\zeta,\zeta)\} \leq \inf_{\eta \in kL(V_\epsilon)} \{(\eta, [\tilde{T}-I]\eta)/(\eta,\eta)\} + 8\epsilon\|\tilde{T}-I\|$, and it follows from Lemma 3 that $\inf_{\zeta \in V} \{(\zeta, [\tilde{T}-I]\zeta)/(\zeta,\zeta)\} \leq 8\epsilon\|\tilde{T}-I\|$. This holds for all sufficiently small positive ϵ, so that $\inf_{\zeta \in V} \{(\zeta, [\tilde{T}-I]\zeta)/(\zeta,\zeta)\} \leq 0$, which implies that $\sup_{V \in \sigma_n} \inf_{\zeta \in V} \{(\zeta, [\tilde{T}-I]\zeta)/(\zeta,\zeta)\} \leq 0$. We now establish the opposite inequality. For any $V \in S_n$, $kL(V) \in \sigma_n$, so that

$$\sup_{V \in \sigma_n} \inf_{\zeta \in V} \frac{(\zeta, [\tilde{T}-I]\zeta)}{(\zeta,\zeta)} \geq \sup_{V \in S_n} \inf_{\zeta \in kL(V)} \frac{(\xi, [\tilde{T}-I]\xi)}{(\xi,\xi)} = 0,$$

and the lemma follows.

Theorem 5: Let (H1) hold and suppose that $\Lambda_n(\Omega) = 0$ for some real Ω and all $n \in U$, where U is some subset of the positive integers. Let $C_\Omega = T - L$, where T is a bounded Hermitian operator from E into E, L admits the positive compact Hermitian inverse K, and $KL = I$ on Δ, $TKT(E) \subset L(\Delta)$, and $\overline{L(\Delta)} = E$. Then U is a finite set, and for each $n \in U$ there exists a nonzero $\zeta_n \in \Delta$ such that $C_\Omega \zeta_n = 0$, where $(\zeta_n, \zeta_m) = 0$ if $m \neq n$, $m, n \in U$.

Proof: Let k denote the positive, compact, Hermitian square root of K. Suppose that $0 = (\phi, kL\zeta)$ for some $\phi \in E$ and all $\zeta \in \Delta$. Then $0 = (k\phi, L\zeta)$ for all $\zeta \in \Delta$, and since $\overline{L(\Delta)} = E$, $k\phi = 0$, so that $\phi = 0$. Therefore, $\overline{kL(\Delta)}^\perp = \{0\}$, and $\overline{kL(\Delta)} = E$. By Lemma 4, $\sup_{V \in \sigma_n} \inf_{\zeta \in V} \{(\zeta, [\tilde{T}-I]\zeta)/(\zeta,\zeta)\} = 0$ for all $n \in U$, and since $\tilde{T} = kTk$ is compact and Hermitian, it follows from Theorem 4 that U must be a finite set, that there exists for each $n \in U$ a nonzero $\xi_n \in E$ such that $(\tilde{T}-I)\xi_n = 0$, and that $\{\xi_n \mid n \in U\}$ is linearly independent. Let $S = \{k\xi_n \mid n \in U\}$. Since $k > 0$, S is linearly independent. Let $\{\zeta_n \mid n \in U\}$ be an orthonormal basis for the span of S. Then $(KT-I)\zeta_n = 0$ for all $n \in U$. Indeed, $(KT-I)\zeta_n = (KT-I)\sum_U \alpha_j k\xi_j = k\sum_U \alpha_j (\tilde{T}-I)\xi_j = 0$. It remains to show that $C_\Omega \zeta_n = 0$, $n \in U$. We have $\zeta_n = KT\zeta_n$, $T\zeta_n = TKT\zeta_n$, and since $TKT(E) \subset L(\Delta)$, there exists $\xi_n \in \Delta$ such that $T\zeta_n = L\xi_n$. Therefore, $\zeta_n = KT\zeta_n = KL\xi_n = \xi_n$, and so $T\zeta_n = L\zeta_n$, i.e., $C_\Omega \zeta_n = 0$.

IV. APPLICATIONS

The results presented in the last two sections have application to a large number of physical systems, as we mentioned in the introduction. Two such examples are discussed herein. The first is the problem of small oscillations about equilibrium of a vertically stratified, viscous, heterogeneous incompressible fluid in a gravitational field. The second concerns the oscillations of a rotating thin annular disk. We show that infinitely many of the eigenfrequencies of these systems are characterized by the minimax principles given in Eqs. (7) and/or (8). We make no attempt here, however, to determine precisely all the eigenfrequencies so described.

We shall have occasion to make use of certain well-known results from the theory of linear differential equations, which for convenience are summarized below.

Proposition: Let L be the nth-order linear differential operator given by

$$L\xi(x) = \sum_{j=0}^n p_j(x) \xi^{(n-j)}(x) \qquad \left(\xi^{(j)} \equiv \frac{d^j \xi}{dx^j} \right),$$

where the $p_j(x)$ are complex-valued functions of class C^{n-j} on the closed interval $a \leq x \leq b$ and $p_0(x) \neq 0$ on $[a, b]$. Let F and G be constant $n \times n$ matrices, $\xi_a \equiv (\xi(a), \xi^{(1)}(a), \ldots, \xi^{(n-1)}(a))^T$, and suppose that L is formally self-adjoint on $S = \{\xi(x) \mid \xi \in C^n[a, b]$, $F\xi_a + G\xi_b = 0\}$. Then the following conclusions hold:

(1) The operator L possesses a denumerably infinite set of eigenfrequencies, all real, with corresponding orthonormal eigenfunctions (in S). The eigenvalues have no finite limit point.

(2) Every $\xi \in S$ can be expanded in a uniformly convergent series of the eigenfunctions.

(3) If 0 is not an eigenvalue of L, then L admits the compact Hermitian inverse K defined on $\mathcal{L}_2[a,b]$ such that $KL=I$ on S, $R_K \subset C[a,b]$, $K(C[a,b])=S$, and $LK=I$ on $C[a,b]$.

For a proof of these results, the reader is referred to Ref. 23.

A. The stratified heterogeneous incompressible fluid

We consider small disturbances about the static equilibrium of a vertically stratified horizontal slab of viscous incompressible fluid, in a gravitational field. The fluid is confined between rigid horizontal walls at $z=0$ and $z=b>0$, where z denotes the vertical coordinate. The pertinent equation and boundary conditions are given in Eqs. (22) and (23) of Ref. 24, viz.:

$$H_\lambda \xi(z) \equiv (\lambda^2 L_1 + \lambda L_2 + L_3)\xi(z) = 0, \quad 0 \leq z \leq b, \tag{14}$$

$$\xi(0) = \xi(b) = \xi'(0) = \xi'(b) = 0. \tag{15}$$

Here $\xi(z)$ denotes the vertical component of the fluid velocity, λ is the eigenfrequency of the disturbance (a time-dependence of the form $e^{\lambda t}$ has been assumed), $L_1 = -d\rho d + k^2\rho$, $L_3 = -k^2 g\rho'$, and $L_2 = d^2\mu d^2 - 2k^2 d\mu d + k^4\mu + k^2\mu''$, where $\mu(z)$ is the fluid viscosity, $\rho(z)$ is the fluid mass density, $g>0$ is the gravitational acceleration, k is the magnitude of the horizontal wave number, $(\,)' \equiv d(\,)/dz$, and d denotes the differential operator d/dz. We assume that $\mu \in C^4[0,b]$, $\rho \in C^2[0,b]$, $\mu>0$ on $[0,b]$, $\rho>0$ on $[0,b]$, $k>0$, and take $E = \mathcal{L}_2[0,b]$ and $\Delta = \{\xi(z) \mid \xi \in C^4[0,b],\ \xi(0)=\xi(b)=\xi'(0) = \xi'(b)=0\}$. For $\xi \in \Delta$, $\xi \neq 0$, we have

$$(\xi, L_1\xi) = \int_0^b \rho(|\xi'|^2 + k^2|\xi|^2)dz > 0, \tag{16}$$

$$(\xi, L_2\xi) = \int_0^b \mu(|\xi'' + k^2\xi|^2 + 4k^2|\xi'|^2)dz > 0. \tag{17}$$

The operators L_1, L_2, and L_3 are all formally self-adjoint on Δ; furthermore, for each real $\lambda \neq 0$, H_λ is a formally self-adjoint fourth-order differential operator on Δ, and the Proposition applies to H_λ with $S=\Delta$,

$$F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad G = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Case 1: Arbitrary ρ'. We set $\omega=\lambda$, $A=L_1$, $B=L_2$, and $C=L_3$. Obviously (H1) and (H3) hold. It is easily seen that for all $\omega<0$, C_ω is bounded above. Indeed, for $\omega<0$, $\xi \in \Delta$, $\hat{\mu} \equiv \min_{[0,b]}\mu$, $\hat{\rho} \equiv \max_{[0,b]}\rho$, $r \equiv \min_{[0,b]}\rho'$, $\alpha \equiv |\omega|\hat{\mu}$, $\beta \equiv 2k^2\alpha - \omega^2\hat{\rho}$, and $\gamma \equiv k^4\alpha - \omega^2 k^2\hat{\rho} + k^2 gr$, Eqs. (16) and (17) give

$$\begin{aligned}(\xi, -C_\omega\xi) &\geq \int_0^b [\alpha(|\xi'' + k^2\xi|^2 + 4k^2|\xi'|^2) \\ &\quad - \omega^2\hat{\rho}(|\xi'|^2 + k^2|\xi|^2) + k^2 rg|\xi|^2]dz \\ &= \int_0^b (\alpha|\xi''|^2 + \beta|\xi'|^2 + \gamma|\xi|^2)dz \\ &= \int_0^b \left[\left|\sqrt{\alpha}\,\xi'' - \frac{\beta}{2\sqrt{\alpha}}\,\xi\right|^2 + \left(\gamma - \frac{\beta^2}{4\alpha}\right)|\xi|^2\right]dz \\ &\geq [\gamma - \beta^2(4\alpha)^{-1}]\|\xi\|^2. \end{aligned} \tag{18}$$

Let $\Omega<0$. Then by the Proposition, C_Ω has an infinite set of real eigenvalues $\{\delta_n\}_1^\infty$ with corresponding orthonormal eigenfunctions $\{\phi_n\}_1^\infty \subset \Delta$, and, since C_Ω is bounded above, the $\{\delta_n\}_1^\infty$ cluster only at $-\infty$, so that there are only a finite number M of positive eigenvalues. We may assume, without loss of generality, that the eigenvalues are enumerated in decreasing order, viz.: $\delta_1 \geq \delta_2 \geq \cdots$. Then $\delta_M>0$, $\delta_n \leq 0$ for $n \geq M+1$, so that for $\Delta_+ \equiv \operatorname{span}\{\phi_1, \ldots, \phi_M\}$, Proposition (2) implies that $(\xi, C_\Omega\xi) = \sum_{M+1}^\infty \delta_n |(\phi_n, \xi)|^2 \leq 0$ for any $\xi \in \Delta_- = \Delta_+^\perp \cap \Delta$. Thus (H2) holds. Consider (H4). Since $A>0$ and $B>0$ on Δ, $Q_-(\xi)<0$ for all nonzero $\xi \in D$ and $\Gamma_- \leq 0$. Let $\xi \in D$, $\xi=0$, and set $x \equiv (\xi, B\xi)/2(\xi, A\xi) > 0$, $y \equiv 2(\xi, C\xi)/(\xi, B\xi)$. If $y \leq 0$, $Q_+(\xi) \geq 0$ while if $y>0$,

$$Q_+(\xi) = -x + \sqrt{x(x-y)} \geq -x + \sqrt{(x-y)^2} = -y \tag{19}$$

with

$$-y = \frac{2k^2 g\int_0^b \rho'|\xi|^2 dz}{(\xi, L_2\xi)} \geq 2\hat{\mu}^{-1} gk^{-2}\min\{0, r\}.$$

Hence in any case, $Q_+(\xi) \geq 2\hat{\mu}^{-1}gk^{-2}\min\{0,r\}$, so that (H4) holds. By a process of estimation similar to that used in Eq. (18), it is readily seen that $2\omega A + B$ is bounded below on Δ for all real ω and that J_- contains all positive integers $n \geq M+1$. We now show that $\Omega_n^- \to -\infty$ as $n \to \infty$. Let $N>0$. Then $B-NA$ is bounded below on Δ, and it follows from the Proposition that $B-NA$ admits an infinite set $\{\alpha_n\}_1^\infty$ of real eigenvalues with corresponding orthonormal eigenfunctions $\{\chi_n\}_1^\infty \subset \Delta$, and that the only limit point of the $\{\alpha_n\}_1^\infty$ is $+\infty$. Thus there are only a finite number m of negative eigenvalues, so that, assuming the eigenvalues to be enumerated in ascending order (i.e., $\alpha_1 \leq \alpha_2 \leq \cdots$), $\alpha_n \geq 0$ for all $n \geq m+1$. Let $E_1 \equiv \operatorname{span}\{\chi_1, \ldots, \chi_m\}$ and $E_2 \equiv \{\eta \mid \eta = \xi_1 + \xi_2,\ \xi_1 \in E_1,\ \xi_2 \in \Delta_+\}$, so that E_2 is a finite-dimensional subspace of Δ with $\dim E \leq m+M$. For $n \leq m+M+1$, given any $V \in S_n$, there exists a nonzero $\xi \in V$ such that $\xi \perp E_2$, and so $\xi \perp E_1$, $\xi \perp \Delta_+$, $\xi \in \Delta_- \subset D$, and $\xi = \Sigma_{m+1}^\infty(\chi_j, \xi)\chi_j$, the last result following from Proposition (2). Hence

$$2Q_-(\xi) \leq -\frac{(\xi, B\xi)}{(\xi, A\xi)} = -N - \frac{(\xi, [B-NA]\xi)}{(\xi, A\xi)} = -N$$

$$- \frac{\sum_{m+1}^\infty \alpha_j |(\chi_j, \xi)|^2}{(\xi, A\xi)} \leq -N,$$

so that $\inf_{\eta \in V \cap D} Q_-(\eta) \leq -N/2$. Since V was an arbitrary element of S_n, $\Omega_n^- \leq -N/2$ for $n \geq m+M+1$, i.e., $\lim_{n\to\infty}\Omega_n^- = -\infty$. Thus there exists a positive integer $l \geq M+1$ such that $\Omega_l^- \leq \Gamma_+$, and it follows immediately from Theorem 3 that $\Lambda_n(\Omega_n^-) = 0$ for all $n \geq l$. For each fixed $n \geq M+1$, $\Omega_n^- < 0$ and $C_{\Omega_n^-}$ is bounded above on Δ, so that for some positive number p, $L \equiv pI - C_{\Omega_n^-}$ is positive on Δ. We conclude from Proposition (3) that L has a positive compact Hermitian inverse K on E such that $KL=I$ on Δ, $C[0,b] \supset K(E)$, and $L(\Delta) = C[0,b]$. Since $\overline{C[0,b]} = \mathcal{L}_2[0,b]$, the hypothesis of Theorem 5 is satisfied (set $T=pI$) for each Ω_n^- with $n \geq l$. We have thus shown that the system defined by Eqs. (14) and (15) possesses an infinite set of negative eigenfrequencies $\lambda_n^- = \Omega_n^-$, $n \geq l$, with $\lambda_n^- \to -\infty$ as $n \to \infty$.

There remains the question of whether any of the Ω_k^+ are eigenvalues. Here we run into the problem that J_+ is empty ($2\omega A + B$ is not bounded above for any real ω) so that Theorem 2 is useless. We can circumvent

this difficulty, however, in the two special cases discussed below.

Case 2: $\rho' \leq 0$ on $[0,b]$. Suppose that $\rho' \leq 0$ on $[0,b]$ and that $\{z \mid \rho'(z)=0,\ 0 \leq z \leq b\}$ has Lebesgue measure zero. Then $(\xi, L_3\xi) = k^2 g \int_0^b (-\rho')|\xi|^2\,dz > 0$ for all nonzero $\xi \in \Delta$, so that the equilibrium is exponentially stable. We set $\omega = \lambda^{-1}$, and rewrite Eq. (14) as

$$\tilde{C}_\omega \xi \equiv (\omega^2 \tilde{A} + \omega B + \tilde{C})\xi = 0, \qquad \xi \in \Delta, \tag{20}$$

with $\tilde{A} \equiv L_3$, $B \equiv L_2$, and $\tilde{C} \equiv L_1$. Obviously (H1) and (H3) hold for the triple of operators $(\tilde{A}, B, \tilde{C})$. For $\omega \neq 0$, $\tilde{C}_\omega = \omega^2 C_{\omega^{-1}}$, and therefore (H2) holds with $\tilde{\Omega} = \Omega^{-1}$ and $\tilde{\Delta}_+ = \Delta_+$, where Ω and Δ_+ are as given in Case 1. Furthermore, Eq. (18) implies that $\tilde{C}_\omega$ is bounded above on Δ for $\omega < 0$. Consider (H4). Since $B > 0$ and $\tilde{A} > 0$ on Δ, $\tilde{Q}_-(\xi) < 0$ for all nonzero $\xi \in \tilde{D} = D$, and so $\tilde{\Gamma}_- \leq 0$. Let $\xi \in D$, $\xi \neq 0$, and set $x \equiv (\xi, B\xi)/2(\xi, \tilde{A}\xi) > 0$, $y \equiv 2(\xi, \tilde{C}\xi)/(\xi, B\xi) > 0$. By Eq. (19), $\tilde{Q}_+(\xi) \geq -y = -2(\xi, L_1\xi)/(\xi, L_2\xi) \geq -2\hat{\rho}k^{-2}\hat{\mu}^{-1}$, and therefore (H4) holds. Obviously $2\omega\tilde{A} + B$ is bounded below on Δ for all real ω, and J_- contains all positive integers $\geq M+1$. We show that $\tilde{\Omega}_n^- \to -\infty$ as $n \to \infty$. The fourth-order differential operator B is formally self-adjoint and positive on Δ, and it follows from the Proposition that B possesses an infinite set of positive eigenvalues $\nu_1 \leq \nu_2 \leq \cdots$ with corresponding orthonormal eigenfunctions $\{\psi_n\}_1^\infty \subset \Delta$. The eigenvalues cluster only at $+\infty$. Given $N > 0$, there exists a positive integer q such that $\nu_n > 2\|\tilde{A}\|N$ for all $n > q$. Let $E_3 \equiv \operatorname{span}\{\psi_1, \ldots, \psi_q\}$ and $E_4 \equiv \{\xi \mid \xi = \xi_1 + \xi_2,\ \xi_1 \in E_3,\ \xi_2 \in \Delta_+\}$, so that the dimension of the subspace $E_4 \subset \Delta$ does not exceed $q+M$. For $n \geq q+M+1$, given any $V \in S_n$, there exists a nonzero $\xi \in V$ such that $\xi \perp E_4$, and so $\xi \perp E_3$, $\xi \perp \Delta_+$, $\xi \in \Delta_- \subset D$, and $\xi = \sum_{q+1}^\infty (\psi_n, \xi)\psi_n$, where the last result follows from Proposition (2). Hence

$$\tilde{Q}_-(\xi) \leq -(2\|\tilde{A}\|)^{-1}\frac{(\xi, B\xi)}{(\xi,\xi)}$$

$$= -(2\|\tilde{A}\|)^{-1}\frac{\sum_{q+1}^\infty \nu_n |(\psi_n, \xi)|^2}{\sum_{q+1}^\infty |(\psi_n, \xi)|^2} \leq -\nu_{q+1}(2\|\tilde{A}\|)^{-1} < -N,$$

and therefore $\inf_{\xi \in V \cap D} \tilde{Q}_-(\xi) < -N$. Since V was an arbitrary element of S_n, we conclude that $\tilde{\Omega}_n^- \leq -N$ for $n \geq q+M+1$, i.e., $\tilde{\Omega}_n^- \to -\infty$ as $n \to \infty$. In particular, there exists a positive integer β such that $\tilde{\Omega}_\beta^- \leq \tilde{\Gamma}_+$, and Theorem 3 implies that $\tilde{\Lambda}_n(\tilde{\Omega}_n^-) = 0$ for all $n \geq \beta$. For any given positive integer $n \geq M+1$, $\tilde{\Omega}_n^- < 0$ and $\tilde{C}_{\tilde{\Omega}_n^-}$ is bounded above on Δ. We proceed exactly as in Case 1 to conclude from Theorem 5 that for each $n \geq \beta$, $\lambda_n^+ \equiv (\tilde{\Omega}_n^-)^{-1}$ is an eigenfrequency of the system described by Eqs. (14) and (15). The λ_n^+ are all negative, and $\lambda_n^+ \to 0$ as $n \to \infty$. It is readily verified that $\tilde{Q}_-(\xi) = [Q_+(\xi)]^{-1} < 0$ for every nonzero $\xi \in D$, and therefore that $(\tilde{\Omega}_n^-)^{-1} = \Omega_n^+$ for all $n \geq M+1$. Hence $\lambda_n^+ = \Omega_n^+$, $n \geq \beta$.

Case 3: $\rho' \geq 0$ on $[0,b]$. Suppose that $\rho' \geq 0$ on $[0,b]$ and that $\{z \mid \rho'(z)=0,\ 0 \leq z \leq b\}$ has measure zero. Then $(\xi, L_3\xi) < 0$ for all nonzero $\xi \in \Delta$, so that, referring back to Case 1, $C < 0$ on Δ and the "overdamped" condition $d(\xi) \geq 0$ on Δ is satisfied. With $\Omega = 0$, we have $M = 0$ and $\Delta_- = D = \Delta$, and since $Q_-(\xi) < 0 < Q_+(\xi)$ for all nonzero $\xi \in \Delta$, $\Gamma_- \leq 0 \leq \Gamma_+$. Thus we may take $l = 1$ in Case 1 and conclude that $\lambda_n^- = \Omega_n^- < 0$ is an eigenfrequency of the system described by Eqs. (14) and (15) for all $n \geq 1$.

We now use our minimax principle to construct an infinite set of positive eigenfrequencies. We set $\omega = \lambda^{-1}$ and rewrite Eq. (14) as

$$\hat{C}_\omega \xi \equiv (\omega^2 \hat{A} + \omega \hat{B} + \hat{C})\xi = 0, \tag{21}$$

with $\hat{A} = -L_3$, $\hat{B} = -L_2$, and $\hat{C} = -L_1$. Obviously (H1) and (H3) hold for the triple of operators $(\hat{A}, \hat{B}, \hat{C})$. Since $\hat{C} < 0$ on Δ, the system is overdamped with $\hat{D} = \Delta$, and (H2) holds with $M = 0$ and $\Delta_- = \Delta$. We have $\hat{Q}_-(\xi) < 0 < \hat{\mu}k^2(2g\max_{[0,b]}\rho')^{-1} \leq \hat{Q}_+(\xi)$ for all nonzero $\xi \in \Delta$, and therefore $\hat{\Gamma}_- \leq 0 < \hat{\Gamma}_+$. The operator $2\omega\hat{A} + \hat{B}$ is clearly bounded above for all real ω, $\hat{J}_+$ contains all the positive integers, and it follows from Theorem 2 that $\hat{\Lambda}_n(\hat{\Omega}_n^+) = 0$ for all $n \geq 1$. The operator $\hat{C}_\omega$ is bounded above for $\omega > 0$, and so we may proceed as in Case 1 to conclude from the Proposition and Theorem 5 that $\lambda_n^+ \equiv (\hat{\Omega}_n^+)^{-1}$ is an eigenfrequency of Eqs. (14) and (15) for each $n \leq 1$. It follows easily from the definition of $\hat{\Omega}_n^+$ and the fact that the eigenvalues $\{\nu_n\}_1^\infty$ of the positive operator $B = -\hat{B}$ cluster at $+\infty$ (see Case 2) that $\hat{\Omega}_n^+ \to \infty$ as $n \to \infty$; thus $\lambda_n^+ \to 0$ as $n \to \infty$. Finally, one readily verifies that $\hat{Q}_+(\xi) = [Q_+(\xi)]^{-1} > 0$ for all nonzero $\xi \in \Delta$, so that the λ_n^+ are given directly in terms of $Q_+(\xi)$ by the max–min principle

$$\lambda_n^+ = \sup_{V \in S_n} \inf_{\xi \in V} Q_+(\xi). \tag{22}$$

Note that $\lambda_n^+ \neq \Omega_n^+$.

This special case of $\rho' \geq 0$ on $[0,b]$ has been investigated by Turner[18] and Eisenfeld.[25] Turner obtains minimax principles for the eigenvalues—however, they are not given directly in terms of L_1, L_2, and L_3. Eisenfeld shows that the eigenfunctions are complete. Their results depend critically on the requirement that $L_3 < 0$ on Δ.

B. The rotating annular disk

We consider the planar oscillations of a thin annular disk rotating about its center with a given angular velocity Ω, and restrict our attention to rotationally symmetric modes. The pertinent equation is Eq. (13) of Ref. 9, viz.:

$$(\omega^2 A + \omega B + C)\xi = 0, \qquad \xi \in \Delta,$$

where $A = I$, $C = I - L$, $\Delta = \Delta_1 \times \Delta_2$,

$$B = 2\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \qquad L = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix},$$

$$L_1 = \gamma_1\left(-\frac{d^2}{dx^2} + \tfrac{3}{4}x^{-2}\right), \qquad L_2 = \gamma_2\left(-\frac{d^2}{dx^2} + \tfrac{3}{4}x^{-2}\right),$$

$$\Delta_1 = \{f(x) \mid f \in C^2[a,1],\ f(a) = 0 = f'(1) - \tfrac{1}{2}(4\gamma_2\gamma_1^{-1} - 1)f(1)\},$$

$$\Delta_2 = \{f(x) \mid f \in C^2[a,1],\ f(a) = 0 = f'(1) - \tfrac{3}{2}f(1)\},$$

$0 < a < 1$, and γ_1 and γ_2 are positive constants satisfying the inequality $4\gamma_2 > \gamma_1 > 2\gamma_2$. The appropriate Hilbert space is $E = \mathcal{L}_2[a,1] \times \mathcal{L}_2[a,1]$. A time dependence of the form $\exp(i\omega\Omega t)$ has been assumed.

It is readily verified that L_1 and L_2 are formally self-adjoint and positive on Δ_1 and Δ_2, respectively. We infer from the Proposition that L_1 and L_2 admit the compact Hermitian inverses K_1 and K_2 defined on

$\mathcal{L}_2[a,1]$ such that $K_nL_n=I$ on Δ_n, $L_n(\Delta_n)=C[a,1]$, and $K_n(\mathcal{L}_2[a,1])\subset C[a,1]$, $n=1,2$. Thus L is formally self-adjoint and positive on Δ, L possesses the compact Hermitian inverse $K=\left(\begin{smallmatrix}K_1 & 0\\ 0 & K_2\end{smallmatrix}\right)$ defined on E such that $KL=I$ on Δ, and $L(\Delta)=C[a,1]\times C[a,1]$. Since $\overline{\Delta}_1=\overline{\Delta}_2=C[a,1]=\mathcal{L}_2[a,1]$, we have $\overline{\Delta}=\overline{L(\Delta)}=E$. For each real ω, $T_\omega KT_\omega(E)\subset C[a,1]\times C[a,1]=L(\Delta)$, where $T_\omega\equiv\omega^2A+\omega B+I=\left(\begin{smallmatrix}\omega^2+1 & i\omega\\ -2i\omega & \omega^2+1\end{smallmatrix}\right)$.

We now show that the minimax principles of Eqs. (7) and (8) each generate infinitely many eigenfrequencies of the disk. The pertinent results of the preceeding sections are collected together in the following theorem, which is convenient and appropriate for application to numerous problems relating to the oscillations of rotating elastic systems.

Theorem 6: Let A, B, and H be bounded Hermitian operators from E into E, $\delta\equiv\inf_E[(\xi,A\xi)/(\xi,\xi)]>0$, let Δ be an infinite-dimensional subspace of E, L be a formally self-adjoint positive operator from Δ into E with compact inverse K on E such that $KL=I$ on Δ, $C=H-L$, $\overline{\Delta}=\overline{L(\Delta)}=E$, $K(E)\subset L(\Delta)$, and $T_\omega KT_\omega\subset L(\Delta)$ for all real ω, where $T_\omega\equiv\omega^2A+\omega B+H$. Then the sets $N_+\equiv\{n\,|\,\Omega_n^+\geq\Gamma_-\}$ and $N_-\equiv\{n\,|\,\Omega_n^-\leq\Gamma_+\}$ are both nonempty, Ω_m^+ and Ω_n^- are eigenvalues of $C_\omega\xi=0$ for all $m\geq\inf N_+$ and $n\geq\inf N_-$, and $\Omega_n^+\to\infty$ and $\Omega_n^-\to-\infty$ as $n\to\infty$.

Proof. By hypothesis, (H1) and (H3) hold. Since L is formally self-adjoint on Δ and $\overline{L(\Delta)}=E$, K is Hermitian, and $L>0$ on Δ and $\overline{\Delta}=E$ imply that $K>0$. It follows from well-known theorems on compact Hermitian operators[22] that K admits an infinite set of positive eigenvalues $\mu_1\geq\mu_2\geq\cdots>0$ with corresponding orthonormal eigenvectors $\{\phi_n\}_1^\infty$ that span$K(E)$, and $\lim_{n\to\infty}\mu_n=0$. We have $\mu_n\phi_n=K\phi_n=L\psi_n$ for some $\psi_n\in\Delta$ ($K(E)\subset L(\Delta)$), so that $\mu_n^2\phi_n=KL\psi_n=\psi_n$, and therefore $L\phi_n=\lambda_n\phi_n$, $\lambda_n\equiv(\mu_n)^{-1}$, $n=1,2,3,\cdots$. Thus every ϕ_n is an eigenvector of L with eigenvalue λ_n, and $\{\phi_n\}_1^\infty$ spans Δ, since $KL=I$ on Δ implies $K(E)\supset\Delta$. Obviously $\lambda_n\to\infty$ as $n\to\infty$ and therefore $\lambda_n>\|H\|$ for all sufficiently large n. Let M be the least integer such that $\lambda_{M+1}\geq\|H\|$. Let $\Delta_+\equiv\text{span}\{\phi_1,\ldots,\phi_M\}$. Then for $\xi\in\Delta_-\equiv\Delta_+^\perp\cap\Delta$, $\xi=\sum_{M+1}^\infty(\phi_n,\xi)\phi_n$, and

$$(\xi,L\xi)=\sum_{M+1}^{\infty}(\xi,\phi_n)(\phi_n,L\xi)=\sum_{M+1}^{\infty}\lambda_n|(\phi_n,\xi)|^2\geq\lambda_{M+1}\|\xi\|^2,$$

so that

$$(\xi,C\xi)=(\xi,H\xi)-(\xi,L\xi)\leq(\|H\|-\lambda_{M+1})\|\xi\|^2\leq 0.$$

Hence (H2) holds with $\Omega=0$ and the above choice of Δ_+ and M. Consider (H4). We have, for all nonzero $\xi\in D$,

$$Q_+(\xi)=\frac{-(\xi,B\xi)+[d(\xi)]^{1/2}}{2(\xi,A\xi)}\geq-\frac{1}{2}\frac{(\xi,B\xi)}{(\xi,A\xi)}\geq-\frac{1}{2}\frac{\|B\|}{\delta},$$

$$Q_-(\xi)=-\frac{(\xi,B\xi)+[d(\xi)]^{1/2}}{2(\xi,A\xi)}\leq-\frac{1}{2}\frac{(\xi,B\xi)}{(\xi,A\xi)}\leq\frac{1}{2}\frac{\|B\|}{\delta},$$

so that $\Gamma_+\geq-\frac{1}{2}\|B\|/\delta$, $\Gamma_-\leq\frac{1}{2}\|B\|/\delta$, and (H4) holds. Since $\|2\omega A+B\|\leq 2|\omega|\,\|A\|+\|B\|$ for all real ω, J_+ and J_- contain all positive integers $n\geq M+1$. We show that $\Omega_n^+\to\infty$ and $\Omega_n^-\to-\infty$ as $n\to\infty$. Given $N>0$, there exists a positive integer $m\geq M$ such that $\lambda_{m+1}>\|H\|+\|A\|(N+\|B\|/2\delta)^2$. Let $V_m\equiv\text{span}\{\phi_1,\ldots,\phi_m\}\supset\Delta_+$, and let $n>m$. Then given any $V\in S_n$, there exists a nonzero $\xi\in V$ such that $\xi\perp V_m$, and so $\xi=\sum_{m+1}^\infty(\phi_k,\xi)\phi_k$, $\|\xi\|^2=\sum_{m+1}^\infty|(\phi_k,\xi)|^2$,

$$\begin{aligned}-(\xi,C\xi)&=(\xi,L\xi)-(\xi,H\xi)\\&\geq\sum_{m+1}^{\infty}\lambda_k|(\phi_k,\xi)|^2\\&\quad-\|H\|\,\|\xi\|^2\geq(\lambda_{m+1}-\|H\|)\,\|\xi\|^2>\|A\|(N\\&\quad+\|B\|/2\delta)^2\|\xi\|^2.\end{aligned}$$

Now $\xi\in V$ and $\xi\perp V_m$ implies $\xi\in\Delta_-\subset D$, and we have

$$Q_+(\xi)\geq-\frac{\|B\|}{2\delta}+\|A\|^{-1/2}\left(\frac{-(\xi,C\xi)}{(\xi,\xi)}\right)^{1/2}>N,$$

$$Q_-(\xi)\leq\frac{\|B\|}{2\delta}-\|A\|^{-1/2}\left(\frac{-(\xi,C\xi)}{(\xi,\xi)}\right)^{1/2}<-N,$$

so that

$$\sup_{\eta\in V\cap D}Q_+(\eta)\geq Q_+(\xi)>N,\qquad\inf_{\eta\in V\cap D}Q_-(\eta)\leq Q_-(\xi)<-N.$$

Since V was an arbitrary element of S_n, we conclude that $\Omega_n^+\geq N$ and $\Omega_n^-\leq-N$ for all $n>m$, i.e., $\Omega_n^+\to\infty$ and $\Omega_n^-\to-\infty$ as $n\to\infty$. In particular, the sets $N_+=\{n\,|\,\Omega_n^+\geq\Gamma_-\}$ and $N_-=\{n\,|\,\Omega_n^-\leq\Gamma_+\}$ are nonempty, and since $J_+\supset N_+$ and $J_-\supset N_-$, we conclude from Theorems 2 and 3 that $\Lambda_k(\Omega_k^+)=0$ for $k\geq\inf N_+$ and $\Lambda_k(\Omega_k^-)=0$ for $k\geq\inf N_-$. The remainder of the theorem now follows immediately from Theorem 5.

*Work supported in part by National Science Foundation Grant NSF-GP33368X.

[1]E.T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., New York, 1937), 4th ed., Chap. 7.
[2]E.T. Routh, *The Advanced Part of a Treatise on the Dynamics of a System of Rigid Bodies* (Macmillan, New York, 1905), 6th ed., Chaps. 3, 6, 7.
[3]W. Thompson (Lord Kelvin) and P.G. Tait, *Treatise on Natural Philosophy*, Part I (Cambridge U.P., 1903).
[4]F.E. Low, Phys. Fluids **4**, 842 (1961).
[5]E.A. Frieman and A. Rotenberg, Rev. Mod. Phys. **22**, 898 (1960).
[6]E.P. Velikhov, Sov. Phys. JETP **9**, 848 (1959).
[7]S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Oxford U.P., Oxford, 1961), Chaps, 7, 9—11.
[8]E.M. Barston, J. Math. Phys. **8**, 523 (1967).
[9]E.M. Barston, J. Math. Phys. **12**, 1867 (1971).
[10]E.M. Barston, J. Fluid Mech. **42**, 97 (1970).
[11]E.M. Barston, "Stability of Dissipative Systems with Applications to Fluids and Magnetofluids" in *Lecture Notes in Physics* (Springer-Verlag, New York, 1973), Vol. 21.
[12]J. Eisenfeld, "Operator Equations and Nonlinear Eigenparameter Problems," J. Funct. Anal. **12**, 475 (1973).
[13]E.M. Barston, J. Math. Phys. **13**, 720 (1972).
[14]G. Laval, R. Pellat, M. Cotsaftis, and M. Trocheris, Nucl. Fusion **4**, 25 (1964).
[15]R.J. Duffin, J. Rat. Mech. Anal. **4**, 221 (1955).
[16]R.J. Duffin, Q. Appl. Math. **18**, 215 (1960).
[17]E.H. Rogers, Arch. Rat. Mech. Anal. **16**, 89 (1964).
[18]R.E.L. Turner, J. Math. Anal. Appl. **17**, 151 (1967).
[19]J. Eisenfeld, J. Funct. Anal. **9**, 777 (1972).
[20]E.M. Barston, Int. J. Engr. Sci. **10**, 947 (1972).
[21]E.M. Barston, Int. J. Engr. Sci. (to be published).
[22]F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1955), Chap. 6.
[23]E.A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955), Chap. 7.
[24]S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Oxford U.P., Oxford, 1961), p. 430.
[25]J. Eisenfeld, J. Math. Anal. Appl. **23**, 58 (1968).

Concerning conservation laws resulting from geometric invariance groups for field theories

Benjamin F. Plybon

Department of Mathematics, Miami University, Oxford, Ohio 45056
(Received 26 December 1973)

A geometric symmetry group is defined as a point transformation of a Riemannian manifold combined with a transformation law for the field as a geometrical object. A covariant definition of invariance of an action integral is given. It is shown that geometric invariance groups can be determined from knowledge of a tensor S^{α}_{β} which can be computed from the density function L in the action integral. A general form for conservation laws due to geometric symmetry is given. Results are applied to electromagnetic fields and it is shown that the Bessel–Hagen conservation laws represent all of the possible conservation laws for electromagnetic fields arising from geometric symmetry.

1. INTRODUCTION

It is well known that if field equations can be derived from a variational principle, then symmetries of the action integral correspond to conservation laws for the physical system. This correspondence between symmetry groups and conservation laws is stated mathematically in the Noether theorems.[1] In the early work by Noether and Bessel-Hagen[2] the concept of a symmetry group was very broad. In fact Bessel-Hagen viewed a symmetry group as simply a change of variables in the action integral which left the integral invariant. In recent years authors concerned with applications of the Noether theorems to physical theories seem to have adopted a more restrictive view of symmetry transformations. Given an integral

$$W=\int dx\, g^{1/2} L(\phi_A, \partial_\alpha \phi_A)$$

where ϕ_A is a geometrical object representing the field, a coordinate transformation in infinitesimal form

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha \tag{1}$$

is introduced and object ϕ_A is transformed according to its mathematical nature as a geometrical object. From the fact that L is assumed to be a scalar, invariance under coordinate transformations from the Lorentz Group results.

So far as the methematical statement of the Noether theorems is concerned, it is possible to transform coordinates x^α and field varibles ϕ_A independently. If in fact transformation (1) represents a coordinate transformation, it seems logical to try transforming ϕ_A as a tensor or vector, perhaps, but it is not necessary to do this if symmetries can be found by transforming ϕ_A in some other way. Hence, in this paper the combined transformation

$$\begin{aligned} &\bar{x}^\alpha = x^\alpha + \Delta x^\alpha, \\ &\phi_A(x) \to \bar{\phi}_A(\bar{x}), \end{aligned} \tag{2}$$

where $\bar{\phi}_A(\bar{x})$ is determined by the transformation properties of a ϕ_A as a geometrical object, will be called a *geometric transformation* of an action integral to indicate its special nature. However, the transformation

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha$$

will not be treated as a coordinate transformation. This may sound contradictory, but it is not. As just pointed out the transformations of x^α and ϕ_A are independent so far as application of the Noether theorems is concerned. Hence, the transformation of ϕ_A selected above was singled out simply because it is the usual choice in the literature and not because of any assumptions about the nature of the transformation of variables x^α. Any transformation such as (1) can be interpreted either as a coordinate transformation or a point transformation.[3] In this paper Eq. (1) will be interpreted as a point transformation for a good reason. In the calculations to follow we assume only a Riemannian manifold as the underlying space. Then to achieve covariant results in our calculations, it becomes apparent that Δx^α in (1) must be a vector. That is the case if (1) is a point transformation, but it does not make sense to treat Δx^α as a vector when (1) is a coordinate transformation (many authors do so, however). Hence ϕ_A in (2) is not being transformed to a new coordinate system, but is being dragged along as we move from point x to point $\bar{x}$ in the manifold.

Specification of our viewpoint toward (1) is crucial when we decide how to handle metric tensor $g_{\alpha\beta}$ in the following calculations. We shall agree not to introduce any local variation of the metric and operate at all times within a fixed coordinate system. Hence, given point transformation $\bar{x}^\alpha = x^\alpha + \Delta x^\alpha$, $g_{\mu\nu}(x)$ transforms to $g_{\mu\nu}(\bar{x})$.

Under the above conditions it is possible to give a very general analysis of invariance of action integrals and a description of all possible conservation laws resulting from geometric symmetry. It must be kept in mind, however, that other kinds of symmetry groups are conceivable. In fact, the arguments in this paper prove that other symmetry groups must exist for electromagnetic fields and perhaps others, because conservation laws are known to exist which do not fit the pattern produced in this paper for conservation laws arising from geometric symmetry.

2. GEOMETRIC SYMMETRY GROUPS FOR INTEGRALS

We shall assume real numbers (x^1, x^2, x^3, x^4) are coordinates of points in a Riemannian manifold with metric tensor $g_{\alpha\beta}$. The coordinate system will be kept fixed and the metric tensor will not be given any local variation. To illustrate the ideas we shall study an integral

$$W=\int_\Omega dx\, g^{1/2} L(\phi_\alpha, \nabla_\beta \phi_\alpha), \tag{3}$$

but similar calculations can be made for other kinds of

geometrical objects ϕ_A. A geometric transformation group is specified by

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha = \bar{x}^\alpha + \xi^\alpha_k \epsilon k$$
$$\bar{\phi}_\alpha(\bar{x}) = \phi_\beta(x) \frac{\partial x^\beta}{\partial \bar{x}^\alpha} \tag{4}$$

where generators ξ^χ_k are functions of the x^α and $\epsilon^1, \ldots, \epsilon^r$ are independent parameters. We assume tensor $\nabla_\beta \phi_\alpha$ is transformed by rule

$$\overline{\nabla_\alpha \phi_\beta} = \nabla_\mu \phi_\nu \frac{\partial x^\mu}{\partial \bar{x}^\alpha} \frac{\partial x^\nu}{\partial \bar{x}^\beta}.$$

It is convenient to let $\nabla_\alpha \phi_\beta = F_{\alpha\beta}$ in the following, so we shall do so. The usual total variations of ϕ_α and $F_{\alpha\beta}$ are defined by

$$\bar{\phi}_\alpha(\bar{x}) = \phi_\alpha(x) + \Delta\phi_\alpha,$$
$$\bar{F}_{\alpha\beta}(\bar{x}) = F_{\alpha\beta}(x) + \Delta F_{\alpha\beta}.$$

Variations $\Delta\phi_\alpha$ and $\Delta F_{\alpha\beta}$ are not vectors or tensors, however. This causes difficulties in attempts to produce convariant results in the following. It was shown by Plybon in a recent paper[4] that covariant forms of the Noether theorems could be produced by redefining total variations as follows. Given $\Delta\phi_\alpha$ as above the local variation of ϕ_α is

$$\delta\phi_\alpha = \bar{\phi}_\alpha(x) - \phi_\alpha(x)$$

and $\delta\phi_\alpha$ can be computed from $\Delta\phi_\alpha$ by $\delta\phi_\alpha = \Delta\phi_\alpha - \partial_\beta \phi_\alpha \Delta x^\beta$. Let

$$(\Delta\phi)_\alpha = \delta\phi_\alpha + \nabla_\lambda \phi_\alpha \Delta x^\lambda. \tag{5}$$

Similarly, let

$$(\Delta F)_{\alpha\beta} = \delta F_{\alpha\beta} + \nabla_\lambda F_{\alpha\beta} \Delta x^\lambda. \tag{6}$$

For geometric transformation (4) we find

$$(\Delta\phi)_\alpha = -\phi_\mu \nabla_\alpha \Delta x^\mu \tag{7}$$

and

$$(\Delta F)_{\alpha\beta} = -F_{\alpha\nu} \nabla_\beta \Delta x^\nu - F_{\nu\beta} \nabla_\alpha \Delta x^\nu. \tag{8}$$

Now using these vector total variations we define the total variation of integral (3) by

$$\Delta W = \int_{\bar{\Omega}} d\bar{x}\ \bar{g}^{1/2} L[\phi_\alpha + (\Delta\phi)_\alpha, F_{\mu\nu} + (\Delta F)_{\mu\nu}] - \int_\Omega dx\ g^{1/2} L(\phi_\alpha, F_{\mu\nu}). \tag{9}$$

We say geometric transformations (4) form a *geometric symmetry group* if $\Delta W = 0$. This is similar to the usual definition of a symmetry group for an integral, but not quite the same due to use of vector total variations. This can be expected to have some effect on the resulting symmetry group. In flat spaces nothing has been changed but in Riemannian spaces in general we have a new concept of a symmetry transformation. The precise nature of a symmetry transformation in a generally covariant theory does not seem clear in current literature, so some definition is in order. The definition selected here seems necessary to this author in order to produce covariant results. We shall now compute ΔW and show how geometric symmetry groups can be determined. We have

$$\bar{g}^{1/2} = g(\bar{x})^{1/2} = g^{1/2}(1 + \Gamma^\rho_{\rho\alpha} \Delta x^\alpha)$$

and

$$d\bar{x} = dx(1 + \partial_\alpha \Delta x^\alpha).$$

Hence,

$$d\bar{x}\,\bar{g}^{1/2} = dx g^{1/2}(1 + \nabla_\alpha \Delta x^\alpha). \tag{10}$$

Substitution of (10) into (9) and some rearrangement yields

$$\Delta W = \int_\Omega dx g^{1/2} \left(L \nabla_\alpha \Delta x^\alpha + \frac{\partial L}{\partial \phi_\alpha} (\Delta\phi)_\alpha + \frac{\partial L}{\partial F_{\lambda\rho}} (\Delta F)_{\lambda\rho} \right). \tag{11}$$

Now using (7) and (8) we find

$$\Delta W = \int_\Omega dx g^{1/2}\, S^\beta{}_\nu \nabla_\beta \Delta x^\nu \tag{12}$$

where

$$S^\beta{}_\nu = L\delta^\beta_\nu - \frac{\partial L}{\partial \phi_\beta} \phi_\nu - \frac{\partial L}{\partial F_{\alpha\beta}} F_{\alpha\nu} - \frac{\partial L}{\partial F_{\beta\alpha}} F_{\nu\alpha}. \tag{13}$$

Given (3) we can compute $S^\beta{}_\nu$ from (13). Tensor $S^\beta{}_\nu$ determines ΔW for a given point transformation. From the form of $S^\beta{}_\nu$ we can determine which point transformations will lead to geometric symmetries.

3. SOME SPECIAL CASES

Suppose $S^{\beta\nu} = S^{\nu\beta}$. Then

$$S^\beta{}_\nu \nabla_\beta \Delta x^\nu = \tfrac{1}{2} S^{\beta\nu}(\nabla_\beta \Delta x_\nu + \nabla_\nu \Delta x_\beta).$$

Hence

$$\Delta W = 0 \quad \text{if} \quad \nabla_\beta \Delta x_\nu + \nabla_\nu \Delta x_\beta = 0. \tag{14}$$

Equation (14) is Killing's equation and we have found $\bar{x}^\alpha = x^\alpha + \Delta x^\alpha$ produces a geometric symmetry if Δx^α is a Killing vector. In this case our point transformation is a motion for the manifold. Such transformations are well known in the literature and have been studied extensively.[5] In the case of a Minkowski space we have the Lorentz group as the group of motions.

Suppose further $S^{\alpha\beta} = S^{\beta\alpha}$ and $S^\alpha{}_\alpha = 0$. Now we find if

$$\nabla_\beta \Delta x_\nu + \nabla_\nu \Delta x_\beta = \psi(x) g_{\nu\beta}$$

for some scalar $\psi(x)$, then

$$S^\beta{}_\nu \nabla_\beta \Delta x^\nu = \tfrac{1}{2} S^\alpha{}_\alpha\, \psi(x) = 0.$$

If Δx^ν satisfies the above condition then

$$\bar{x}^\alpha = x^\alpha + \Delta x^\alpha$$

is a conformal point transformation. Hence the group of conformal transformations yields a geometric symmetry group if $S^{\alpha\beta} = S^{\beta\alpha}$ and $S^\alpha{}_\alpha = 0$.

The preceding results look familiar since all of the above is well known if $S^\alpha{}_\nu$ is the energy–momentum tensor for a field. However, $S^\alpha{}_\nu$ as defined above has not been shown to be such an energy–momentum tensor. In fact, in covariant theories it is not easy to explain what an energy–momentum tensor should be in general.

The special cases of S^{α}_{β} examined above are probably the most important ones for classical field theories, but it is conceivable that some other special properties of S^{α}_{β} might arise in a given theory leading to other geometric symmetry groups.

A particularly interesting application of the preceding to electromagnetic fields can be made. Let ϕ_{α} be the 4-potential and let

$$f_{\mu\nu}=\nabla_{\mu}\phi_{\nu}-\nabla_{\nu}\phi_{\mu}.$$

Let $L=-\frac{1}{4}f_{\mu\nu}f^{\mu\nu}$.

Then

$$W=\int dxg\tfrac{1}{2}L \tag{15}$$

is an action integral for the electromagnetic field in vacuum. From (13) we find

$$S^{\beta}{}_{\nu}=-\tfrac{1}{4}f_{\lambda\rho}f^{\lambda\rho}\delta^{\beta}{}_{\nu}+f^{\alpha\beta}f_{\alpha\nu} \tag{16}$$

which we recognize as the conventional energy—momentum tensor for the field. Since $S^{\beta\nu}=S^{\nu\beta}$ and $S^{\alpha}{}_{\alpha}=0$ in this case, then the conformal group of point transformations yields a geometric symmetry group for the field by our preceding arguments. This result is well known in flat spaces. The argument given here shows it to be true in any Riemannian manifold. The nature of the conformal group varies from one manifold to another, however, and in fact some manifolds do not admit any conformal point transformations.

4. FORM OF THE CONSERVATION LAWS

When a symmetry group exists this implies existence of conservation laws. It will be shown now that a rather general form can be given for all the conservation laws arising from (3) because of geometric symmetries. It was shown by the author[4] that if any symmetry transformation as defined above exists, then $\nabla_{\alpha}P^{\alpha}=0$ where

$$P^{\alpha}=T^{\alpha}{}_{\lambda}\Delta x^{\lambda}+\frac{\partial L}{\partial F_{\alpha\beta}}(\Delta\phi)_{\beta} \tag{17}$$

and

$$T^{\alpha}{}_{\lambda}=L\delta^{\alpha}_{\lambda}-\frac{\partial L}{\partial F_{\alpha\beta}}F_{\lambda\beta}. \tag{18}$$

Suppose

$$\Delta x^{\lambda}=\xi^{\lambda}_{k}\epsilon^{k}.$$

Then

$$(\Delta\phi)_{\beta}=-\phi_{\mu}\nabla_{\beta}\xi^{\mu}_{k}\epsilon^{k}$$

and

$$P^{\alpha}=\left(T^{\alpha}{}_{\lambda}\xi^{\lambda}_{k}-\frac{\partial L}{\partial F_{\alpha\beta}}\phi_{\mu}\nabla_{\beta}\xi^{\mu}_{k}\right)\epsilon^{k}=P^{\alpha}_{k}\epsilon^{k}. \tag{19}$$

Due to independence of parameters ϵ^{k} we conclude

$$\nabla_{\alpha}P^{\alpha}_{k}=0 \quad \text{for } k=1,2,\ldots,r.$$

These are the desired conservation laws. It is worthwhile to express vectors P^{α}_{k} in terms of S^{α}_{β} to illustrate the explicit nature of these conservation laws.

In the following let

$$\frac{\partial L}{\partial F_{\alpha\beta}}=L^{\alpha\beta}.$$

Now from (19) we have

$$P^{\alpha}_{k}=T^{\alpha}{}_{\lambda}\xi^{\lambda}_{k}-L^{\alpha\beta}\phi_{\mu}\nabla_{\beta}\xi^{\mu}_{k}.$$

But

$$L^{\alpha\beta}\phi_{\mu}\nabla_{\beta}\xi^{\mu}_{k}=L^{\alpha\beta}[\nabla_{\beta}(\phi_{\mu}\xi^{\mu}_{k})-\nabla_{\beta}\phi_{\mu}\xi^{\mu}_{k}].$$

Hence

$$P^{\alpha}_{k}=T^{\alpha}{}_{\lambda}\xi^{\lambda}_{k}+L^{\alpha\beta}F_{\beta\lambda}\xi^{\lambda}_{k}-L^{\alpha\beta}\nabla_{\beta}(\phi_{\mu}\xi^{\mu}_{k})$$

or

$$P^{\alpha}_{k}=(L\delta^{\alpha}_{\lambda}-L^{\alpha\beta}F_{\lambda\beta}+L^{\alpha\beta}F_{\beta\lambda})\xi^{\lambda}_{k}-L^{\alpha\beta}\nabla_{\beta}(\phi_{\mu}\xi^{\mu}_{k}).$$

Using the definition of S^{α}_{λ}, we see

$$P^{\alpha}_{k}=(S^{\alpha}{}_{\lambda}\xi^{\lambda}_{k})+B^{\rho\alpha}F_{\rho\lambda}\xi^{\lambda}_{k}+\frac{\partial L}{\partial\phi_{\alpha}}\phi_{\lambda}\xi^{\lambda}_{k}$$

$$-\nabla_{\beta}(L^{\alpha\beta}\phi_{\mu}\xi^{\mu}_{k})+\nabla_{\beta}L^{\alpha\beta}(\phi_{\mu}\xi^{\mu}_{k})$$

where $B^{\rho\alpha}=L^{\rho\alpha}+L^{\alpha\rho}$.

Assuming ϕ_{α} satisfies the field equations, then

$$\frac{\partial L}{\partial\phi_{\alpha}}=\nabla_{\beta}L^{\beta\alpha};$$

so we find

$$P^{\alpha}_{k}=(S^{\alpha}{}_{\lambda}\xi^{\lambda}_{k})+B^{\rho\alpha}F_{\rho\lambda}\xi^{\lambda}_{k}+\nabla_{\beta}B^{\beta\alpha}(\phi_{\lambda}\xi^{\lambda}_{k})-\nabla_{\beta}(L^{\alpha\beta}\phi_{\lambda}\xi^{\lambda}_{k}). \tag{20}$$

Derivation of (20) was motivated by the author's observation that most conserved vectors for field theories have the form

$$S^{\alpha}{}_{\lambda}\xi^{\lambda}_{k}$$

which appears as the first term of (20). So it was felt that possibly all conserved vectors associated with geometric symmetries were of this form. In the attempt to show this (20) arose. We see the situation is not so simple as expected. Tensor S^{α}_{β} and generators ξ^{λ}_{k} do not determine conserved vectors by themselves. Although more complicated than expected (20) does represent all possible conserved vectors arising from geometric symmetry and can be useful in recognizing such conservation laws.

For instance, in recent years several papers have appeared describing infinitely many divergenceless expressions believed by some to represent new conserved quantities for electromagnetic fields in vacuum.[6] These so-called conservation laws should be related to some symmetry transformations for the action integral. In the case of electromagnetic fields $L^{\alpha\beta}=-L^{\beta\alpha}$ so $B^{\alpha\beta}=0$ and (20) reduces to

$$P^{\alpha}_{k}=S^{\alpha}_{\lambda}\xi^{\lambda}_{k}-\nabla_{\beta}(L^{\alpha\beta}\phi_{\lambda}\xi^{\lambda}_{k}). \tag{21}$$

All conserved vectors for electromagnetic fields in vacuum connected with geometric symmetries must be of this form. In Minkowski space we have

$$\partial_{\alpha}P^{\alpha}_{k}=\partial_{\alpha}(S^{\alpha}{}_{\lambda}\xi^{\lambda}_{k})-\partial_{\alpha}\partial_{\beta}(L^{\alpha\beta}\phi_{\lambda}\xi^{\lambda}_{k})$$

and the second term vanishes because of antisymmetry

of $L^{\alpha\beta}$. Hence all conservation laws due to geometric symmetries have the form

$$\partial_\alpha(S^\alpha{}_\lambda \xi^\lambda_k) = 0. \tag{22}$$

The symmetry group in this case is the conformal group and (22) yields the fifteen Bessel-Hagen conservation laws. These are the only possible conservation laws for electromagnetic fields in vacuum related to geometric symmetries. The recent conservation laws due to Lipkin and others are not of this form so they cannot be related to geometric symmetries.

5. SUMMARY OF RESULTS

It has been pointed out that the symmetry transformations employed by most authors today when applying the Noether theorems are more restricted than necessary. This suggests defining geometric symmetry transformations as those in current use with some clarification concerning the real meaning of these transformations. It has been shown the geometric symmetry transformations for a given integral are determined by tensor $S^\alpha{}_\beta$ defined by (13). By use of this tensor we find once more the geometric symmetry group for electromagnetic fields is the conformal group in any Riemannian manifold. If we do not insist on geometric transformations then other symmetry transformations may exist and apparently do. It is possible to write down a general form (20) for all conserved vectors related to geometric symmetry. In the case of electromagnetic fields we find the Bessel-Hagen conservation laws are the only ones which can arise from geometric symmetry.

[1]E. Noether, Nachr. Ges. Wiss. Gottingen II, 235 (1918); E. L. Hill, Rev. Mod. Phys. 23, 253 (1951).
[2]E. Bessel-Hagen, Math. Annalen. LXXXIV, 258 (1921).
[3]T. Fulton, F. Rohrlich, and L. Witten, Rev. of Mod. Phys. 34, 443 (1962).
[4]B. F. Plybon, J. Math. Phys. 12, 57 (1971).
[5]K. Yano and S. Bochner, *Curvature and Betti Numbers* (Princeton U. P., 1953).
[6]D. M. Lipkin, J. Math. Phys. 5, 696 (1964); T. A. Morgan, J. Math. Phys. 5, 1659 (1964); D. B. Fairlie, Nuovo Cimento 37, 897 (1965).

Eight-vertex model on the honeycomb lattice*

F. Y. Wu†

Research School of Physical Sciences, The Australian National University, Canberra, ACT. 2601, Australia
(Received 20 December 1973)

The most general vertex model defined on a honeycomb lattice is the eight-vertex model. In this paper it is shown that the symmetric eight-vertex model reduces to an Ising model with a nonzero real or pure imaginary magnetic field H. The equivalent Ising model is either ferromagnetic with $e^{2H/kT}$ real or antiferromagnetic with $e^{2H/kT}$ unimodular. The exact transition temperature and the order of phase transition in the former case are determined. As an application of the result we verify the absence of a phase transition in the monomer-dimer system on the honeycomb lattice.

1. INTRODUCTION

The vertex model in statistical mechanics plays an important role in the study of phase transitions in lattice systems. A case of current interest is the eight-vertex model on a square lattice.[1,2] This is a rather special model in which only a limited number of the possible vertex types are allowed. The most general one on a square lattice would be the sixteen-vertex model.[3] Unfortunately, except in some special cases,[4,5] the behavior of this general model is not known.

In this paper we consider the counterpart of the sixteen-vertex model of a square lattice for the honeycomb lattice. That is, we consider an eight-vertex model defined on the hexagonal lattice. It turns out that we can say a lot more in this case. While the exact solution of this model still proves to be elusive in most cases, we can make definite statements about its phase transition. In particular, the exact transition temperature can be quite generally determined. An application of our result is the verification of the absence of a phase transition in the monomer-dimer system on the honeycomb lattice.

2. DEFINITION OF THE MODEL

In the study of a vertex model one is interested in the evaluation of a graph generating function. Consider a honeycomb lattice and draw bonds (graphs) along the lattice edges such that each edge can be independently "traced" or left "open." Denote the traced (resp. open) edges by solid (resp. broken) lines; then, as shown in Fig. 1, there are eight possible vertex configurations. With each type of vertex configuration we associate a vertex weight a, b, c, or d (see Fig. 1). Our object is to evaluate the generating partition function

$$Z = Z(a, b, c, d) = \sum_G a^{n_0} b^{n_1} c^{n_2} d^{n_3}, \tag{1}$$

where the summation is over all possible graphs on the lattice and, for a given graph G, n_i is the number of vertices having i solid lines (or bonds). This defines an "eight-vertex" model for the honeycomb lattice.

Since all possible vertex types are allowed, this eight-vertex model is the counterpart of the sixteen-vertex model of a square lattice. Note that we do not distinguish the bonds in different directions. Whereas it is possible to consider the further generalization of eight different weights, we shall not go into this complication in this paper. As a motivation we point out some special cases of interest. When $c = d = 0$, the partition function (1) becomes the monomer-dimer generating function for the honeycomb lattice. When $b = d = 0$, Z reduces to the partition function of a zero-field Ising model, which can be evaluated by pfaffians.

In a statistical model of phase transitions, the vertex weights are the Boltzmann factors

$$\begin{aligned} a &= \exp(-\epsilon_0/kT), \quad b = \exp(-\epsilon_1/kT), \\ c &= \exp(-\epsilon_2/kT), \quad d = \exp(-\epsilon_3/kT) \end{aligned} \tag{2}$$

where ϵ_i is the energy of a vertex having i bonds. While the weights (2) are always positive, the symmetry relations to be derived below are valid more generally for any real or complex weights.

3. SYMMETRY RELATIONS

The partition function (1) possesses a number of symmetry properties. Interchanging the solid and broken lines in Fig. 1, we obtain the symmetry relation

$$Z(a, b, c, d) = Z(d, c, b, a). \tag{3}$$

Also since both the total number of vertices, N, and the number of vertices with odd number of bonds are even, we have the negation symmetry

$$\begin{aligned} Z(a, b, c, d) &= Z(-a, -b, -c, -d) \\ &= Z(-a, b, -c, d) \\ &= Z(a, -b, c, -d). \end{aligned} \tag{4}$$

The weak graph expansion[6] yields an additional symmetry relation. For its derivation it is most convenient to use Wegner's formulation[7] of the weak-graph expansion. Denote the vertex weights by $\omega(i,j,k)$, where $i,j,k = \pm 1$ are the edge indices such that $+1$ corresponds to no bond and -1 corresponds to a bond on the edge. I.e., $\omega(+,+,+) = a$, $\omega(+,+,-) = \omega(+,-,+) = \omega(-,+,+) = b$, $\omega(+,-,-) = \omega(-,+,-) = \omega(-,-,+) = c$, and $\omega(-,-,-) = d$. Define a set of new vertex weights $\omega^*(+,+,+) = a^*$, etc. by

$$\omega^*(\alpha, \beta, \gamma) = \sum_{ijk} V_{\alpha i} V_{\beta j} V_{\gamma k} \omega(i, j, k), \tag{5}$$

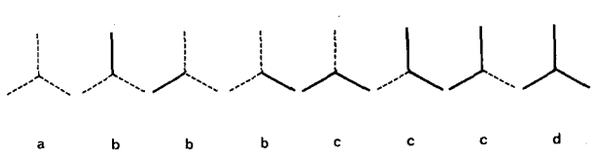


FIG. 1. The eight vertex configurations and the associated weights for a honeycomb lattice.

where the 2×2 matrix V having elements $V_{\alpha i}$ satisfies

$$V\tilde{V}=I, \tag{6}$$

I being the identity matrix. We then have the weak-graph symmetry

$$Z(a,b,c,d)=Z(a^*,b^*,c^*,d^*). \tag{7}$$

There are two possible choices for V:

$$V(y)=(1+y^2)^{-1/2}\begin{pmatrix}1 & y\\ y & -1\end{pmatrix} \tag{8}$$

or

$$U(y)=(1+y^2)^{-1/2}\begin{pmatrix}1 & y\\ -y & 1\end{pmatrix} \tag{9}$$

for aribitrary (real or complex) y. The explicit transformation generated by (8) is

$$\begin{aligned}
a^* &= (1+y^2)^{-3/2}[a+3yb+3y^2c+y^3d],\\
b^* &= (1+y^2)^{-3/2}[ya-(1-2y^2)b+(y^3-2y)c-y^2d],\\
c^* &= (1+y^2)^{-3/2}[y^2a+(y^3-2y)b+(1-2y^2)c+yd],\\
d^* &= (1+y^2)^{-3/2}[y^3a-3y^2b+3yc-d].
\end{aligned} \tag{10}$$

The transformation generated by (9) leads to identical vertex weights subject to the negation symmetry $b^*\to -b^*$; $d^*\to -d^*$ hence is not independent. We shall write (10) in the short-hand notation

$$\omega^*(y)=V(y)\omega. \tag{11}$$

It is also seen that two consecutive transformations are equivalent to a single one:

$$V(x)V(y)=U\left(\frac{y-x}{1+xy}\right). \tag{12}$$

In particular we have

$$V(y)V(y)=I. \tag{13}$$

4. SPECIAL SOLUTIONS

Before we consider the model with general weights, it is useful to first consider some special cases whose solutions are known

A. $b=ua$, $c=u^2a$, $d=u^3a$

The vertex weights in this case can be converted into the bond weight u^2. Since *all* graphs are included in (1), we then obtain

$$\begin{aligned}
Z &= a^NZ(1,u,u^2,u^3)\\
&= a^N(1+u^2)^{3N/2}.
\end{aligned} \tag{14}$$

Here we see a simple example for which the partition function (1) does not exhibit a phase transition.

B. $b=d=0$

Here only the vertices with even number (0 or 2) of bonds are allowed. The graphs in (1) are then precisely those encountered in the high-temperature expansion of a zero-field Ising model. Writing

$$c/a=\tanh K, \tag{15}$$

we then obtain

$$\begin{aligned}
Z &= Z(a,0,c,0)\\
&= a^NZ(1,0,\tanh K,0)\\
&= a^N2^{-N}(\cosh K)^{-3N/2}Z_{\text{Ising}}(0,K),
\end{aligned} \tag{16}$$

where more generally $Z_{\text{Ising}}(L,K)$ is the partition function of an Ising model on the honeycomb lattice with interactions $-kTK$ and a magnetic field $-kTL$. From the known expression[8] of $Z_{\text{Ising}}(0,K)$ given by (A1) we obtain, in the large N limit,

$$\begin{aligned}
\frac{1}{N}\ln Z &= (16\pi^2)^{-1}\int_0^{2\pi}d\theta\int_0^{2\pi}d\phi\,\ln\{a^4+3c^4+2(c^4-a^2c^2)\\
&\quad\times[\cos\theta+\cos\phi+\cos(\theta+\phi)]\}.
\end{aligned} \tag{17}$$

We remark that (17) is valid for arbitrary (real or complex) a and c, although the physical range of an Ising model is restricted to real values satisfying $|c/a|\le 1$. The expression (17) is nonanalytic at

$$a/c=\pm\sqrt{3}. \tag{18}$$

Other established properties of $Z_{\text{Ising}}(L,K)$ for $L\neq 0$ are summarized in the Appendix.

C. $a=d$, $b=c$

The vertex weights are now symmetric under the interchange of the solid and the broken lines in Fig. 1. In this case we can again reduce the partition function to the form of (16). Indeed, taking $y=1$ in (10), we obtain

$$Z=Z((a+3b)/\sqrt{2},0,(a-b)/\sqrt{2},0). \tag{19}$$

The phase transition now occurs at

$$a/b=3\pm2\sqrt{3}. \tag{20}$$

D. $ad=bc$

In this case we define the Ising parameters L and K by

$$z=\tanh K=c/a,\quad \tau=\tanh L=b/\sqrt{ac}. \tag{21}$$

Then

$$\begin{aligned}
Z &= a^NZ(1,\sqrt{z}\tau,z,z^{3/2}\tau)\\
&= a^N2^{-N}(\cosh L)^{-N}(\cosh K)^{-3N/2}Z_{\text{Ising}}(L,K)\\
&= (2a^3c)^{-N}(ac-b^2)^N(a^2-c^2)^{3N/2}Z_{\text{Ising}}(L,K).
\end{aligned} \tag{22}$$

Here the second step follows from the generalization of (16) to the high-temperature expansion of $Z_{\text{Ising}}(L,K)$.

E. $b^2=ac$

In this case we have

$$Z=a^NZ(1,u^{-1},u^{-2},d/a), \tag{23}$$

where $u=a/b$. The partition function on the rhs of (23) is in a form similar to that considered in Ref. 5. We then obtain in a similar fashion[9]

$$Z=(b/a)^{2N}(1+a^2/b^2)^{3N/8}(ad/bc-1)^{N/2}Z_{\text{Ising}}(L,K), \tag{24}$$

where

$$\begin{aligned}
\exp(4K) &= 1+a^2/b^2,\\
\exp(2L) &= (1+a^2/b^2)^{3/2}(ad/bc-1)^{-1}.
\end{aligned} \tag{25}$$

We see that the Ising model is ferromagnetic for real

a/b. For the Boltzmann weights (2) (subject to $2\epsilon_1=\epsilon_0+\epsilon_2$), we find the model in general exhibits no phase transition, except for $\epsilon_0<\epsilon_1 (a>b)$ and $\gamma_0<(\epsilon_3-\epsilon_0)(\epsilon_1-\epsilon_0)^{-1}<0$ the model has a first-order phase transition at $\exp(2L)=1$ or

$$(a^2+b^2)^{3/2}=a^2d-b^3. \tag{26}$$

Here $\gamma_0=3-2\ln(27+15\sqrt{3})/\ln(6+4\sqrt{3})=-0.102\,2204\cdots$.

5. GENERAL CASE

We are now in a position to discuss the general solution for arbitrary (positive) vertex weights (2). The idea is to introduce the weak-graph transformation (10) and choose y to make the new vertex weights satisfying either $a^*d^*=b^*c^*$ or $b^{*2}=a^*c^*$. We can then use the results of the Appendix to determine the critical behavior of the vertex model. For clarity we use subscripts 1 and 2 to distinguish the two cases. That is, in analogy to (11), we write

$$\omega_i^*\equiv\omega^*(y_i)=V(y_i)\omega,\quad i=1,2, \tag{27}$$

and consider the two cases separately.

(i) $a_1^*d_1^*=b_1^*c_1^*$: From (27) and (10) we find y_1 given by

$$y_1^2-2Ay_1-1=0, \tag{28}$$

where $A=(b^2-ac+bd-c^2)/(ad-bc)$. The new vertex weights $\omega_i^*=\{a_1^*,b_1^*,c_1^*,d_1^*\}$ are real if we take the positive solution

$$y_1=A+(A^2+1)^{1/2}>0. \tag{29}$$

Then, from (10), $a_1^*>0$. Also c_1^* is real since

$$a_1^*+c_1^*=(1+y_1^2)^{-1/2}(a+by_1+c+dy_1)>0. \tag{30}$$

The partition function is now

$$Z=(2a_1^{*3}c_1^*)^{-N}(a_1^*c_1^*-b_1^{*2})^N(a_1^{*2}-c_1^{*2})^{3N/2}\times Z_{\text{Ising}}(L_1^*,K_1^*), \tag{31}$$

where

$$\exp(2K_1^*)=(a_1^*+c_1^*)/(a_1^*-c_1^*),$$
$$\exp(2L_1^*)=[(a_1^*c_1^*)^{1/2}+b_1^*]/[(a_1^*c_1^*)^{1/2}-b_1^*]. \tag{32}$$

We observe that $\exp(2K_1^*)<1$, $L_1^*=$pure imaginary if $c_1^*<0$. Since not much is known about $Z_{\text{Ising}}(L_1^*,K_1^*)$ for K_1^* and L_1^* in these ranges, we shall be interested only in $c_1^*>0$. We observe in particular that, for a_1^* and c_1^* positive, $\exp(2L_1^*)\neq-1$.

(ii) $b_2^{*2}=a_2^*c_2^*$: From (27) and (10) we find y_2 given by

$$(bd-c^2)y_2^2+(ad-bc)y_2+(ac-b^2)=0. \tag{33}$$

The partition function is then

$$Z=(b_2^*/a_2^*)^{2N}(1+a_2^{*2}/b_2^{*2})^{3N/8}\times(a_2^*d_2^*/b_2^*c_2^*-1)^{N/2}Z_{\text{Ising}}(L_2^*,K_2^*). \tag{34}$$

Here the weights a_2^*,b_2^*,c_2^*,d_2^* are real if the discriminant

$$\Delta\equiv(ad-bc)^2-4(bd-c^2)(ac-b^2) \tag{35}$$

is positive. The parameters K_2^* and L_2^* are given by (25) with $a\to a_2^*$, etc. After some steps we find the simple result:

$$\exp(4K_2^*)=1+\Delta/(bd-c^2+ac-b^2)^2>0. \tag{36}$$

We shall consider $\Delta>0$ which corresponds to K_2^* being ferromagnetic. The similar expression of L_2^*, which is not needed for our discussions, is rather complicated and will not be given.

The two transformations (i) and (ii) are obviously related. To see the relationship, we observe from (27), (12), and (13) that

$$\omega_2^*=V(y_2)V(y_1)\omega_1^*=U\left(\frac{y_1-y_2}{1+y_1y_2}\right)\omega_1^*. \tag{37}$$

Since (34) is invariant under the negation of b_2^* and d_2^*, there exists a single transformation which relates ω_1^* to ω_2^*. To effect this transformation, we set $ad=bc$ in (33) and obtain $y_2=(a/c)^{1/2}$. The new weights are then

$$\begin{aligned}a_2^*&=4(1+a/c)^{-3/2}(a/c)^{1/2}(b+\sqrt{ac}),\\ b_2^*&=2(1+a/c)^{-3/2}(a/c-1)(b+\sqrt{ac}),\\ c_2^*&=b_2^{*2}/a_2^*,\\ d_2^*&=(1+a/c)^{-3/2}(a^{5/2}/c^{3/2}-3ab/c+3\sqrt{ac}-bc/a).\end{aligned} \tag{38}$$

Now (36) becomes, for $ad=bc$,

$$\exp(4K_2^*)=[(a+c)/(a-c)]^2. \tag{39a}$$

Also using (38), we find

$$\begin{aligned}\exp(2L_2^*)&=(\sqrt{ac}+b)/(\sqrt{ac}-b),\quad \text{if } a/c>1,\\ &=(b+\sqrt{ac})/(b-\sqrt{ac}),\quad \text{if } a/c<1.\end{aligned} \tag{39b}$$

Letting $a=a_1^*$, $b=b_1^*$, $c=c_1^*$, $d=d_1^*$ in (39) and comparing with (32), we then obtain the relation

$$\exp(4K_2^*)=\exp(4K_1^*),$$
$$\exp(2L_2^*)=\pm\exp(2L_1^*),\quad \text{for } a_1^*/c_1^*\gtrless 1. \tag{40}$$

Note that while $\exp(2K_2^*)$ can be taken to be positive, $\exp(2K_1^*)$ can be either positive or negative. We observe from (40), (32), and (36) that $\Delta>0$ and $c_1^*>0$ are equivalent. Hence, for $\Delta>0$, K_2^* is ferromagnetic and $\exp(2L_2^*)$ is real.

Using the results of the Appendix, we conclude that, for $\Delta>0$, the nonanalyticity of Z can occur only at $\exp(2L_2^*)=+1$ or -1. To distinguish these two cases, we turn to L_1^*. Since $\exp(2K_1^*)$ may be negative, it is then convenient to consider the following situations separately:

(i) $a_1^*>c_1^*>0$: From (40) and $\exp(2L_1^*)\neq-1$, the nonanalyticity can occur only at $\exp(2L_1^*)=\exp(2L_2^*)=1$. By using (32) this is equivalent to

$$b_1^*=d_1^*=0. \tag{41}$$

A little algebra using (28) reduces (41) to

$$2(ab-cd)[(b^2-ac+bd-c^2)^2-(ad-bc)^2]+(ad-bc)(b^2-ac+bd-c^2)\times(a^2+d^2-3b^2-3c^2-2ac-2bd)=0 \tag{42}$$

which defines $T=T_c$. To see whether indeed a phase transition occurs at T_c, we observe that K_1^* and K_2^* are

equal and positive. Then from the result of the Appendix we need to compute $z_c = (c_1^*/a_1^*)_{T=T_c}$. The vertex model will exhibit a first-order transition if $z_c > 1/\sqrt{3}$, a second-order transition with an infinite specific heat if $z_c = 1/\sqrt{3}$, and no transition at all if $z_c < 1/\sqrt{3}$, even if (42) has a solution. The following useful expression of z_c is obtained by combining (29), (10) and (41):

$$z_c = \left[\frac{4(ac+bd)A^2 + 4(ab-cd)A + (3b+d)(d-b) + (a+3c)(a-c)}{4(a^2+d^2)A^2 + 12(ab-cd)A + (3b+d)^2 + (a+3c)^2}\right]_{T_c}. \tag{43}$$

(ii) $c_1^* > a_1^* > 0$: In this case the nonanalyticity occurs only at $\exp(2L_2^*) = -\exp(2L_1^*) = -1$. Then T_c is again given by (41) or (42). Now $K_2^* > 0$ and $\exp(2L_2^*) = -1$; hence the vertex model always has a first-order transition. Note that we can reach the same conclusion by considering K_1^*. In this case $\exp(2K_1^*) < -1$ and $\exp(2L_1^*) = 1$. We need only to reverse the signs of $\exp(2K_1^*)$ and $\exp(2L_1^*)$ which leaves $Z_{\text{Ising}}(L_1^*, K_1^*)$ unchanged, as can be seen from the low-temperature expansion.

Combining the results in (i) and (ii), we conclude that a phase transition occurs for $\Delta > 0$ only if $z_c \geq 1/\sqrt{3}$.

A special case is that (41) or (42) is an identity. Then, for *all* Δ, $L_1^* = 0$ and Z reduces to that of a zero-field Ising model. The vertex model now exhibits the Ising-type transition (logarithmic specific heat singularity) at T_c defined by

$$\Delta/(bd - c^2 + ac - b^2)^2 = (2+\sqrt{3})^{\pm 2} - 1. \tag{44}$$

Unfortunately we are unable to make any general statement for $\Delta < 0$. For $\Delta < 0$, K_2^* is antiferromagnetic and $\exp(2L_2^*)$ is unimodular and lies on the unit circle. Presumably the zeros of an Ising antiferromagnet also distribute along the unit circle in the thermodynamic limit.[10] The vertex model then in general shows a unique transition.

6. SUMMARY

We have established the following results for the vertex model (2):

(i) If (42) is an identity, then an Ising-type transition occurs at T_c defined by (44), where Δ is given in (35).

(ii) For $\Delta \geq 0$ and (42) not an identity, a phase transition occurs at T_c defined by (42) if $z_c \geq 1/\sqrt{3}$, where z_c is given in (43). Otherwise ($z_c < 1/\sqrt{3}$) there is no phase transition. The transition is of first-order except that the specific heat diverges for $z_c = 1/\sqrt{3}$.

(iii) For $\Delta < 0$ and (42) not an identity, the vertex model is related to an Ising antiferromagnet with a pure imaginary magnetic field. Nature of the transition is not known.

It is instructive to illustrate with some examples.

(i) $a = d$, $b = c$: Since (42) is an identity, we find from (44) the critical condition

$$(a^2 + 2ab - 3b^2)/4b^2 = (2 \pm \sqrt{3})^2 - 1,$$

which agrees with (20).

(ii) $b^2 = ac$: We find $\Delta = (a^2 d - b^3)^2/a^2 > 0$ and $\exp(4K_2^*) = 1 + a^2/b^2$. This is in agreement with (25). It can be verified that the condition (42) is the same as that obtained from $\exp(2L) = 1$ in (25).

(iii) $b = c = d$: We find $\Delta = b^2(a-b)^2 > 0$ and $\exp(4K_2^*) = 2$. Since K_2^* is a constant with $z_c^{-1} = 3 + 2\sqrt{2} > \sqrt{3}$, there is no phase transition.

(iv) *Monomer-dimer system*: For $c = d = 0$ the partition function (1) becomes the monomer-dimer generating function $Z_{\text{MD}}(a, b^2)$ where a and b^2 are, respectively, the monomer and dimer activities. It is known that this system does not have a phase transition.[11] We verify this by observing that $\Delta = 0$, $K_2^* = 0$. Also (42) has no solution for $c = d = 0$, $ab \neq 0$.

To obtain a closed expression for Z_{MD}, we find that, for $c = d = 0$, either $\exp(2K_2^*) = 1$, $\exp(2L_2^*) = -1$ or $\exp(2K_1^*) = -1$, $\exp(2L_1^*) = 1$. In either case the Ising partition function is identically zero. Therefore we must take the limit $c = d \to 0$ appropriately. This leads to the expression

$$Z_{\text{MD}}(a, b^2) = \lim_{c \to 0} (b/4c)^N Z_{\text{Ising}}(L_2^*, K_2^*) \tag{45}$$

where (for small c)

$$\exp(2K_2^*) = 1 + 4c/b,$$

$$\exp(2L_2^*) = -1 \pm 2a\sqrt{c}/b^{3/2}.$$

ACKNOWLEDGMENTS

I wish to thank Professor K.J. Le Couteur for his hospitality at The Australian National University, and Dr. R.J. Baxter for a discussion on the weak-graph expansion. The support of the Australian—American Educational Foundation is also gratefully acknowledged.

APPENDIX: ISING PARTITION FUNCTION

We summarize in this Appendix the relevant properties of the Ising partition function $Z_{\text{Ising}}(L, K)$.

A closed expression is known for $L = 0$. In the large N limit, one has[8]

$$\frac{1}{N}\ln Z_{\text{Ising}}(0, K) = \tfrac{3}{4}\ln 2 + \frac{1}{16\pi^2}\int_0^{2\pi} d\theta \int_0^{2\pi} d\phi \times \ln[c^3 + 1 - s^2(\cos\theta + \cos\phi + \cos(\theta+\phi))], \tag{A1}$$

where

$$c = \cosh 2K, \quad s = \sinh 2K.$$

The second derivative of (A1) diverges logarithmically at $\tanh K = \pm 1/\sqrt{3}$.

A unique property of the honeycomb lattice (coordination number = odd) is that the partition functions at $L = i\frac{1}{2}\pi$ and $L = 0$ are related. To see this connection, consider the high-temperature expansion of $Z_{\text{Ising}}(L, K)$. Using the identities for $L = i\frac{1}{2}\pi$,

$$\sum_{\sigma = \pm 1} \sigma \exp(L\sigma) = 2\sinh L = 2i, \tag{A2}$$

$$\sum_{\sigma=\pm 1} \exp(L\sigma) = 2\cosh L = 0,$$

we see that only the vertices with odd number of bonds contribute in the expansion. Thus we obtain

$$Z_{\text{Ising}}(i\tfrac{1}{2}\pi, K) = (2i)^N (\cosh K)^{3N/2} Z(0, \sqrt{z}, 0, z^{3/2}) \tag{A3}$$
$$= Z_{\text{Ising}}(0, \tilde{K}),$$

where

$$\tanh\tilde{K} \tanh K = 1.$$

The last step follows from the symmetry relation (3) and (16). Note that $Z_{\text{Ising}}(i\frac{1}{2}\pi, K)$ is analytic for real K.

Most of the established properties for $L \neq 0$ are for ferromagnetic interactions ($K > 0$). For $K > 0$, $Z_{\text{Ising}}(L, K)$ can be nonanalytic in L or K only at $|\exp(2L)| = 1$.[12,13] This means $\exp(2L) = \pm 1$ for real $\exp(2L)$. At $\exp(2L) = 1$ the analyticity extends to all $0 < z < 1/\sqrt{3}$ while the first derivative w.r.t. L is discontinous for all $1/\sqrt{3} < z < 1$. At $\exp(2L) = -1$ this first derivative is presumably discontinuous for all $0 < z < 1$. This is similar to the result of a square lattice[14] and can be easily seen to hold in both the high and low temperature limits. We hope to return in the future for an exact calculation of this discontinuity.

*Work supported in part by the NSF Grant No. GH-35822 at Northeastern University.

†Senior Fulbright Scholar on leave of absence from Department of Physics, Northeastern University, Boston, Massachusetts 02115 (permanent address).

[1]C. Fan and F.Y. Wu, Phys. Rev. B 2, 723 (1970).
[2]R.J. Baxter, Ann. Phys. (N.Y.) 70, 193 (1972).
[3]For a general introduction of the sixteen-vertex model see E.H. Lieb and F.Y. Wu, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M.S. Green (Academic, London, 1972), Vol. I.
[4]F.Y. Wu, Solid State Commun. 10, 115 (1972).
[5]F.Y. Wu, Phys. Rev. B 6, 1810 (1972).
[6]J.F. Nagle, J. Math. Phys. 8, 1007 (1968).
[7]F. Wegner, Physica 68, 570 (1973).
[8]See, e.g., Eq. (131) in C. Domb, Adv. Phys. 9, 149 (1960).
[9]F.Y. Wu, Phys. Rev. Lett. 32, 460 (1974).
[10]For a discussion on a plausible distribution of zeros for an Ising antiferromagnet (due to M.E. Fisher), see M. Suzuki, C. Kawabata, S. Ono, Y. Karaki, and M. Ikeda, J. Phys. Soc. Japan 29, 837 (1970).
[11]O.J. Heilmann and E.H. Lieb, Commun. Math. Phys. 25, 190 (1972); Phys. Rev. Lett. 24, 1412 (1970).
[12]T.D. Lee and C.N. Yang, Phys. Rev. 87, 410 (1952).
[13]J.L. Lebowitz and O. Penrose, Commun. Math. Phys. 11, 99 (1968).
[14]Closed expressions for the free energy and the magnetization of a square Ising lattice at $L = i\frac{1}{2}\pi$ were first given in Ref. 12. See also B.M. McCoy and T.T. Wu, Phys. Rev. 155, 438 (1967).

Heat conduction and sound transmission in isotopically disordered harmonic crystals*

A. J. O'Connor† and J. L. Lebowitz

Belfer Graduate School of Science, Yeshiva University, New York, New York 10033
(Received 1 November 1973)

We investigate some kinetic properties of an isotopically disordered harmonic crystal. We prove rigorously that for almost all disordered chains the transmission coefficient of a plane wave with frequency ω, $t_N(\omega)$, decays exponentially in N, the length of the disordered chain, with the decay constant proportional to ω^2 for small ω. The response of this system to an incident wave is related to the nature of the heat flux $J(N)$ in a disordered chain of length N placed between heat reservoirs whose temperatures differ by $\Delta T > 0$. We clarify the relationship between the works of various authors in the heat conduction problem and establish that for all models $J(N) \to 0$ as $N \to \infty$ in a disordered system. The exact asymptotic dependence of $J(N)$ on N eludes us, however. We also investigate the heat flow in a simple stochastic model for which Fourier's law is shown to hold. Similar results are proven for two-dimensional systems disordered in one direction.

1. INTRODUCTION

There does not exist at the present time any dynamical system for which kinetic laws can be proven to hold. A kinetic law relates fluxes to gradients, e.g., Fourier's law of heat conduction. Indeed the two "standard models" of equilibrium statistical mechanics, that of the noninteracting gas for an ideal fluid and the perfect harmonic crystal for the ideal solid do not obey any macroscopic kinetic laws. The next dynamical model, in order of complexity, is the isotopically disordered harmonic system where the masses of the individual particles are independent identically distributed random variables. This paper studies the transport properties of such a system, particularly those of the disordered harmonic chain: We give new rigorous proofs of some already known (or conjectured) results and derive a few new ones.

We consider a disordered chain in which left and right end particles are coupled by some mechanism to heat baths at different temperatures; call them T_L and T_R, $T_L - T_R = \Delta T$. Using some description of the coupling to the heat baths,[1] we can compute the steady state energy flow across the chain. If $J(N)$ is the flow across a particular chain of length N and $\langle J(N)\rangle$ the average of $J(N)$ over the different choices of the N masses we identify $N^{-1}\,\Delta T$ as the "temperature gradient" across the chain and define the average conductivity of chains with length N by $K(N) = \langle J(N)\rangle/(\Delta T/N)$. Fourier's law will hold if $K(N) \to K$ as $N \to \infty$ with K a finite, strictly positive, constant. Fourier's law certainly fails for periodic systems. In these $J(N)$ tends to a nonzero constant as N increases, i.e., $K(N)$ grows linearly with N. This was proven for the homogeneous chain in[1] and for the general periodic chain in.[2] The behavior of $J(N)$ does not depend on the dimensionality of the system: Hellemann[3] investigated two-dimensional homogeneous cylindrical systems with general couplings and found the same behavior for $J(N)$ (see also Nakazawa[4]). For truly disordered chains the situation is entirely different. Here Casher and Lebowitz[2] proved that $J(N) \to 0$ as $N \to \infty$ for almost every random chain (almost all is defined here with respect to the probability measure on the chains constructed from the individual distribution of each mass). Ideally we would like to decide if $\lim_{N\to\infty} K(N)$ is finite, zero, or infinite. Regretably we still cannot do this in a definite way. Some heuristic arguments suggest[5] that for the linear chain $K(N)$ decreases as $N^{-1/2}$. This, if true, implies that the random chain is an even poorer heat conductor than a real system and would presumably be a peculiarity related to the chain being one-dimensional. We would then have to look at two- and three-dimensional random harmonic systems to obtain models in which Fourier's law holds.

Intuitively, we picture the heat baths exciting the ends of the chain and setting up vibrations which travel along the chain. These vibrations are linear combinations of the chain's normal modes. The energy flow therefore depends on the fraction of normal modes which have significant amplitudes at both ends of the chain. We could say that a normal mode which has significant amplitude at both ends is an efficient heat carrier. In periodic systems (i.e., m_i is periodic in i) nearly every mode is efficient and so that heat flow $J(N)$ through a periodic chain of length N approaches a nonzero limit with increasing chain length. In a disordered system on the other hand nearly every mode is "localized" and so relatively few are efficient heat conductors. This leads to $J(N) \to 0$ as $N \to \infty$ for these systems.

The difference between the normal modes in periodic and disordered systems is reflected in the spectrum of the corresponding infinite chains and the character of plane wave solutions to the lattice equations of motion [a plane wave solution is one of the type $\mathbf{u}(t) = \mathbf{u}(0)e^{i\omega t}$]. In a periodic chain the frequency spectrum consists of allowed bands separated by gaps. The bands are actually the spectrum of an infinite-dimensional self-adjoint matrix operator. The spectrum of this operator is absolutely continuous. At allowed frequencies the plane wave solutions are bounded periodic functions on the chain. For frequencies in the band gaps the plane wave solutions grow or decrease exponentially. In a disordered chain the corresponding spectrum is more complicated. For almost all chains the corresponding infinite matrix operator does not have any absolutely continuous spectrum.[2] Indeed for every frequency $\omega > 0$ the plane wave solutions of the equations of motion of the semiinfinite chain (1-2) grow exponentially for almost all chains. Borland[6] was the first to appreciate this exponential growth (he used it to explain the frequent occurence of localized modes in random systems). A rigorous proof of the existence of exponentially

growing solutions for infinite chains was, however, only obtained when Matsuda and Ishii[5] proved that the powerful results of Furstenberg[7] apply to this system.

The harmonic chain

We will now specify more precisely the dynamical system with which we are primarily concerned here. The harmonic chain is a one-dimensional system of particles coupled together by harmonic springs. The force between two adjacent particles is proportional to the change in the length of the connecting spring: When both particles are in their equilibrium position this force is zero. At the nth site there is a particle with mass m_n whose displacement from its equilibrium position at time t is $u_n(t)$. The center of mass movement of a finite chain of length N can be removed by constraining the first and last particles by additional harmonic forces. The Hamiltonian for the system with the spring constant set equal to one, is then

$$H=\sum_{n=1}^{N}\tfrac{1}{2}m_n\dot{u}_n^2+\sum_{n=1}^{N-1}\tfrac{1}{2}(u_n-u_{n+1})^2+\tfrac{1}{2}u_1^2+\tfrac{1}{2}u_N^2. \qquad (1.1)$$

This is often described as a chain with fixed boundary conditions because it is also obtained by considering a chain beginning with a particle labelled 0 and ending with one labelled $N+1$ and demanding that $u_0=u_{N+1}\equiv 0$. The equation of motion for the chain is

$$M_N\ddot{\mathbf{u}}(t)+\Phi_N\mathbf{u}(t)=\mathbf{0}. \qquad (1.2)$$

$\mathbf{u}(t)$ is the column vector $[u_1(t),\ldots,u_N(t)]$, M_N is the diagonal matrix with entries $m_1\cdots m_N$. Φ_N is the $N\times N$ tridiagonal matrix with entries $(\Phi_N)_{ii}=2$, $(\Phi_N)_{ij}=-1$ for $|i-j|=1$ and $(\Phi_N)_{ij}=0$ otherwise.

This harmonic chain has N normal modes, i.e., solutions of the form $\mathbf{u}(t)=\mathbf{u}(0)\,e^{i\omega t}$. Any solution of (1.2) with specified initial conditions is a linear combination of these solutions. For a normal mode (1.2) becomes

$$M_N\omega^2\mathbf{u}(0)=\Phi_N\,\mathbf{u}(0)$$

or

$$\omega^2(M_N^{1/2}\mathbf{u}(0))=M_N^{-1/2}\,\Phi_N(M_N^{-1/2}\,M_N^{1/2}\,\mathbf{u}(0)) \qquad (1.3)$$

The normal mode frequencies are thus (determined by) the eigenvalues of the symmetric matrix $H_N=M_N^{-1/2}\Phi_N M_N^{-1/2}$.

These ideas extend to a semiinfinite (or infinite) harmonic chain. Again m_n and u_n are the mass of the nth particle and its displacement from its equilibrium position; n runs from 1 (or $-\infty$) to ∞. We will assume that all m_i are bounded above and below. The equations of motion for the semiinfinite chain are

$$m_n\ddot{u}_n+2u_n-u_{n-1}-u_{n+1}=0 \quad (n>1), \qquad (1.4)$$

$$m_1\ddot{u}_1+2u_1-u_2=0.$$

The energy

$$E(t)=\sum_{n=1}^{\infty}\tfrac{1}{2}m_n\dot{u}_n^2+\sum_{n=1}^{\infty}\tfrac{1}{2}(u_n-u_{n+1})^2+\tfrac{1}{2}u_1^2$$

is a conserved quantity and so the set of solutions to (1.4) with finite initial energy span a Hilbert space h whose norm is just the energy functional. For these solutions (1.4) can be written

$$M\ddot{\mathbf{u}}(t)+\Phi\mathbf{u}(t)=0. \qquad (1.5)$$

Φ is the bounded self-adjoint tridiagonal matrix operator on h with entries $\Phi_{ii}=2$, $\Phi_{ij}=-1$ if $|i-j|=1$ and $\Phi_{ij}=0$ otherwise. M is the infinite diagonal matrix with entries m_i. It is a bounded operator on h. The "allowed" frequencies ω^2 are points in the spectrum of the symmetric operator $H=M^{-1/2}\Phi M^{-1/2}$ with $\|H\|\leq 4/\overline{m}$, $\overline{m}=\min_i\{m_i\}$.

Models of stationary heat flow

In this paper we are interested primarily in the behavior of the thermal conductivity of a random harmonic crystal. Since we are interested in a stationary flow we need to have our system coupled at its ends (left and right) to some kind of inexhaustible heat reservoirs which are maintained at temperatures T_L and T_R so that energy will flow steadily across the system from left to right due to the temperature difference $T_L-T_R=\Delta T>0$.

In this note we use two models for the heat bath and its coupling to the chain. These were developed by Lebowitz *et al.*[1,2] and by Rubin and Greer.[8] In Lebowitz's model the heat bath is a Maxwellian gas of very light molecules. These gas molecules collide with the end particles of the chain. At each collision the momentum of the end particle is altered in a discontinuous way. Using the Maxwell–Boltzmann distribution for the velocities of the gas particles prior to a collision we can compute the probability per unit of time that the momentum of an end (chain) particle will jump from p to p'. This will depend on the gas temperature and the frequency of these collisions. The frequency is incorporated into a constant λ measuring the coupling between the particle and the heat bath. Finally we get a modified Liouville equation for the Gibbs ensemble density $\mu(u_1\ldots u_N, p_1\ldots p_N, t)$ of the system,[1,2] $(p_i=m_i\dot{u}_i)$,

$$\frac{\partial\mu}{\partial t}=\sum_{i,j=1}^{2N}\left[\frac{\partial}{\partial x_i}(a_{ij}x_j\mu)+\tfrac{1}{2}d_{ij}\left(\frac{\partial^2\mu}{\partial x_i\partial x_j}\right)\right]. \qquad (1.6)$$

$x=(x_1,\ldots x_{2N})=(u_1\ldots u_N,\ p_1\ldots p_N)$ and a_{ij} and d_{ij} are entries in the $2N\times 2N$ matrix

$$a=\begin{pmatrix}0 & -M_N^{-1}\\ \Phi_N & L\end{pmatrix},\qquad d=\begin{pmatrix}0 & 0\\ 0 & 2M_NLT\end{pmatrix}.$$

L and T are diagonal $N\times N$ matrices with entries

$$L_{ii}=\lambda(\delta_{i1}+\delta_{iN}).$$

$$T_{ii}=T_L\,\delta_{i1}+T_R\,\delta_{iN}.$$

λ represents the coupling of the baths to the system ($\lambda\geq 0$). When $\lambda=0$ the system is isolated and follows the equations of motion given in (1.2). The solution of Liouville's equation (1.6) with $\lambda=0$ has the form

$$\mu[x_1\cdots x_{2N};t]=\mu[x_1(-t)\cdots x_{2N}(-t);0]$$

where $x_i(t)$ is the solution of (1.2) with initial values $x_1\cdots x_{2N}$. In this case μ will not approach any stationary state. For $\lambda>0$, however, any initial distribution ap-

proaches a unique stationary distribution which is a generalized Gaussian. The expected value of the heat flow across the system in the stationary state, for a specified set of masses $\{m_i\}$, is[2]

$$J(N) = \pi^{-1}(T_L - T_R)\lambda^2 m_1 m_N \int_{-\infty}^{\infty} \omega^2 \, |Z(\omega)^{-1}_{1N}|^2 \, d\omega$$
$$= \pi^{-1}(T_L - T_R)\lambda^2 m_1 m_N \int_{-\infty}^{\infty} \omega^2 j_N(\omega)\, d\omega. \qquad (1.7)$$

Here $Z(\omega)$ is the $N\times N$ matrix $\Phi_N - \omega^2 M_N - i\omega M_N L$ and

$$j_N(\omega) = \{2m_1 m_N \lambda^2\omega^2 + K^2_{1,N} + \lambda^2\omega^2(m_N^2 K^2_{1,N-1} + m_1^2 K^2_{2,N}) + \lambda^4\omega^4 m_1^2 m_N^2 K^2_{2,N-1}\}^{-1}. \qquad (1.8)$$

$K_{ij}(\omega^2)$ is the determinant of the submatrix of $\Phi - \omega^2 M$ beginning with the ith row and column and ending at the jth row and column.

Rubin and Greer's model[8] is rather different. In it the chain of N particles (which constitutes the system) is connected at either end to semiinfinite chains of identical particles. Initially the left-and right-hand chains are in thermal equilibrium at temperatures T_L and T_R, respectively. We can follow the time developement of the infinite system from a specified initial state and at any later time we can compute such quantities as the local temperature or energy flow. More interestingly we can find their average values over the ensemble of initial states and then calculate the steady state value approached as $t\to\infty$, of these averaged quantities. In the next section we give a simpler rederivation of Rubin's result relating the stationary heat flow $\hat{J}(N)$ in his model to the integral of the square of the transmission coefficient $t_N^2(\omega)$. This uses a method introduced by Ford, Kac, and Mazur.[9] Our approach is quite similar to that of Casher and Lebowitz.[2]

In Sec. 3 we use Furstenberg's theorem to prove rigorously an earlier result of Rubin, based on an explicit but not entirely rigorous computation that, in a chain with random masses, $N^{-1}\lim|t_N(\omega)|\to -\gamma(\omega)$ as $N\to\infty$, with $\gamma(\omega)>0$ for $\omega\neq 0$. It follows from this that $\hat{J}(N)$ like $J(N)\to 0$ as $N\to\infty$ for almost all random chains. We also show that $\gamma(\omega)$ is a continuous function of ω for small ω and $\gamma(\omega)/\omega^2\to$ const for $\omega\to 0$. The latter result was proven earlier by Matsuda and Ishii[5] using a perturbation expansion.

In Sec. 4 we use the Casher—Lebowitz expression for the heat flux, $J(N)$ to derive an explicit expression for the nonvanishing heat flow in an infinite periodic diatomic chain and in a uniform chain containing a single impurity. We then, in Sec. 5, derive rigorously an expression for the weak coupling limit of the heat flux $J(N,\lambda)$ where λ is the coupling to the heat reservoirs, i.e., we compute $\lim_{\lambda\to 0}\lambda^{-1}J(N,\lambda)$ and find it in agreement with the perturbation result of Matsuda and Ishii.[5] We note however that the interchange of the limits $\lambda\to 0$ and $N\to\infty$ should not be expected to be valid when the heat flow vanishes as $N\to\infty$. This is shown explicitly in Sec. 6 where we construct a nondynamical model which obeys Fourier's law of heat conduction.

Section 7 discusses the generalization of the Rubin formalism to a two-dimensional harmonic square lattice in which the masses in each column are the same and the heat flow is along the x-axis. We find, as expected,[2] that when the system is periodic the heat flux per unit cross-sectional area does not vanish when the length of the system becomes infinite. When the masses in the different columns are random, then the analog of the Casher—Lebowitz argument for chains, based on the Furstenberg theorem, shows that the flux vanishes. The difference between the two flows is, as in the case of chains, a reflection of the difference between the spectral measure of periodic and random harmonic systems and, in Sec. 8, we give an explicit proof that the spectrum of a simple harmonic chain is absolutely continuous.[2]

Finally in Sec. 9 we discuss briefly the relation between the heat flow in Rubin's and Lebowitz's model. We also discuss there what strengthening of the Furstenberg theorem is needed for obtaining the asymptotic N-dependence of $J(N)$ or $\hat{J}(N)$. Appendices A—C contain some technical details.

2. THE HEAT FLOW IN RUBIN'S MODEL

The first step is to look at a finite analog of the infinite chain. Particles of unit mass are placed at sites $-S$ to 0, from sites 1 to N particles of random mass and from sites $N+1$ to $N+S+2$ particles of unit mass are placed. The random masses are assumed for simplicity to be all greater than one and are identically distributed, independent random variables. At $t=0$ we know the position and momenta of every particle in the chain. The left-hand segment of unit masses is just a chain driven by an external force u_1. Explicitly

$$\eta(t) + \Omega^2\eta(t) = g(t), \qquad (2.1)$$

where

$$\eta(t) = (u_0, u_{-1}, \ldots, u_{-S}),\ \Omega^2 = \Phi_{S+1},$$
$$g(t) = (u_1(t), 0, \ldots, 0).$$

Ω^2 has the spectral representation

$$\Omega^2 = \sum_{a=1}^{S+1} \omega_a^2 \, |\xi_a\rangle\langle\xi_a| \qquad (2.2)$$

where

$$\omega_a^2 = 4\sin^2\left(\frac{a\pi}{S+2}\right) \quad\text{and}\quad \xi_a(j) = \left(\frac{2}{S+2}\right)^{1/2}\sin\left(\frac{ja\pi}{S+2}\right),$$
$$1\le j\le S+1.$$

If

$$\eta(0) = \sum_{a=1}^{S+1} b_a\xi_a, \qquad (2.3)$$
$$\dot{\eta}(0) = \sum_{a=1}^{S+1} v_a\xi_a,$$

then as is known

$$u_0(t) = g_1(t) + \int_0^t A_S(t-s)u_1(s)\,ds, \qquad (2.4)$$

where

$$g_1(t) = \sum_a [(\cos\omega_a t)b_a + \omega_a^{-1}(\sin\omega_a t)v_a]\xi_a(1),$$
$$A_a(t) = \sum_a \omega_a^{-1}(\sin\omega_a t)\xi_a^2(1). \qquad (2.5)$$

The initial energy of the particles $-S,\ldots,0$ is $\frac{1}{2}\sum_a(v_a^2+\omega_a^2 b_a^2)$. The b_a and v_a have a Boltzmann (Gaussian) distribution at temperature T_L and so we can compute the statistical properties of g. When $S\to\infty$

$$u_0(t) = g_1(t) + \int_0^t A(t-s)\,u_1(s)\,ds, \tag{2.6}$$

$$A(t) = (2/\pi)\int_0^\pi \omega^{-1}(k)\,(\sin^2 k)\,\sin[t\,\omega(k)]\,dk, \tag{2.7}$$

$$\omega(k) = 2\sin(k/2),$$

$$\langle g_1(t)\rangle = 0,$$

$$\langle g_1(t)\,g_1(t+s)\rangle = (T_L/\pi)\int_0^\pi \omega^{-2}(k)\,(\sin^2 k)\cos[\omega(k)s]\,dk, \tag{2.8}$$

where we have set Boltzmann's constant equal to one. Similarly,

$$u_{N+1}(t) = g_N(t) + \int_0^\infty A(t-s)\,u_N(s)\,ds. \tag{2.9}$$

g_N has identical properties to g_1 when T_R replaces T_L.

Here $g_1(t)$ and $g_N(t)$ are to be interpreted as "independent Gaussian random variables" with mean zero and covariances given by (2.8) for g_1 and a corresponding expression with T_R replacing T_L for g_N. We set $\mathbf{g} = (g_1, 0, \ldots, g_N)$.

Using (2.6) and (2.9) we have a closed set of equations for the particles 1 to N:

$$m_1\ddot{u}_1 + 2u_1 - u_2 = u_0 = g_1 + A^* u_1,$$

$$m_2\ddot{u}_2 + 2u_2 - u_3 - u_1 = 0$$

$$\vdots$$

$$m_{N-1}\ddot{u}_{N-1} + 2u_{N-1} - u_{N-2} - u_N = 0,$$

$$m_N\ddot{u}_N + 2u_N - u_{N-1} = u_{N+1} = g_N + A^* u_N. \tag{2.10}$$

The same set of equations was obtained by Magalinskii.[10]

In the Fourier representation (2.10) takes the form

$$[\Phi_N - M_N\omega^2 - A(\omega)\,U_N]\;\mathbf{u}(\omega) = \mathbf{g}(\omega) \tag{2.11}$$

where U_N is the N by N diagonal matrix with entries $(U_N)_{ii} = (\delta_{i1} + \delta_{iN})$. In Appendix A we show that the $N\times N$ matrix $Y_N(\omega) = \Phi_N - M_N\omega^2 - A(\omega)\,U_N$ is nonsingular for all real values of ω except $\omega^2 = 0$ and 4. These singularities are integrable. We specify that $\mathbf{u}(t)$ and $\mathbf{g}(t)$ vanish when $t<0$. It is important that we include this carefully in the calculation. $\mathbf{u}(t)$ is the sum of a particular solution of the inhomogeneous equation (2.10) and a general solution of the homogeneous equation which matches the initial values of $\mathbf{u}$ and $\dot{\mathbf{u}}$. In Appendix B we show that the general solution decays at least as fast as $t^{-1/2}$. This represents the diffusion of energy into the chain and so initial data on the N particles does not contribute to the steady state heat flow. A solution of the inhomogeneous equation is

$$\mathbf{u}(t) = (2\pi)^{-1}\int_{-\infty}^{\infty} \exp(i\omega t)\,Y_N(\omega)^{-1}\,\mathbf{g}(\omega)\,d\omega. \tag{2.12}$$

The value of $A(\omega)$ and the statistical properties of $\mathbf{g}(\omega)$ are easily computed as continuations of the associated Laplace transforms. We obtain

$$A(\omega) = \tfrac{1}{2}[2 - \omega^2 - i\omega(4-\omega^2)^{1/2}]$$

$$\langle \mathbf{g}(\omega)\rangle = 0,$$

$$\langle g_i(\omega)\,g_j(\sigma)\rangle = (T_L\delta_{i1} + T_R\delta_{iN})\,\delta_{ij}\,[\bar{g}(\omega) + \bar{g}(\sigma)]\,\hat{\theta}(\sigma+\omega),$$

$$\bar{g}(\omega) = \tfrac{1}{2}[(4-\omega^2)^{1/2} - i\omega],$$

$$\hat{\theta}(\omega) = \lim_{\epsilon\to 0_+}(\epsilon + i\omega)^{-1}. \tag{2.13}$$

We choose the branch of $z^{1/2}$ which is cut from 0 to ∞ along the positive real axis. For later use we note that when ω is real

$$\begin{aligned} &A(\omega) - A(-\omega) = -i\omega\,(4-\omega^2)_+^{1/2}, && |\omega| < 2,\\ &A(\omega) - A(-\omega) = 0, && |\omega| \geq 2,\\ &\bar{g}(\omega) + \bar{g}(-\omega) = (4-\omega^2)_+^{1/2}, && |\omega| < 2,\\ &\bar{g}(\omega) + \bar{g}(-\omega) = 0, && |\omega| \geq 2, \end{aligned} \tag{2.14}$$

where $(x)_+^{1/2}$ is the positive square root of $x \geq 0$.

We can now compute the average heat flow past particle 1 at time t:

$$\begin{aligned} \hat{J}(N,t) &= \langle \dot{u}_1(u_1 - u_0)\rangle \\ &= \langle \dot{u}_1(u_1 - g_1 - A^* u_1)\rangle \end{aligned} \tag{2.15}$$

Substituting from (2.12) and using (2.13) and (2.14) gives the following expression for the stationary heat flux in Rubin's model,

$$\hat{J}(N) \equiv \lim_{t\to\infty}\hat{J}(N,t) = \pi\,(T_L - T_R)\int_0^2 \omega^2\,(4-\omega^2)\,|\Delta_N(\omega)|^{-2}\,d\omega, \tag{2.16}$$

$$\Delta_N(\omega) = \det[Y_N(\omega)].$$

In deriving this we use the result

$$\lim_{t\to\infty}\lim_{\epsilon\to 0_+}\int_{-\infty}^{\infty} \exp[i(\omega+\sigma)t]\;[\epsilon + i(\omega+\sigma)]^{-1}\;f(\sigma)\,d\sigma = \pi f(-\omega) \tag{2.17}$$

when f is a continuous, integrable function.

It is interesting to note that only the frequencies in the allowed band of the infinite chain which, because we considered only heavy impurities, contains all the characteristic frequencies of the finite chain, contribute to $\hat{J}(N)$. Any solution $\mathbf{u}(t)$ which vanishes when $t<0$ must contain contributions from almost all real frequencies [because $\mathbf{u}(\omega)$ is a nonzero analytic function in the lower half plane and so its boundary values $\mathbf{u}(\omega)$ can only vanish on a set of measure zero]. As time increases however the contribution from frequencies outside the allowed band falls to zero.

We can relate $\hat{J}(N)$ to the transmission coefficients of the segment $1,\ldots,N$ for plane waves with frequencies from 0 to 2. To calculate the transmission coefficient of an incoming plane wave with frequency ω we only need to find a solution of the equations of motion which to the right of the segment $1,\ldots,N$ is a combination of an incoming and a reflected wave and to the left is a pure outgoing wave, i.e.,

$$\begin{aligned} &u_j(t) = D\,\exp[-i(\omega t + kj)] + R\exp[-i(\omega t - kj)], && j \geq N,\\ &u_j(t) = \exp[-i(\omega t + kj)], && j \leq 0,\\ &\omega = \omega(k) = 2\sin(k/2). \end{aligned} \tag{2.18}$$

Clearly $|D|^{-1}$ is the transmission coefficient $t_N(\omega)$ and the argument of D^{-1} is the phase shift of the plane wave. Also R/D is the reflection coefficient with $|R/D|^2 = 1 - t_N^2$. Using the transfer matrix approach, we find

$$\begin{bmatrix} u_{N+1} \\ u_N \end{bmatrix} = \begin{bmatrix} \exp[-ik(N+1)] & \exp[+ik(N+1)] \\ \exp(-ikN) & \exp(+ikN) \end{bmatrix}\begin{bmatrix} D \\ R \end{bmatrix}$$

$$= T_N T_1 \cdots T_0 \begin{pmatrix} u_0 \\ u_1 \end{pmatrix}, \tag{2.19}$$

$$\begin{bmatrix} u_0 \\ u_{-1} \end{bmatrix} = \begin{bmatrix} 1 \\ \exp(+ik) \end{bmatrix}$$

where T_j is the "transfer matrix,"

$$T_j = \begin{bmatrix} 2 - m_j \omega^2 & -1 \\ 1 & 0 \end{bmatrix}. \tag{2.20}$$

We find that

$$\begin{aligned} |D(k)|^{-1} &= |2 \sin k| \; |K_{1,N} - \exp(-ik)(K_{2,N} + K_{1,N-1}) \\ &\quad + \exp(-2ik) K_{2,N-1}|^{-1} \\ &= t_N(\omega), \end{aligned} \tag{2.21}$$

where $K_{i,j}(\omega^2)$ is defined in (1.8).

The expression for $t_N(\omega)$ is related simply to $\Delta_N(\omega)$ in (2.16),

$$\Delta_N(\omega) = K_{1,N} - \exp(-ik)\,(K_{2,N} + K_{1,N-1}) + \exp(-2ik) K_{2,N-1}. \tag{2.22}$$

The final result is then

$$\hat{J}(N) = (4\pi)^{-1} (T_L - T_R) \int_0^2 d\omega\, t_N^2(\omega). \tag{2.23}$$

This agrees with the result of Rubin and Greer.[8]

For periodic chains $t_N(\omega)$ approaches, as $N \to \infty$, a finite value different from zero for ω in the spectrum of this chain. This spectrum consists of bands in the interval $\omega \in [0, 2]$. For ω not in the spectrum $t_N(\omega)$ vanishes as $\exp[-N\delta(\omega)]$, where $\delta(\omega) = (\omega - \omega_0)^2$ and ω_0 is band edge nearest ω. Indeed, for $m_i = 1$, for all i, $t_N(\omega) = 1$. Thus $\hat{J}(N)$ will approach a finite positive value as $N \to \infty$ in periodic chains. The situation is quite different in random chains where, as will be shown in the next section, $t_N(\omega)$ goes to zero, exponentially in N for almost all chains.

3. GROWTH OF SOLUTIONS TO THE LATTICE EQUATIONS

We consider a semiinfinite chain with masses m_i, $i \geq 1$. A plane wave solution with frequency ω satisfies the equation [cf (1.3)–(1.5)]

$$(2 - m_N \omega^2) u_N = u_{N+1} + u_{N-1}. \tag{3.1}$$

This is more conveniently written in the transfer matrix notation

$$\begin{aligned} \begin{bmatrix} u_{N+1} \\ u_N \end{bmatrix} &= \begin{bmatrix} 2 - m_N \omega^2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_N \\ u_{N-1} \end{bmatrix} \\ &= T_N(\omega) \begin{bmatrix} u_N \\ u_{N-1} \end{bmatrix}. \end{aligned} \tag{3.2}$$

T_N is in the matrix group $SL(2, R)$. In disordered chains the sequence $\{u_N(\omega)\}$ grows exponentially with N for almost every sequence of masses and almost all initial values of u_0 and u_1. This was first proven by Matsuda and Ishii[5] using a theorem of Furstenberg.[7] Here a brief summary of the theorem and its application is given.

We see from (3.2) that the asymptotic behavior of $u_N(\omega)$ is determined by the behavior of products of the transfer matrices T_j associated with the chain.

Furstenberg's theorem deals with products of matrices in the groups $SL(m, R)$ when the matrices themselves are random variables.

Theorem (Furstenberg,[7] Theorem 8.5): Suppose that G is a subgroup of $SL(m, R)$ such that

(i) G is not compact;

(ii) no subgroup of G with finite index in G is reducible;

(iii) there is a probability measure μ on G.

Then for almost all sequences $\{g_N : N \geq 1\}$ chosen from G we have

$$\lim_{N \to \infty} \frac{1}{N} \log \| g_N \cdots g_1 u \| = \gamma > 0 \tag{3.3}$$

for any nonzero vector u in R^N.

Remark: A real analytic group G is reducible if it has a faithful finite-dimensional continuous representation and if every finite-dimensional continuous representation of G is semisimple, i.e., if G has a representation as linear transformations on a finite-dimensional vector space V the only subspaces of V which are invariant under the action of G are $\{0\}$ and V itself. In our case G is a subgroup of $SL(2, R)$ and so there is always a faithful representation as matrices acting on R^2. The remaining condition must be checked explicitly. γ can be explicitly calculated in terms of certain measures on the projective space P^{m-1}. These measures are determined by μ and the induced action of G on P^{m-1}. In the statement of this theorem almost all is meant in the sense of the standard measure on the product of a countable number of copies of G which can be obtained from the basic measure μ on G.

Matsuda and Ishii have proven the following result. Their argument has been greatly simplified by Yoshioka.[11]

Theorem (Matsuda and Ishii, Theorem 1): If there are at least two different masses present, the subgroup of $SL(2, R)$ generated by the transfer matrices $\left(\begin{smallmatrix} 2 - m_1 \omega^2 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ obeys conditions (i) and (ii) (for $\omega^2 > 0$).

The mass m is a random variable with probability distribution $dp(\cdot)$ and the measure μ on the subgroup is determined by $dp(\cdot)$. The corresponding γ is written as $\gamma(\omega)$ and by (3.3) $\gamma(\omega) > 0$ for $\omega \neq 0$. $\gamma(\omega)$ can be calculated from the following equations, when ω^2 is small, i.e., η defined in (3.4b) is real,[5]

$$\gamma(\omega) = \int_{-\pi/2}^{\pi/2} \log \left| \frac{\cos(\theta + \eta)}{\cos \theta} \right| dG(\theta), \tag{3.4a}$$

where

$$\begin{aligned} 2 \cos \eta &= 2 - \langle m \rangle \omega^2, \\ \langle m \rangle &= \int_0^\infty m\, dp(m). \end{aligned} \tag{3.4b}$$

$dG(\cdot)$ is a probability measure on $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$ which satisfies

$$G(A)=\int_0^\infty G[\Psi(A,m)]\,dp(m) \tag{3.5}$$

for every measurable set A in $(-\frac{1}{2}\pi, \frac{1}{2}\pi)$. Here $\Psi(A,m)=\{\Psi(\theta,m):\theta\in A\}$ and

$$\tan[\Psi(\theta,m)+\eta]=\tan\theta-2\left[\frac{m-\langle m\rangle}{\langle m\rangle}\right]\tan\frac{\eta}{2}\,. \tag{3.6}$$

The integral in (3.4) is absolutely convergent and can be shown to be independent of the particular measure dG used, provided that $dG(\cdot)$ obeys (3.5). These are Eqs. (3.14) and (3.15) in Ref. 5.

We now give a nonperturbative proof of Theorem 2 in Ref. 5. This deals with the low frequency behavior of $\gamma(\omega)$.

Theorem: For small values of ω, $\gamma(\omega)$ is continuous in ω and

$$\lim_{\omega\to 0_+}\frac{\gamma(\omega)}{\omega^2}=\frac{1}{8}\,\frac{\langle m^2\rangle-\langle m\rangle^2}{\langle m\rangle}\,.$$

Proof: We always choose ω so small that η is real. Then the stationary measures $dG(\theta,\eta)$ (explicitly showing the dependence on η) can be chosen to depend continuously on η (see Lemmas 1.2, 2.2, and 2.3 in Ref. 7). Then because the integral defining $\gamma(\omega)$ converges absolutely we can write $\gamma(\omega)$ as the sum of terms similar to

$$\begin{aligned}&A(\omega)=\int_0^{1/2\pi-\epsilon}\log|\cos\theta|\,dG(\theta,\eta)+a(\epsilon,\eta),\\&a(\epsilon,\eta)=\lim_{\delta\to 0_+}\int_{1/2\pi-\epsilon}^{1/2\pi-\delta}\log|\cos\theta|\,dG(\theta,\eta)\end{aligned} \tag{3.7}$$

and absolute convergence also means that $\lim_{\epsilon\to 0_+} a(\epsilon,\eta)=0$. So by choosing η and η' (corresponding to ω and ω') sufficiently close together and also choosing ϵ small enough, we can make $A(\omega)-A(\omega')$ arbitrarily small. Consequently, $\gamma(\omega)$ is a continuous function of ω.

$$\begin{aligned}2\gamma(\omega)&=\int_{-\pi/2}^{\pi/2}\log\left(\frac{\cos^2(\theta+\eta)}{\cos^2\theta}\right)dG(\theta,\eta)\\&=\int_{-\pi/2}^{\pi/2}\log\cos^2[\Psi(\theta,m)+\eta]\,dG[\Psi(\theta,m),\eta]\\&=\int_{-\pi/2}^{\pi/2}\log\cos^2\theta\,dG(\theta,\eta).\end{aligned} \tag{3.8}$$

$\Psi(\theta,m)$ can be calculated from (3.6) to second order in η. Using this gives

$$\begin{aligned}&\log\cos^2[\Psi(\theta,m)+\eta]=\log\cos^2\theta-a\sin 2\theta\\&\qquad\qquad+a^2f(\theta)+O(a^3),\\&f(\theta)=-\sin\theta\cos^3\theta.\\&a=2\left(\frac{m-\langle m\rangle}{\langle m\rangle}\right)\tan(\tfrac{1}{2}\eta).\end{aligned} \tag{3.9}$$

Since $\gamma(\omega)$ is independent of m, (3.8) is not changed if we integrate it over $dp(m)$. We obtain

$$\begin{aligned}2\gamma(\omega)=&\int_0^\infty dp(m)\int_{-\pi/2}^{\pi/2}-a\sin^2\Phi(\theta,m)\,dG(\theta,\eta)\\&+\int_0^\infty dp(m)\int_{-\pi/2}^{\pi/2}a^2f[\Phi(\theta,m)]\,dG(\theta,\eta)\\&+O(\eta^3).\end{aligned} \tag{3.10}$$

$\Phi(\theta,m)$ is the inverse of Ψ and expanding in powers of a gives

$$\begin{aligned}\sin 2\Phi(\theta,m)=&\sin 2(\theta+\eta)+2a\cos^2(\theta+\eta)\cos 2(\theta+\eta)\\&+O(a^2).\end{aligned} \tag{3.11}$$

Then to second order in a

$$\begin{aligned}2\gamma(\omega)=&\int dp(m)\cdot a\int\{\sin 2\theta+2\cos 2\theta[\eta+a\cos^2\theta]\}\\&dG(\theta,\eta)\\&-\int dp(m)\,a^2\int\cos^2\theta\cos 2\theta\,dG(\theta,\eta).\end{aligned} \tag{3.12}$$

Finally, we find

$$\begin{aligned}\lim_{\omega\to 0_+}[\gamma(\omega)/\omega^2]=&\tfrac{1}{2}\cdot\langle m\rangle\int dp(m)\left(\frac{m-\langle m\rangle}{\langle m\rangle}\right)^2\\&\times\int_{-\pi/2}^{\pi/2}\cos 2\theta\cos^2\theta\,dG(\theta,0).\end{aligned} \tag{3.13}$$

We can choose $dG(\theta,0)$ to be $\pi^{-1}\,d\theta$ and so finally get the result of the theorem.

We can use this result to connect Sec. 2 with earlier work by Rubin[12] on the transmission of plane waves through random chains. We can rewrite (2.21) as

$$2|\sin k|\,t_N^{-1}(\omega)=\left|(1,-e^{-ik})\cdot T_1\cdots T_N\begin{bmatrix}1\\e^{ik}\end{bmatrix}\right|. \tag{3.14}$$

Furstenberg[7] has not only shown that the norm of $T_1\cdots T_N u$ grows exponentially with N but also that the vector converges to a fixed direction (depending on u). Consequently, we can use (3.13) to calculate

$$\lim_{N\to\infty}[-(1/N)\log t_N^2(\omega)]=2\gamma(\omega). \tag{3.15}$$

This relationship was previously proven by Minami and Hori[13] using a different method. If each mass can take the values m and $m(1+Q)$ with probabilities q and p, we see from (2.13) that

$$\lim_{\omega\to 0_+}\frac{2\gamma(\omega)}{\omega^2}=\left(\frac{m}{4}\right)\frac{pqQ^2}{1+pQ}\,. \tag{3.15'}$$

Rubin's normalization of frequency is equivalent to taking $m=4$. The mean spacing between the heavy particles is $\sum_{r=0}^\infty(r+1)q^rp=p^{-1}$. In Rubin's notation this is C^{-1} and so for small ω, $2\gamma(\omega)$ behaves as $C(1-C)Q^2(1+QC)^{-1}\omega^2$. This agrees with Rubin's result (Eq. 4.4, Ref. 12]. (Note, however, that Rubin's N differ from ours by a factor of C.) Because $t_N^2(\omega)$ decays exponentially with N for almost all random chains, we can use the argument of Ref. 2 to conclude that $\langle\hat{J}(N)\rangle\to 0$ as $N\to\infty$.

Sulem and Frisch[14] have recently examined the transmission of light through a one-dimensional system in which the refractive index takes different constant values on successive intervals. These values are independent, identically distributed random variables. They used an argument based on the random ergodic theorem to show that almost all such systems are totally reflecting. The method of Ref. 12 shows that Furstenberg's theorem applies to their model.[15]

4. CALCULATION OF HEAT FLOW IN SPECIAL CASES

The heat flow through an arbitrary chain of masses given by (1.7) can only rarely be explicitly calculated. In Ref. 2, Casher and Lebowitz checked that it agrees with Ref. 1 for the infinite isotropic chain. Two more examples are given here; the infinite diatomic chain and the infinite isotropic chain in which a single impurity is imbedded. The spectrum of both systems contains an

absolutely continuous part and so the limiting value of the heat flux, J, is nonzero (Ref. 2 and Sec. 8).

A. The infinite diatomic chain

This is an infinite periodic chain in which m is m_1 when j is odd and m_2 when j is even. From Ref. 2 the heat flow through an infinite periodic chain whose unit cell contains the masses $m_1 \cdots m_c$ is just

$$J = \pi^{-1} m_1 m_c \lambda \, \Delta T \int d\omega \, |\omega \sin q| \; |c(\omega)|,$$
$$c(\omega)^{-1} = (1 + \lambda^2 \omega^2 m_1 m_c)\,(m_c K_{1,c-1} + m_1 K_{2,c}). \tag{4.1}$$

Only frequencies in the allowed bands will contribute to J. These are just the values of ω^2 for which

$$|K_{1,c}(\omega) - K_{2,c-1}(\omega)| \leq 2. \tag{4.2}$$

For a wave vector q they are the solutions of

$$K_{1,c}(\omega) - K_{2,c-1}(\omega) = 2\cos q$$

as q ranges from 0 to π. For the diatomic chain these are

$$\omega^2 = (m_1^{-1} + m_2^{-1})\,(1 \pm \phi(q)).$$
$$\phi(q)^2 = 1 - \mu(1 - \cos q), \tag{4.3}$$
$$\mu = 2m_1 m_2 (m_1 + m_2)^{-2}.$$

The acoustical branch of the spectrum is given by the negative sign and the optical branch by the positive sign. Then (4.1) reduces to

$$J = 2(1 + M\lambda^2)^{-1} \; \pi^{-1} \lambda \, \Delta T \int_0^{\pi} dq \sin^2 q \, |\phi(q)|^{-2}$$
$$\times [(1 + M\lambda^2)^2 - (M\lambda^2 \phi(q)^2)]^{-1},$$
$$M = m_1 + m_2. \tag{4.4}$$

A partial fraction expansion of the integrand gives

$$\frac{1}{x^2} + \left(\frac{2\mu - 1}{(1+x)^2}\right)(\phi^2)^{-1}$$
$$- \left(\frac{(2\mu x^2 + 2x + 1)(2x+1)}{x^2(1+x)^2}\right)[(1+x)^2 - x^2\phi^2]^{-1}.$$

Using $\phi^2 = \cos^2(\tfrac{1}{2}q) + (1 - 2\mu)\sin^2(\tfrac{1}{2}q)$, each term becomes a simple trigonometric integral and we get

$$J = \frac{8\lambda^{-3}\Delta T}{M^2(1+x)(1-\delta^2)} \{(1+x)^2 - x^2\delta - (1+2x)^{1/2}[1$$
$$+ 2x + x^2(1-\delta^2)]\} \tag{4.5}$$

where $x = M\lambda^2$ and we have set $|m_2 - m_1| = M\delta$. This agrees with (3.13) in Ref. 1c when $m_1 = m_2 = m$. Near $\delta = 0$, J is a decreasing function of δ so that starting from a monatomic chain and keeping M fixed J will initially decrease as $|m_1 - m_2|$ is increased.

B. A single impurity in an infinite isotropic chain

When a finite number of impurities are added to an infinite isotropic chain the spectrum of the new chain still contains an absolutely continuous piece. At most a finite number of isolated eigenvalues will be added to the original spectrum. These eigenvalues correspond when the impurities are light to highly localized normal modes. Using the techniques of Ref. 2 it is clear that they will not contribute to the heat flux through the infinite chain. In principle we can use (1.7) to calculate the limiting value, J, for any set of impurities. There seems to be a simple expression for J only for a single impurity and even this cannot be evaluated in terms of elementary functions. It was found more convenient to use techniques similar to those of Sec. 2 on the Langevin equation approach used by Ishii.[16] [The results will be exactly equivalent to those obtained from (1.7).] The final result, for a single impurity of mass m in the middle of a chain of unit masses, is (for details see Appendix D)

$$J = (2\pi)^{-1} \lambda \, \Delta T \int_0^{\pi} d\theta \sin^2\theta \,(1 + u^2)\,[(1+u^2)^2 + u^2 v^2]^{-1},$$
$$u^2 = \lambda^2 \omega^2, \quad v = (m-1)\omega^2, \quad \omega^2 = 2(1 - \cos\theta). \tag{4.6}$$

When $m = 1$, this agrees with (3.13) in Ref. 1c.

5. HEAT FLOW IN A WEAKLY COUPLED CHAIN

We give here an exact derivation of the asymptotic behavior of the heat flow in Lebowitz's model when the coupling constant λ of the chain to the heat baths is small. This is of interest because in the limit $\lambda \to 0$ the dependence of the heat flux $J(N, \lambda)$ [where we have indicated the explicit dependence of $J(N)$ on λ] on the amplitude of the normal modes at the ends of the chain becomes transparent. For small λ the integrand in (1.7) is large when $K_{1,N}(\omega^2) = 0$, i.e., at the normal mode frequencies of the chain. We will only treat those chains whose normal modes are distinct. This is not an important restriction (see Ref. 17 for a discussion of this point). Our theorem is also Theorem 7 in Ref. 5.

Theorem: Consider a chain of masses $\{m_i : i = 1, \cdots, N\}$ which has distinct normal mode frequencies $\{\omega_i^2 : i = 1 \cdots N\}$ and corresponding normal modes $\{\mathbf{u}_i : i = 1 \cdots N\}$. $\mathbf{u}_i$ is normalized by $\sum_{j=1}^{N} m_j u_i(j)^2 = 1$. Then

$$\lim_{\lambda \to 0} \lambda^{-1} J(N, \lambda) = \pi^{-1} m_1 m_N \Delta T \sum_{i=1}^{N} \frac{m_1 u_i(1)^2 m_N u_i(N)^2}{m_1 u_i(1)^2 + m_N u_i(N)^2}. \tag{5.1}$$

Proof: If ω_i^2 is a simple zero of $K_{1,N}(\omega^2)$, then when ω^2 is near ω_i^2

$$K_{1,N}(\omega^2) = (\omega^2 - \omega_i^2) K'_{1,N}(\omega_i^2) + O(\omega^2 - \omega_i^2)^2 \tag{5.2}$$

where the prime indicates derivative. Hence the contribution of ω_i^2 to (1.7) is

$$J^{(i)}(N, \lambda) = m_1 m_N \, \Delta T \, |K'_{1,N}(\omega_i^2)| \, \tau(\omega_i^2)^{-1/2}, \tag{5.3}$$

where

$$\tau(\omega^2) = 2m_1 m_N + m_1^2 K_{2,N}^2(\omega^2) + m_N^2 \, K_{1,N-1}^2(\omega^2).$$

Using the identity

$$K_{1,N}(\omega^2) K_{2,N-1}(\omega^2) - K_{1,N-1}(\omega^2) K_{2,N}(\omega^2) = -1, \tag{5.4}$$

(5.3) reduces to

$$J^{(i)}(N, \lambda) = m_1 m_N \, \Delta T \, |K'_{1,N}(\omega_i^2)|^{-1} \; |K_{1,N-1}(\omega_i)|$$
$$[m_N K_{1,N-1}^2 + m_1]^{-1}. \tag{5.5}$$

The normal mode associated with ω_i^2 is just $\mathbf{u}_i$ and

$$u_i(j) = K_{1,j-1}(\omega_i^2) N(\omega_i)^{-1},$$
$$N(\omega_i)^2 = \sum_{j=1}^{N} m_j K_{1,j-1}^2(\omega_i). \tag{5.6}$$

Considering the equations

$$(2-m_{j+1}\omega^2)K_{1j}=K_{1j+1}+K_{ij-1},$$
$$-m_{j+1}K_{1j}+(2-m_{j+1}\omega^2)K'_{1j}=K'_{1j+1}+K'_{1j-1} \tag{5.7}$$

and multiplying the first by K'_{1j} and the second by K_{1j} and substracting, we get

$$m_{j+1}K^2_{1j}=(K_{1j-1}K'_{1j}-K'_{1j-1}K_{1j})-(K_{1j}K'_{1j+1}-K'_{1j}K_{1j+1})$$
$$=\phi_j-\phi_{j+1}. \tag{5.8}$$

Thus

$$\sum_{j=0}^{N-1} m_{j+1}K^2_{1j}=\phi_0-\phi_{N+1},$$

$$\phi_0=K_{1,-1}-K'_{10}-K'_{11}K_{10}=0,$$

and if $\omega^2=\omega_i^2$, where $K_{1,N}(\omega_i^2)=0$, then

$$\sum_{j=0}^{N-1} m_{n+1}K^2_{1j}(\omega_i^2)=N(\omega_i^2)=K'_{1N}(\omega_i^2)K_{1N+1}(\omega_i^2). \tag{5.9}$$

So (5.5) reduces to the term in (5.1) associated with ω_i^2 and the proof is complete.

Matsuda and Ishii have argued in Ref. 5 that this supports the conjecture that $\langle J(N,\lambda)\rangle$ decreases as $N^{-3/2}$. We want to point out, however, that even if one could establish that, for random chains, the right side of (5.1) behaves as $N^{-3/2}$ when $N\to\infty$ this would not necessarily tell us anything about the behavior of $J(N,\lambda)$ as $N\to\infty$ for any fixed $\lambda>0$. What (5.1) gives is the large N behavior of $\lim_{\lambda\to 0}\lambda^{-1}J(N,\lambda)$ and this need not be the same as the large N behavior of $\lambda^{-1}J(N,\lambda)$ for $\lambda>0$. They will agree for periodic chains where $\lambda^{-1}J(N,\lambda)$ approaches a finite nonzero limit as $N\to\infty$ for any $\lambda>0$. We surmise that for a system obeying Fourier's law the asymptotic form of the heat flux may be of the form

$$J(N,\lambda)\sim\lambda(\Delta T)/(1+\gamma\lambda N)^{-1}, \tag{5.10}$$

where γ is related to the resistance to heat flow in the interior of the system, e.g., the degree of anharmonicity in a anharmonic crystal, or the "degree of disorder" $\langle(m-\langle m\rangle)^2\rangle$ in a random crystal if indeed such a system obeys Fourier's law. If this surmise is right then the two asymptotic behaviors will not be the same. This surmise is based (or strengthened) by the behavior of the heat flow in a simple stochastic model system discussed in the next section.

6. RANDOM REFLECTION MODEL

This is a simple system which transports energy and has a Fourier law behavior. It is a variation of one originally considered by Lebowitz and Frisch.[18] It is a dilute gas of noninteracting particles which move linearly along a cylinder. At either end of the cylinder is a heat bath and barriers are placed at random positions along the cylinder. When a gas particle meets a barrier it will either pass through without changing its velocity or it will be reflected with its velocity exactly reversed. The probability of reflection is r and of transmission $1-r$. At each end it can be directly reflected with probability $1-\lambda$ or with probability λ it is reflected back with a random velocity. This random velocity is independent of the incident velocity and has a Maxwellian distribution characterized by the temperature of the heat baths. These are T_0 on the left and T_1 on the right ($T_0>T_1$). Thus λ plays the role of coupling to the heat baths as before, $0\le\lambda\le 1$. Since the total number of particles is constant, the total number flowing to the right at any point will, in a steady state exactly balance the total number flowing to the left. The first group will presumably be more energetic and so energy will be carried along the cylinder.

We look at the steady state situation. There are N barriers and between barriers i and $i+1$, the number of particles in a unit volume with velocities between v and $v+dv$ is $f_i(v)\,dv$.

Let

$$f_i^+(v)=f_i(v), \qquad v>0,$$
$$f_i^-(v)=f_i(-v), \qquad v>0, \tag{6.1}$$

be the densities for those particles flowing to the right and those flowing to the left. $f_0^\pm$ and $f_N^\pm$ are the corresponding densities for the particles between the heat baths and the first and last barriers. At each barrier the net flux of particles with velocities near v must be zero. So

$$f_i^+(v)=rf_i^-(v)+(1-r)f_{i-1}^+(v), \qquad N\ge i>0.$$
$$f_i^-(v)=rf_i^+(v)+(1-r)f_{i+1}^-(v), \qquad N>i\ge 0. \tag{6.2}$$

At the left the flux of particles incident on the heat bath with velocities near v is $vf_0^-(v)\,dv$. This is redistributed by direct and diffuse reflection so

$$vf_0^+(v)=(1-\lambda)\,vf_0^-(v)+\lambda\,v\,g_0(v)\int_0^\infty uf_0^-(u)\,du. \tag{6.3}$$

$g_0(v)$ is proportional to the Maxwellian distribution of particles in the heat bath at temperature T_0. It is normalized so as to conserve the total flux striking the edge of the cylinder. So $g_0(v)=\beta_0 m\exp(-\beta_0 mv^2)$, $\beta_0=(kT_0)^{-1}$. g_1 is defined similarly. So

$$f_0^+(v)=(1-\lambda)f_0^-(v)+\lambda\,g_0(v)\int_0^\infty uf_0^-(u)\,du \tag{6.4}$$

and

$$f_N^-(v)=(1-\lambda)f_N^+(v)+(\lambda)\,g_1(v)\int_0^\infty(u)\,du. \tag{6.5}$$

The solution of these equations, for $0\le i\le N$, is

$$f_i^+=\nu[g_0+C_i(g_1-g_0)],$$
$$f_i^-=\nu[g_1+C_{N-i}(g_0-g_1)], \tag{6.6}$$

where $C_j=\alpha+j\beta$ with,

$$\alpha=(1-\lambda)(1-r)/[(2+\lambda)(1-r)+r\lambda N],$$
$$\beta=\lambda r/[(2-\lambda)(1-r)+r\lambda N] \tag{6.7}$$

and ν is the total flux of particles flowing in either direction, i.e., $\int_0^\infty vf_i^+(v)\,dv=\int_0^\infty vf_i^-(v)\,dv=\nu$ for $N\ge i\ge 0$. The next flux of energy from left to right is

$$\int_0^\infty \tfrac{1}{2}m\,v^3[f_i^+(v)-f_i^-(v)]\,dv=\frac{\nu\lambda(1-r)}{(2-\lambda)(1-r)+\lambda rN}$$
$$\times k(T_0-T_1)=\bar{J}(N,\lambda). \tag{6.8}$$

The heat flux $J(N,\lambda)$ has the form conjectured in (5.10). It depends only on the number of barriers present and does not depend at all on their spacing (because there is no attenuation between adjacent barriers). In particular, it does not depend on the length of the cylinder. However, if we suppose that in a cylinder of

length L the number of barriers present is randomly distributed with a Poisson distribution whose mean is ρL the average heat flow will be

$$\langle J(L,\lambda)\rangle = \sum_{n=0}^{\infty} \frac{(\rho L)^n}{n!} \exp(-\rho L)\bar{J}(n,\lambda). \tag{6.9}$$

In the limit

$$\lim_{L\to\infty} \left[\frac{L\langle J(L,\lambda)\rangle}{T_0 - T_1}\right] = \nu k(1-r)/r\rho, \tag{6.10}$$

is independent of λ, Eq. (6.10) depends on the asymptotic expansion

$$\sum_{n=0}^{\infty} \frac{t^n}{n!}\frac{e^{-t}}{n+b} = \frac{1}{t} + 0\left(\frac{1}{t^2}\right) \tag{6.11}$$

as $t\to\infty$, for fixed positive b.

Exactly the same result holds if the barriers are placed with a constant density ρ along the cylinder. So we can say that the thermal conductivity of this model is

$$K = \nu k(1-r)\cdot(r\rho)^{-1}. \tag{6.12}$$

Decreasing the barrier spacing means that ρ increases and then K will tend to zero.

7. A SIMPLE TWO-DIMENSIONAL MODEL

This is a cylindrical system in which the masses in each column are identical although from column to column the mass may vary randomly. The cylindrical analog of Lebowitz's model has been examined by Nakazawa.[4] We will combine the method of Sec. 2 with his method to examine the cylindrical analog of Rubin's model. Each column of the cylinder contains p masses and the displacement of the particle in the ith column and the ath row is x_{ia}. Its mass is m_i. When $-S \le i \le 0$ and $N+1 \le i \le N+S+2$, $m_i = 1$; when $1 \le i \le N$ the m_i are independent, identically distributed random variables. The particle at site (i,a) is coupled by harmonic forces of unit strength to those at $(i-1,a)$ and $(i+1,a)$ and by forces of strength μ to those at $(i,a-1)$ and $(i,a+1)$. The equation of motion of this system, with $x_{i,p+1} = x_{i,1}$, is

$$m_i\ddot{x}_{i,a} + (2x_{i,a} - x_{i-1,a} - x_{i+1,a}) + \mu(2x_{i,a} - x_{i,a-1} - x_{i,a+1}) = 0. \tag{7.1}$$

We exploit cylindrical symmetry by forming the sums

$$x_i(\theta_l) = p^{-1/2}\sum_{a=1}^{p} x_{i,a}\exp(i\,\theta_l a) \tag{7.2}$$

where $1 \le l \le p$, $\theta_l = 2\pi l/p$. Then (7.1) becomes

$$M\ddot{\mathbf{x}}(\theta_l) + \Omega^2(\theta_l)\mathbf{x}(\theta_l) = 0, \tag{7.3}$$

where M is the diagonal matrix with entries m_i, $\Omega^2(\theta)$ the tridiagonal matrix with diagonal entries $2 + 4\mu\sin^2\frac{1}{2}\theta$ and off diagonal entries -1 and $\mathbf{x}(\theta_l)$ is the column vector whose ith entry is $x_i(\theta_l)$, $-S \le i \le N+S+2$. These are identical to the equations of motion (1.2) of a linear chain with different coupling strengths between adjacent particles.

We have the inversion formula

$$x_{i,a} = p^{-1/2}\sum_{l=1}^{p} x_i(\theta_l)\eta^{-la}. \tag{7.4}$$

In terms of these coordinates the energy in the left-hand piece of the cylinder is

$$\frac{1}{2}\sum_{l=1}^{p}\{\dot{\mathbf{x}}'(\theta_l)\cdot\dot{\mathbf{x}}'(\theta_l) + \mathbf{x}'(\theta_l)\cdot\Omega^2(\theta_l)\mathbf{x}'(\theta_l)\}. \tag{7.5}$$

Here $\mathbf{x}'(\theta_l) = \{x_i(\theta_l) : -S \le i \le 0\}$.

We can repeat the analysis of Sec. 2 to obtain a closed set of equations of motion for the columns 1 through N. If the left-and right-hand pieces are initially in thermal equilibrium at temperatures T_L and T_R we find, after letting $S\to\infty$ and then $p\to\infty$, that the analog of (2.6) and (2.7) are

$$x_0(\theta,t) = g_1(\theta,t) + \int_0^T A(t-s)\,x_1(\theta,s)\,ds,$$

$$A(t) = \frac{2}{\pi}\int_0^{\pi} \omega(k)^{-1}\sin[t\omega(k)]\sin^2 k\,dk,$$

$$\langle g_1(t)\rangle = 0,$$

$$\langle g_1(t)\,g_1(t+s)\rangle = \pi^{-1}k\;T_L\int_0^{\pi}\omega(k)^2\sin^2 k\cos[s\omega(k)]\,dk,$$

$$\omega(k)^2 = 4(\sin^2\tfrac{1}{2}k + \mu\sin^2\tfrac{1}{2}\theta) \tag{7.6}$$

(in the limit $p\to\infty$, θ_l becomes a continuous parameter θ ranging from 0 to 2π). In deriving these we note that the $S\times S$ matrix $\Omega^2(\theta)$ has eigenvalues $4(\sin^2\frac{1}{2}\phi_j + \mu\sin^2\frac{1}{2}\theta]$ with $\phi_j = j\pi/S+1$ $(1 \le j \le S)$ and eigenvectors $\xi_j = [2/(S+1)]^{1/2}(\sin\phi_j, \cdots, \sin S\phi_j)$.

The average energy flowing past the particle at site $(1,a)$ (from left to right) is

$$\hat{J}(N,t) = \langle\dot{x}_{1,a}(x_{1,a} - x_{0,a})\rangle = p^{-1}\sum_{l=1}^{p}\langle\dot{x}_1(\theta_l) - x_0(\theta_l)]\rangle. \tag{7.7}$$

When $p\to\infty$, $\hat{J}(N,t)$ becomes

$$\hat{J}(N,t) = (2\pi)^{-1}\int_0^{2\pi}\dot{x}_1(\theta)[x_1(\theta) - x_0(\theta)]\,d\theta. \tag{7.8}$$

This is just a superposition of currents from harmonic chains with coupling matrices $\Omega^2(\theta)$ so that repeating the analysis of Sec. 2 we obtain, when $t\to\infty$ (setting Boltzmann's constant equal to unity),

$$\hat{J}(N) = (4\pi)^{-1}\Delta T\int_0^{\pi}d\theta\int_0^{\infty}d\omega\,|\det Y(\omega,\theta)|^{-2} \times i\omega[\hat{A}(\omega) - \hat{A}(-\omega)]\,[\hat{g}(\omega) + \hat{g}(-\omega)]. \tag{7.9}$$

$\hat{A}$ and $\hat{g}$ are the Fourier transforms of A and g and are obtained by analytic continuation of their Laplace transforms. $Y(\omega,\theta)$ is the $N\times N$ matrix $\Omega^2(\theta) - \omega^2 M - \hat{A}(\omega)L$. We find that when $4\mu\sin^2\frac{1}{2}\theta \le \omega^2 \le 4 + 4\mu\sin^2\frac{1}{2}\theta$, then

$$\hat{A}(\omega) - \hat{A}(-\omega) = -i(\omega^2 - 4\mu\sin^2\tfrac{1}{2}\theta)^{1/2}(4 + 4\mu\sin^2\tfrac{1}{2}\theta - \omega^2)^{1/2},$$

$$\hat{g}(\omega) + \hat{g}(-\omega) = \omega^{-1}(\omega^2 - 4\mu\sin^2\tfrac{1}{2}\theta)^{1/2}(4 + 4\mu\sin^2\tfrac{1}{2}\theta - \omega^2)^{1/2} \tag{7.10}$$

and that they are zero otherwise. Thus, calling R the range of the ω integration, we obtain

$$\hat{J}(N) = \Delta T\int_0^{\pi}d\theta\int_R d\omega\,|\det Y(\omega,\theta)|^2(\omega^2 - 4\mu\sin^2\tfrac{1}{2}\theta) \times(4 + 4\mu\sin^2\tfrac{1}{2}\theta - \omega^2). \tag{7.11}$$

We simplify this by introducing the parametrization $\omega^2 = 4\sin^2\frac{1}{2}k_1 + 4\mu\sin^2\frac{1}{2}k_2$ which is valid for all ω^2 in the range of integration of (7.11); then

$$\hat{J}(N)=\Delta T\int_0^\pi\int_0^\pi dk_1\,dk_2\cos\tfrac{1}{2}k_1\left(\frac{\sin^2\frac{1}{2}k_1}{\sin^2\frac{1}{2}k_1+\mu\sin^2\frac{1}{2}k_2}\right)^{1/2}$$

$$|f(k_1,k_2)|^2,$$

$$f(k_1,k_2)=2\sin k_1\cdot(\det Z_N(k_1,k_2))^{-1},$$

$$Z_N=\Omega^2(k_2)-M\omega^2-\exp(ik_1)L. \tag{7.12}$$

L is the diagonal matrix with entries $L_{ii}=(\delta_{i1}+\delta_{iN})$ and in this representation of ω^2, $\hat{A}(\omega)=\exp(ik_1)$.

We can relate $\hat{J}(N)$ to the transmission properties of the columns 1 to N. The incident and transmitted plane waves are

$$\begin{aligned}x_{j,a}&=\exp[-i(\omega t+k_1 j+k_2 a)]\\ &\qquad,\ j\le 1,\\ x_{j,a}&=D\exp[-i(\omega t+k_1 j+k_2 a)]\\ &\quad+R\exp[-i(\omega t-k_1 j+k_2 a)]\\ &\qquad,\ j\ge N.\end{aligned} \tag{7.13}$$

Using the equations of motion (7.3) for a plane wave with frequency ω, we find

$$\begin{bmatrix}x_{j+1}(\theta)\\ x_j(\theta)\end{bmatrix}=\begin{bmatrix}2+4\mu\sin^2\frac{1}{2}\theta-m_j\omega^2 & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}x_j(\theta)\\ x_{j-1}(\theta)\end{bmatrix}. \tag{7.14}$$

Applying (7.13) yields

$$\begin{bmatrix}\exp(-ik_1(N+1) & \exp[ik_1(N+1)\\ \exp(-ik_1N) & \exp(ik_1N)\end{bmatrix}\begin{bmatrix}D\\ R\end{bmatrix}$$

$$=T_N\cdots T_1(k_1,k_2)\begin{bmatrix}\exp(-ik_1)\\ 1\end{bmatrix}, \tag{7.15}$$

where

$$T_j(k_1,k_2)=\begin{bmatrix}2+4\mu\sin^2\frac{1}{2}k_2-m_j\omega^2 & -1\\ 1 & 0\end{bmatrix}.$$

This yields for the transmission coefficient, t_N, in analogy with (2.21),

$$|D_N(k_1,k_2)|^{-1}=|f(k_1,k_2)|=|t_N(k_1,k_2)|. \tag{7.16}$$

Substituting in (7.12) gives

$$\hat{J}(N)=k\Delta T\int_0^\pi\int_0^\pi dk_1\,dk_2\cos\tfrac{1}{2}k_1\times\left[\frac{\sin^2\frac{1}{2}k_1}{\sin^2\frac{1}{2}k_1+\mu\sin^2\frac{1}{2}k_2}\right]^{1/2}|t_N(k_1,k_2)|^2. \tag{7.17}$$

We can compare this with Nakazawa's result[4] for the heat flow in the cylindrical analog of Lebowitz's model

$$J(N)=\pi^{-2}k\,m_1\,m_N\lambda^2\Delta T\int_0^\pi d\theta\int_{-\infty}^{\infty}\omega^2|\det Y(\omega,\theta)|^{-2}d\omega,$$

$$Y(\omega,\theta)=\Omega^2(\theta)-\omega^2M-i\omega\lambda ML. \tag{7.18}$$

When the mass sequence m_i is periodic, the results of the next section extend to show that the semiinfinite matrix $K(\theta)=M^{-1/2}\Omega^2(\theta)M^{-1/2}$ has only an absolutely continuous spectrum and so from Ref. 2 we see that both $J(N)$ and $\hat{J}(N)$ have strictly positive limits as $N\to\infty$. An examination of Theorem 1 in Ref. 5 shows that Furstenberg's theorem holds for any subgroup of $SL(2,R)$ which is generated by two noncommuting matrices $\left[\begin{smallmatrix}a_i & -1\\ 1 & 0\end{smallmatrix}\right]$ $(i=1,2)$. (See also Ref. 11.) Consequently, for a disordered cylindrical system $J(N)$ and $\hat{J}(N)$ will tend to zero as $N\to\infty$ for almost every choice of masses in the columns. We can use Sec. 3 to find the asymptotic behavior of $t_N(k_1,k_2)$ for large N. Equation (7.15) gives

$$|t_N(k_1,k_2)|^{-1}=|2\sin k_1|^{-1}\left|[1,-\exp(ik_1)]\,T_N\cdots T_1\times\begin{pmatrix}1\\ e^{ik_1}\end{pmatrix}\right|.$$

An extension of Theorem 8.1 in Ref. 7 shows that the angle between the rows and the angle between the columns of $T_N\cdots T_1$ converges to zero as N increases. So for any k_1, the angle between the vector $T_N\cdots T_1\binom{1}{e^{ik_1}}$ and the vector $\binom{1}{1}$ tends to zero. If $k_1\neq 0$, this shows that $|t_N(k_1,k_2)|^{-1}$ grows exponentially with N. The analog of (3.15) is now as follows: when $k_1\neq 0$, $\lim_{N\to\infty}-(1/N)\log|t_N(k_1,k_2)|=\gamma(k_1,k_2)$. When $k_1=0$, t_N is identically zero. The analog of $\gamma(\omega)$ (3.4a) is now $\gamma(\mathbf{k})$.

Theorem: For small $|\mathbf{k}|$, $\gamma(\mathbf{k})$ is continuous in $\mathbf{k}$ and

$$\lim_{|\mathbf{k}|\to 0}\frac{\gamma(\mathbf{k})}{g(\mathbf{k})}=\frac{1}{8}\frac{\langle m^2\rangle-\langle m\rangle^2}{\langle m\rangle}, \tag{7.19}$$

where

$$g(\mathbf{k})=(k_1^2+\mu\,k_2^2)^2\ k_1^2+\mu\left(1-\frac{1}{\langle m\rangle}\ k_2^2\right)^{-1}$$

(we assume $m_i\ge 1$, all i).

Proof: Adapting the calculation of Sec. 3 to the family of transfer matrices $T_j(k_1,k_2)$ yields

$$\gamma(\mathbf{k})=\int_{-\pi/2}^{\pi/2}\log\left|\frac{\cos(\theta+\eta)}{\cos\theta}\right|dG(\theta,\eta) \tag{7.20}$$

where

$$\begin{aligned}&G(A)=\int_0^\infty G[\Psi(A,m)]\,dp(m),\\ &\tan\phi=\tan[\Psi(\phi,m)+\eta]+[(m-\langle m\rangle)/\sin\eta]\,\omega^2,\\ &2\cos\eta=2+4\mu\sin^2\tfrac{1}{2}k_2-4\langle m\rangle(\sin^2\tfrac{1}{2}k_1+\mu\sin^2\tfrac{1}{2}k_2).\end{aligned} \tag{7.21}$$

If $a=[(m-\langle m\rangle)/\sin\eta]\,\omega^2$ then (3.12) gives

$$\gamma(\mathbf{k})=\int_0^\infty dp(m)\cdot a^2\int_{-\pi/2}^{\pi/2}\cos 2\theta\cos^2\theta\,dG(\theta,\eta)+0(\eta^3). \tag{7.22}$$

This gives the result.

8. SPECTRUM OF PERIODIC CHAINS

We consider a infinite periodic chain whose basic cell contains the masses $m_1\cdots m_A$. It is easy to see that the allowed bands for the chain are specified by the algebraic condition

$$|\mathrm{Tr}\ T(\omega)|\le 2, \tag{8.1}$$

where $T(\omega)=T_A\cdots T_1(\omega)$ is the transfer matrix for one cell of the chain. These bands form the spectrum of the self-adjoint infinite matrix operator $H=M^{-1/2}\Phi M^{-1/2}$ introduced in Sec. 1. H acts on the Hilbert space l^2 and has a cyclic vector, viz. $\psi=(1,0,0\cdots)$. So H has a simple spectrum whose spectral type (Ref. 19, Chap. VII) is precisely the type of the measure μ_Ψ on $[0,\infty)$ determined by

$$(\Psi, H^k\Psi)=\int_0^\infty x^k\, d\mu_\Psi \quad (k\geq 0). \tag{8.2}$$

Theorem: The spectrum of a semiinfinite periodic lattice is absolutely continuous.

Proof: We will evaluate the left side of (8.2) for finite periodic systems and then let the length tend to infinity. We will see that there is a unique measure satisfying (4.2) and that it is absolutely continuous. We also normalize the lightest mass to 1 so that all $m_i\geq 1$. The spectrum of H will therefore be in $[0,4]$.

Let P_N be the projection on l^2 which projects any vector onto its first NA entries. Then $H_N=P_NHP_N$ is the operator introduced in Sec. 1 corresponding to a periodic lattice with N cells. For each $k\geq 1$, the operators H_N^k converge strongly to H^k. Suppose that a chain containing N cells has normal modes with frequencies ω_a^2, $a=1\cdots NA$. The corresponding displacements are

$$\begin{aligned} &u(a,jA+p)=d_p(\omega_a)\sin[(jA+p)K_a],\\ &K_a=\pi a(NA+1)^{-1},\quad 1\leq p\leq A,\quad 0\leq j\leq N-1. \end{aligned} \tag{8.3}$$

$d_p(a)$ are certain constants depending on a. Using the transfer matrix method (8.2) can be a solution to the lattice equations only if

$$\mathrm{Tr}\ T(\omega_a^2)=2\cos AK_a. \tag{8.4}$$

The normalization condition for (4.3) is

$$1=|u(a)|^2=\frac{1}{2}\sum_{p=1}^{A}|d_p(a)|^2[N-\cos 2K_a(p-1)]. \tag{8.5}$$

So

$$(\Psi, H_N^k\Psi)=2N^{-1}\sum_{a=1}^{NA}F(\omega_a^2)\,\omega_a^{2k}+O(N^{-2})$$

with

$$F(\omega^2)=\sum_{p=1}^{A}d_p(\omega)^2. \tag{8.6}$$

Each $d_p(\omega)$ is a cofactor in a certain determinant and so $F(\omega^2)$ is a bounded continuous function. The spacing of the wave vectors K_a in a long finite chain is very nearly $\pi(NA)^{-1}$. Rewriting this in terms of the frequencies ω_a^2 and letting $N\to\infty$ gives

$$(\Psi, H^k\Psi)=2A\pi^{-1}\int\left|\frac{dK}{d\omega^2}\right|\omega^{2k}F(\omega^2)\,d(\omega^2) \tag{8.7}$$

provided that $dK/d\omega^2$ makes sense. The integral is over the values of ω^2 for which $|\mathrm{tr}\,T(\omega)|\leq 2$, i.e., the allowed frequency bands. Using (4.4) in the limit $N\to\infty$ gives

$$2\cos AK=\mathrm{Tr}\ T(\omega^2)=g(\omega^2), \tag{8.8}$$

So

$$\begin{aligned}\left|\frac{dK}{d\omega^2}\right| &=(2A)^{-1}g'(\omega^2)\,(4-g^2)^{-1/2}\\ &=(2A)^{-1}h(\omega^2).\end{aligned} \tag{8.9}$$

Any singularities occuring in this ratio are integrable and so

$$(\Psi, H^k\Psi)=\pi^{-1}\int h(\omega^2)\,F(\omega^2)\,\omega^{2k}\,d\omega^2=\int x^k p(x)\,dx. \tag{8.10}$$

$p(x)$ is an integrable function over the spectrum of the chain. Since H is bounded we also have

$$(\Psi, f(H)\Psi)=\int f(x)\,p(x)\,dx \tag{8.11}$$

for any polynomial f and consequently for all measurable functions f. Spectral theory now tells us that $p(x)\,dx$ is unique and consequently the spectrum of H is absolutely continuous.

9. DISCUSSION

The expression for the heat flux $J(N)$ in (1.7) and for $\hat{J}(N)$ in (2.16) [or (2.23)] differ essentially in that (1.7) contains an integral over all ω while the integration in (2.16) is restricted to the spectrum of the homogeneous chain with unit masses. (The difference in integrands is presumably due to the nature of the coupling between the system and heat baths in the two models.) It seems intuitively clear that the reason why frequencies outside the spectrum do not contribute to $\hat{J}(N)$ is that all such modes would be damped out in the homogeneous stretches of the side chains when $S\to\infty$. Indeed the integral in (1.7) will reduce to an integral only over the spectrum of the chain when we take a chain of length $N+2S$ in which $m_j=1$ if $1\leq j\leq S$ and $N+S+1\leq j\leq 2S+N$. Using expansions similar to those in Appendix C, it is easy to see that $\lim_{S\to\infty}J(N+2S)$ becomes an integral over $[0,4]$ involving only the determinants $K_{1,r}$, etc. There does not, however, seem to be a compact expression for this flux.

In any case, as we have seen, both $J(N)$ and $\hat{J}(N)$ approach nonvanishing limits when $N\to\infty$ in periodic systems and go to zero in random systems. The latter result follows from the behavior of the integrands $j_N(\omega)$ and $t_N^2(\omega)$ which, by Furstenberg's theorem, vanish for almost all chains as $\exp[-N\gamma(\omega)]$ as $N\to\infty$ for fixed ω, with $\gamma(\omega)>0$ for $\omega\neq 0$. The difficulty with using Furstenberg's theorem for the evaluation of the asymptotic form of $J(N)$ or $\hat{J}(N)$ (the latter ought to be easier since the integration is over a finite range) is that the approach to the limit in Furstenberg's theorem, i.e., in (3.15), is not known to be uniform in ω for $\omega\neq 0$. We need some such kind of uniformity to decide for certain whether Fourier's law is obeyed by random harmonic systems.

ACKNOWLEDGMENTS

This work was partly supported by USAFOSR Grant No. 73-2430. It was completed when one of us, A. O'Connor, was at the Dublin Institute For Advanced Studies and he thanks Professor L. O' Raifeartaigh for his hospitality. It is also a pleasure to thank Professor E. Lieb and Dr. R.J. Rubin for valuable discussions.

APPENDIX A

We show that the matrix $Y(\omega)=\Phi-M\omega^2-A(\omega)U$ in Sec. 1 is nonsingular for real ω except at $\omega^2=0$ and 4. If $\omega^2\leq 4$, we write $\omega=2\sin(\theta/2)$ and if $D(\omega)=\det Y(\omega)$

vanishes we have

$$\cos\theta(K_{1,r}+K_{2,r-1})=K_{2,r}+K_{1,r-1},$$
$$\sin\theta(K_{1,r}-K_{2,r-1})=0. \tag{A1}$$

If θ is different from 0 and π, then $K_{2,r-1}=K_{1,r}$ and (3.4) implies $K_{2,r}K_{1,r-1}=1+K_{1,r}^2$ so that $|\cos\theta|\geq 1$. Equality is only possible if $\theta=0$ or π, i.e., $\omega^2=0$ or 4. When $\omega^2=0$, $A(\omega)=1$ and $D=0$. When $\omega^2=4$, D need not vanish and will only have a simple zero. If $\omega^2>4$, we consider a square matrix C with $2N+r$ rows and columns:

$$C=\Phi-M\omega^2 \tag{A2}$$

with Φ the usual tridiagonal matrix and M and diagonal matrix with entries $M_{ii}=1$ if $1\leq i\leq N$ and $N+r+1\leq i\leq 2N+r$ and $M_{ii}=m_j$ if $i=N+j$, $j=1\cdots r$. Then

$$\det C=d(N-2)^2K_{1,r}-d(N)\,d(N-2)\,(K_{1,r-1}+K_{2,r})+d(N)^2K_{2,r-1}, \tag{A3}$$

$$d(N)=(-1)^N\frac{\sinh(N+1)\theta}{\sinh\theta}\quad\text{when }\omega^2=4\cosh^2\theta.$$

Also, using Rayleigh's theorem,[20] we can find lower bounds on the eigenvalues of the matrix $M^{-1/2}\Phi M^{-1/2}$ and so prove that when N is large

$$|\det C|\geq(\omega^2)^r\;d(2N)\,m_1\cdots m_r. \tag{A4}$$

So $\lim_{N\to\infty}\exp(-2N\theta)\;|C|=|D(\omega)|>0$. The only zeros in $D(\omega)$ then are cancelled by the zeros in the numerator of (1.24).

APPENDIX B

We show that for an infinite harmonic chain with $m_j=1$ except possibly when $N\geq j\geq 1$ that if $u_j=0=\dot{u}_j$ at $t=0$ except when $N\geq j\geq 1$ then $|u_j(t)|$ falls off as $t^{-1/2}$ when $N\geq j\geq 1$. We check the case when $\dot{\mathbf{u}}(0)=0$ but $\mathbf{u}(0)\neq 0$. Then

$$\mathbf{u}(t)=(2\pi)^{-1}\int_C\exp(i\omega t)\,(\Phi-M\omega^2-a(\omega)L)^{-1}\mathbf{u}(0)\,d\omega. \tag{B1}$$

C is a contour obtained as the limit of semicircles in the upper half plane with radius R and centre $-i\epsilon$ (ϵ very small and positive). The integrand may have some poles in the upper half plane and has a cut along the real axis from -2 to $+2$. The poles contribute exponentially decreasing terms and the cut a term of type

$$\int_{C_1}\omega^2(4-\omega^2)\,[P(\omega)+iQ(\omega)\,(4-\omega^2)^{1/2}]^{-1}\exp(i\omega t)\,d\omega. \tag{B2}$$

P and Q are polynomials in ω and the integrand has only the singularity due to the branch in the square root. C_1 is a contour enclosing the interval $(-2,2)$. We can easily check that if f and f' are integrable $\int_{-2}^{2}f(\omega)\exp(i\omega t)\,d\omega$ falls off at least as t^{-1} and that $\int_{-2}^{2}f(\omega)\,(4-\omega^2)^{1/2}\exp(i\omega t)\,d\omega$ falls off at least as $t^{-1/2}$. This proves the claim.

APPENDIX C

We give here the detailed computation for the heat flow along an infinite isotropic chain containing a single impurity. Using the Langevin equation approach of Ishii[5,16] and the method of Sec. 2 we find that the heat flow across a segment of r masses, $m_1\cdots m_r$, embeded in isotropic chains of length N is just

$$J=\lambda^2kTm_1m_r\int_{-2}^{2}\omega^2|b(\omega)|^4\,|\det Z(\omega)|^{-2}\,d\omega. \tag{C1}$$

$Z(\omega)=\Phi-M\omega^2-a(\omega)L$ is the standard $r\times r$ matrix. $a(\omega)$ and $b(\omega)$ are given by

$$(d_{N+1}-i\omega\lambda\,d_N)\,a(\omega)=d_N-i\omega\lambda\,d_{N-1},$$
$$(d_{N+1}-i\omega\lambda\,d_N)\,b(\omega)=1, \tag{C2}$$

where $d_p(\omega)=\sin(p+1)\theta/\sin\theta$ if $\omega^2=2(1-\cos\theta)$.

We can use the methods of Ref. 2 to show that as $N\to\infty$ this reduces to an integral over the spectrum of the infinite homogeneous chain (if any $m_j<1$ then there are some localized modes with frequencies greater than 2 but an explicit examination shows that these do not contribute to J as $N\to\infty$). This integral is only tractable when $r=1$. In that case putting $\phi=(2N+1)\theta$ and letting $N\to\infty$, we get

$$J=(2\pi)^{-1}m\lambda^2kT\int_0^{2\pi}d\omega\,\omega^2(\theta)\sin^2\theta\times\int_0^{2\pi}d\phi\,|F(\theta,\phi)|^{-2}, \tag{C3}$$

$$F(\theta,\phi)=a\cos2\phi+b\sin2\phi+e-if,$$
$$a=\cos2\theta+u^2-v\sin2\theta,$$
$$b=\sin2\theta+v(\cos2\theta+u^2),$$
$$e-if=-v(1+u^2)-2iu\sin\theta. \tag{C4}$$

Standard manipulations will reduce (C3) to (4.6).

*Research supported by USAFOSR Grant No. 73-2430.
†Present address: Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland.

[1](a) P.G. Bergmann and J.L. Lebowitz, Phys. Rev. 99, 578 (1965). (b) J.L. Lebowitz, Phys. Rev. 114, 1192 (1959). (c) Z. Rieder, J.L. Lebowitz, and E. Lieb, J. Math. Phys. 8, 1073 (1967).
[2]A. Casher and J.L. Lebowitz, J. Math Phys. 12, 1701 (1971).
[3]R.H.G. Gellemann, "Heat Transport in Harmonic and Anharmonic Lattices," thesis, Yeshiva University, 1971.
[4]H. Nakazawa, Suppl. Progr. Theor. Phys. 45, 231 (1970).
[5]H. Matsuda and K. Ishii, Suppl. Prog. Theor. Phys. 45, 56 (1970).
[6]R.E. Borland, Proc. Roy. Soc. A274, 529 (1963).
[7]H. Furstenberg, Trans. Am. Math. Soc. 108, 377 (1963).
[8]R.J. Rubin and W.L. Greer, J. Math. Phys. 12, 1686 (1971).
[9]G.W. Ford, M. Kac, and P. Mazur, J. Math. Phys. 6, 504 (1965).
[10]V.B. Magalinskii, Zh. Eksp. Teor. Fiz. 36, 1942 (1959) [Sov. Phys. JETP 9, 1381 (1959)].
[11]Y. Yoshioka, "On the Singularity of the Spectral Measures of a Semi-Infinite Random System" (to be published).
[12]R.J. Rubin, J. Math. Phys. 9, 2252 (1968); J. Math. Phys. 11, 1857 (1970).
[13]S. Minami and J. Hori; Suppl. Prog. Theor. Phys. 45, 87 (1970).
[14]P.L. Sulem and H. Frisch, J. Plasma Phys. 8, 217 (1972).
[15]U. Frisch, C. Froeschle, J.P. Scheidecker, and P.L. Sulem, Phys. Rev. A 8, 1416 (1973).
[16]K. Ishii, "Localization of eigenstates and transport phenomena in one-dimensional disordered systems," preprint, Research Institute for Fundamental Physics, Kyoto University, 1972.
[17]W.M. Visscher, Prog. Theor. Phys. 46, 729 (1971).
[18]J.L. Lebowitz and H.L. Frisch, Phys. Rev. 107, 917 (1957).
[19]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Vol. I.
[20]A. Maradudin, E. Montroll, and G. Weiss, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic, New York, 1963).

Lattice Green's function for the face centered cubic lattice

Michiko Inoue

RCA Research Laboratories, Inc., P.O. Box 5151, Tokyo International 100-31, Japan
(Received 12 February 1974)

We have proved that Green's function $G(l,m,n)$ at an arbitrary lattice site (l,m,n) in face centered cubic lattice with nearest neighbor interactions is, in general, expressed in terms of linear combinations of products of complete elliptic integrals of the first and second kinds.

1. INTRODUCTION

The method of Green's function has proved to be very powerful for quantitative studies of a variety of problems in solid state physics. Extensive investigations have been made to evaluate analytically as well as numerically lattice Green's functions for different crystal structures. For face centered cubic (fcc) lattice with nearest neighbor interactions, a lattice Green's function at a point (l,m,n) is represented by the following integral:

$$G(l,m,n)=\frac{1}{\pi^3}\int_0^\pi\int_0^\pi\int_0^\pi dx\,dy\,dz \times\frac{\cos lx\cos my\cos nz}{E-i\delta-\cos x\cos y-\cos y\cos z-\cos z\cos x}, \tag{1.1}$$

where $l+m+n$ is zero or an even integer, δ is an infinitesimal number and E is a real number between $-\infty$ and $+\infty$.

Iwata[1] has shown that the above function at the origin, $G(0,0,0)$, can be expressed in a compact form as a product of complete elliptic integrals of the first kind.

In the present paper we show that $G(l,m,n)$ at an arbitrary site (l,m,n) is given, in general, by a linear combination of products of complete elliptic integrals of the first and second kinds. In the region where $E>3$ and $E<-1$, $G(l,m,n)$ is real and given by a function of $K(k)$ and $E(k)$ with real moduli. For $3>E>-1$, $G(l,m,n)$ has a nonvanishing imaginary part and is expressed as a function of $K(k)$ and $E(k)$ with complex moduli. The method of analytic continuation for $K(k)$ and $E(k)$, as described by Morita and Horiguchi,[2] is applicable to the present case to evaluate $G(l,m,n)$ for the whole range of values of E.

In Sec. 2 we derive the recurrence relations for $G(l,m,n)$, which shows that the knowledge of a finite group of Green's functions, $G(2p,0,0)$ ($p=0$ or a positive integer) and $G(2,2,0)$ is enough to determine the entire family of $G(l,m,n)$. In Sec. 3 we present explicit evaluations of $G(2p,0,0)$ and $G(2,2,0)$ for $E>3$, and the results are easily transformed into expressions valid for $E<-1$.

2. RECURRENCE RELATIONS FOR $G(l, m, n)$

With the representation (1.1), the following relation holds for $G(l,m,n)$:

$$\begin{aligned}&G(l+1,m+1,n)+G(l-1,m-1,n)+G(l+1,m-1,n)\\&+G(l-1,m+1,n)+G(l,m+1,n+1)\\&+G(l,m-1,n-1)+G(l,m+1,n-1)\\&+G(l,m-1,n+1)+G(l+1,m,n+1)\\&+G(l-1,m,n-1)+G(l+1,m,n-1)\\&+G(l-1,m,n+1)=4EG(l,m,n)-4\delta_{0l}\delta_{0m}\delta_{0n},\end{aligned} \tag{2.1}$$

where δ_{l0} is Kronecker's delta. Furthermore, since the number of nonequivalent lattice point in a cubic lattice lie inside and on the surfaces of a space bounded by three planes (0,0,1), (1,−1,0) and (0,1,−1), a portion corresponding to 1/48 of the whole space, it is sufficient to determine Green's functions $G(l,m,n)$ at these lattice points. Then, it is readily found that a function $G(l,m,n+1)$ for a positive integer n can be obtained successively from (2.1) if $G(l,m,0)$ and $G(l+1,m+1,1)$ are determined. In other words, it is necessary to obtain Green's function at lattice points on the two adjacent layers for $z=0$ and 1 parallel to the (0,0,1) planes, illustrated by a shaded region including the axes $y=0$ and $x=y$ in Fig. 1.

Now we derive a recurrence relation for $G(l,m,0)$ by adopting the procedure due to Morita[3] for two-dimensional lattice. For simplicity we assume $E>3$, and hence can neglect a factor $i\delta$ in the denominator of (1.1). When we integrate (1.1) with respect to z for $n=0$, we get

$$G(l,m,0)=\frac{1}{\pi^2}\int_0^\pi\int_0^\pi dx\,dy \times\frac{\cos lx\cos my}{[(E-\cos x\cos y)^2-(\cos x+\cos y)^2]^{1/2}}. \tag{2.2}$$

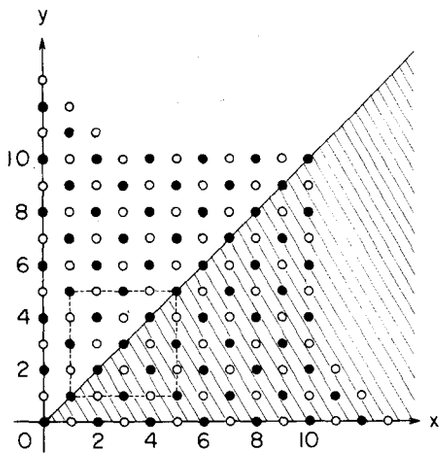


FIG. 1. Lattice points for fcc lattice on the two adjacent layers parallel to the (0, 0, 1) plane. ● denotes a point on the first layer and ○ a point on the second layer. A square by dotted line encloses 13 lattice point whose $G(l, m, n)$ are related by Eq. (2.9).

Introducing a definition

$$F_m(E,x)=\int_0^\pi dy\,\frac{\cos my}{[(E-\cos x\cos y)^2-(\cos x+\cos y)^2]^{1/2}},\tag{2.3}$$

we rewrite (2.2) as

$$G(l,m,0)=\frac{1}{\pi^2}\int_0^\pi dx\cos lx\,F_m(E,x).\tag{2.4}$$

Let's consider an integral for $m\neq 0$

$$f_m(E,x)\equiv\int_0^\pi dy\cos my[(E-\cos x\cos y)^2-(\cos x+\cos y^2]^{1/2}.\tag{2.5a}$$

which is expressed in terms of $F_m(E,x)$ as

$$\begin{aligned}f_m(E,x)=&\int_0^\pi dy\\&\times\frac{\cos my[E^2-\cos^2x-2(E+1)\cos x\cos y-\sin^2x\cos^2y]}{[(E-\cos x\cos y)^2-(\cos x+\cos y)^2]^{1/2}}\\&=(E^2-\cos^2x)\,F_m(E,x)-(E+1)\\&\quad\times\cos x[F_{m+1}(E,x)+F_{m-1}(E,x)]\\&\quad-\frac{\sin^2x}{4}[F_{m+2}(E,x)+F_{m-2}(E,x)+2F_m(E,x)].\end{aligned}\tag{2.5b}$$

Integrating (2.5a) by parts, we get

$$\begin{aligned}f_m(E,x)=&-\frac{1}{m}\int_0^\pi dy\\&\times\frac{\sin my[(1+E)\cos x\sin y+\sin^2x\cos y\sin y]}{[(E-\cos x\cos y)^2-(\cos x+\cos y)^2]^{1/2}}\\&=\frac{(1+E)\cos x}{2m}[F_{m+1}(E,x)-F_{m-1}(E,x)]\\&\quad+\frac{\sin^2x}{4m}[F_{m+2}(E,x)-F_{m-2}(E,x)].\end{aligned}\tag{2.6}$$

Equating the two expressions for $f_m(E,x)$, we obtain the following relation for $F_m(E,x)$s:

$$\begin{aligned}&(m+1)\sin^2x\,F_{m+2}(E,x)+2(1+2m)(1+E)\cos x\,F_{m+1}(E,x)\\&\quad-2m(2E^2-2+\sin^2x)\,F_m(E,x)\\&\quad-2(1-2m)(1+E)\cos x\,F_{m-1}(E,x)\\&\quad-(1-m)\sin^2x\,F_{m-2}(E,x)=0.\end{aligned}\tag{2.7}$$

Then, by use of (2.4),

$$\begin{aligned}&G(l+1+2(p+1),l+1,0)+G(l+1+2(p-1),l+1,0)\\&\quad=2\int_0^\pi dx\cos(l+2p+1)x[1-2\sin^2x]\,F_{l+1}(E,x).\end{aligned}\tag{2.8}$$

Substitution of (2.7) into (2.8) with some rearrangement of terms deduce the recurrence relation for $G(l,m,0)$,

$$\begin{aligned}&G(l+2p+3,l+1,0)+G(l+2p-1,l+1,0)\\&-2G(l+2p+1,l+1,0)\\&-\frac{4(1+E)(2l-1)}{l}[G(l+2p+2,l,0)+G(l+2p,l,0)]\\&+\frac{4(4E^2-3)(l-1)}{l}G(l+2p+1,l-1,0)\\&-\frac{2(l-1)}{l}[G(l+2p+3,l-1,0)+G(l+2p-1,l-1,0)]\\&-\frac{4(1+E)(2l-3)}{l}[G(l+2p+2,l-2,0)+G(l+2p,l-2,0)]\\&-\frac{2(l-2)}{l}G(l+2p+1,l-3,0)\\&+\frac{(l-2)}{l}[G(l+2p+3,l-3,0)\\&+G(l+2p-1,l-3,0)]=0,\end{aligned}\tag{2.9}$$

where l is a positive integer and p is zero or an integer. Eq. (2.9) gives rise to a relation among Green's function for 13 lattice points lying on the edges and inside a square formed by four lines connecting the lth with the $(l+5)$th sites along the x and y axes (see Fig. 1), except the square whose center is at the origin, i.e., $l=0$ in (2.9).

We show below that proper application of (2.1) and (2.9) makes it possible to obtain any function $G(l,m,0)$ or $G(l+1,m,1)$ for arbitrary integers l,m, provided that we know the functions $G(2,2,0)$ and $G(2p,0,0)$ where p is zero or a positive integer. Suppose that $G(2,2,0)$ and $G(2p,0,0)$ have been obtained. Then $G(1,1,0)$ is easily obtained from (2.1) as

$$G(1,1,0)=\tfrac{1}{3}[EG(0,0,0)-1].$$

Moreover, with (2.2) we can always find $G(l+m,m-l+1,1)$ where l changes from 1 to m for a positive integer m by

$$\begin{aligned}2G(l+m,m-l+1,1)=&4EG(l+m-1,m-l+1,0)\\&-G(l+m,m-l+2,0)-G(l+m-2,m-l,0)\\&-G(l+m,m-l,0)-G(l+m-2,m-l+2,0)\\&-2G(l+m-1,m-l+2,1)-2G(l+m-1,m-l,1)\\&-2G(l+m-2,m-l+1,1).\end{aligned}\tag{2.10}$$

Now, let's place the center of a square determined by (2.9) at the point $(2q,0,0)$ $(q\neq 0)$ and move it to the adjacent point along the y axis up to $(2q,2q,0)$, so that we can determined successively $G(2q+2,2n+2,0)$ for $n=0,1,2\cdots q$ in terms of the known functions $G(l,m,0)$ for $l,m<2q+2$. Similarly, by moving the position of the center of the square from $(2q-1,1,0)$ to $(2q-1,2q-1,0)$, we can obtain successively $G(2q+1,2n'+1,0)$ for $n'=1,2\cdots q$, in terms of the functions $G(2q+1,1,0)$ and $G(l,m,0)$ for $l,m<2q+1$. A function $G(2q+1,1,0)$ is in turn given from (2.1) by $G(2q+1,1,0)=EG(2q,0,0)-G(2q-1,1,0)-G(2q,1,1)$ where $G(2q,1,1)$ can be independently obtained from (2.10). As the above procedure is valid for $q=1$ it holds for any positive integer q.

3. CALCULATIONS OF $G(2p,0,0)$ AND $G(2,2,0)$

In this section we shall derive the explicit expressions of $G(2p,0,0)$ and $G(2,2,0)$. Let's start with $G(2p,0,0)$:

$$G(2p,0,0)=\frac{1}{\pi^2}\int_0^\pi\int_0^\pi dx\,dy$$

$$\times \frac{\cos 2px}{[(E-\cos x\cos y)^2-(\cos x+\cos y)^2]^{1/2}}. \quad (3.1)$$

We perform the integration over y in a straightforward manner and

$$G(2p,0,0)=\frac{2}{\pi^2(1+E)}\int_0^\pi dx\cos 2px\,K(k_0), \quad (3.2)$$

where $K(k_0)$ is the complete elliptic integral of the first kind with modulus given by $k_0=2(E+\cos^2 x)^{1/2}/(1+E)$. If we replace $K(k_0)$ by a hypergeometric series and exchange the order of integral and summation, we get

$$G(2p,0,0)=\frac{1}{\pi(1+E)}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\left(\frac{2}{1+E}\right)^{2n}\times\int_0^\pi dx\cos 2px(E+\cos^2 x)^n, \quad (3.3)$$

where the standard symbol $(m)_n=\Gamma(n+m)/\Gamma(m)$ is used. We note that the integral in (3.3) gives rise to associated Legendre function $P_n^p(\xi)$ as[4]

$$\int_0^\pi dx\cos 2px(E+\cos^2 x)^n=\frac{\pi[E(1+E)]^{n/2}}{(n+1)_p}P_n^p(\xi), \quad (3.4)$$

with

$$\xi=(E+\tfrac{1}{2})/[E(1+E)]^{1/2}.$$

The associated Legendre function is in turn related to Jacobi polynomials $P_n^{(m,m)}(\xi)$,[4]

$$P_n^p(\xi)=\frac{(\xi^2-1)^{p/2}}{2^p}(n+1)_p\,P_{n-p}^{(p,p)}(\xi). \quad (3.5)$$

A successive application of recurrence formula for $P_{n-p}^{(p,p)}$ leads to

$$P_{n-p}^{(p,p)}(\xi)=\sum_{j=0}^{p}(-1)^j\binom{p}{j}P_n^{(p-j,j)}(\xi). \quad (3.6)$$

Thus by use of (3.3), (3.4), and (3.5), we have

$$G(2p,0,0)=\frac{1}{2^{2p}(1+E)^{1+p/2}E^{p/2}}\sum_{j=0}^{p}(-1)^j\binom{p}{j}\sum_{n=0}^{\infty}\frac{(\frac{1}{2})_n(\frac{1}{2})_n}{(1)_n(1)_n}\times\left(\frac{2}{1+E}\right)^{2n}[E(1+E)]^{n/2}P_n^{(p-j,j)}(\xi). \quad (3.7)$$

Furthermore, using the identity for an arbitrary inger q,

$$\frac{(\frac{1}{2})_n}{(1)_n}=\frac{1}{q!(q+1)_n}\sum_{\mu=0}^{q}\binom{q}{\mu}(\tfrac{1}{2})_{q-\mu}(\tfrac{1}{2})_\mu(\mu+\tfrac{1}{2})_n, \quad (3.8)$$

we can replace the summation over n in (3.7) by a hypergeometric function of Appell's type, $F_4(\alpha,\beta,\gamma,\gamma',X,Y)$.[4] Thus

$$G(2p,0,0)=\frac{1}{2^{2p}(1+E)^{1+p/2}E^{p/2}2p!}\sum_{j=0}^{p}\sum_{\mu=0}^{p-j}\sum_{\mu'=0}^{j}C_{\mu\mu'}^{p-jj}F_4(\mu+\tfrac{1}{2},\mu'+\tfrac{1}{2},p-j+1,j+1,X_+,X_-), \quad (3.9)$$

where

$$C_{\mu\mu'}^{p-jj}=(-1)^j\binom{p}{j}^2\binom{p-j}{\mu}(\tfrac{1}{2})_\mu(\tfrac{1}{2})_{p-j-\mu}\binom{j}{\mu'}(\tfrac{1}{2})_{\mu'}(\tfrac{1}{2})_{j-\mu'}, \quad (3.9a)$$

and

$$X_\pm=\frac{2}{(1+E)^2}[E+\tfrac{1}{2}\mp E^{1/2}(1+E)^{1/2}]. \quad (3.9b)$$

We note that it is possible to expand $F_4(\alpha,\beta,\gamma,\gamma',x,y)$ in a finite series of products of simple hypergeometric function, if $\gamma+\gamma'-\alpha-\beta-1$ is zero or positive integer, as it is in the present case. Thus

$$F_4(\alpha,\beta,\gamma,\gamma',x(1-y),y(1-x))=\sum_{k=0}^{\nu}\binom{\nu}{k}(-y)^k\frac{(\beta)_k}{(\gamma')_k}\,{}_2F_1(\alpha,\beta+k;\gamma;x)\,{}_2F_1(\gamma+\gamma'-\beta-1,\beta+k;\gamma'+k;y), \quad (3.10)$$

with $\nu=\gamma+\gamma'-\alpha-\beta-1$. The relation (3.10) is easily shown if one expands the integral representation for $F_4(\alpha,\beta,\gamma,\gamma',x(1-y),y(1-x))$ in terms of products of integrals for simple hypergeometric functions. Finally it should be mentioned that a hypergeometric function of the form ${}_2F_1(a-\frac{1}{2},b-\frac{1}{2};c;x)$ for positive integers a,b,c, is reduced to linear combinations of ${}_2F_1(\frac{1}{2},\frac{1}{2},1;x)$ and ${}_2F_1(\frac{1}{2},-\frac{1}{2};1;x)$ after a successive application of Gauss' relations.

Hence we conclude that $G(2p,0,0)$ can always be expressed as a finite sum of products of the complete elliptic integrals of the first and the second kinds,

$$K(k_\pm)=(\pi/2)\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;k_\pm^2),$$

$$E(k_\pm)=(\pi/2)\,{}_2F_1(\tfrac{1}{2},-\tfrac{1}{2};1;k_\pm^2),$$

where the moduli $k_\pm$ are calculated to be

$$k_\pm^2=\frac{1}{2}\left(1\mp\frac{4E^{1/2}(1+E)^{1/2}}{(1+E)^2}-\frac{(E-1)(E+1)^{1/2}(E-3)^{1/2}}{(1+E)^2}\right). \quad (3.11)$$

Next, we derive an expression for $G(2,2,0)$. We rewrite $G(2,2,0)$ in the following form

$$G(2,2,0)=G(0,2,0)-\frac{2}{\pi^2}\int_0^\pi dx\sin^2 x\,F_2(E,x). \quad (3.12)$$

A straightforward calculation of $F_2(E,x)$ from (2.3) is carried out to give the following result:

$$F_2(E,x)=\frac{2}{\sin^2 x}\left(\frac{4\sin^2 x}{(1+E)(1+k_0')}K(k)-(1+E)(1+k_0')E(k)+\frac{2(1+E)k_0'}{1+k_0'}\Pi(\mu^2,k)\right)-F_0(E,x), \quad (3.13)$$

where

$$k_0=2(E+\cos^2 x)^{1/2}/(1+E),$$

$$k_0'=(1-k_0^2)^{1/2},$$

$$k=\frac{1-k_0'}{1+k_0'},$$

$$\mu^2=[E-1-(E+1)k_0']/[E-1+(E+1)k_0'].$$

and $\Pi(\mu^2,k)$ is the complete elliptic integral of the third kind. Substituting (3.13) into (3.12), we get

$$G(2,2,0)=G(0,0,0)-\frac{4}{\pi^2}\int_0^{\pi}dx\left(\frac{4\sin^2x}{(1+E)(1+k_0')}K(k)-(1+E)(1+k_0')E(k)+\frac{4(1+E)k_0'}{1+k_0'}\Pi(\mu^2,k)\right). \tag{3.14}$$

A similar calculation is made to obtain $G(1,1,0)$

$$G(1,1,0)=\frac{2}{\pi^2}\int_0^{\pi}dx\left(\frac{E-1+(E+1)k_0'}{(1+E)(1+k_0')}K(k)-\frac{2k_0'}{1+k_0'}\Pi(\mu^2,k)\right). \tag{3.15}$$

Eliminating the integral involving $\Pi(\mu^2,k)$ from (3.14) and (3.15), and applying transformations for $K(k)$ and $E(k)$, we get

$$G(2,2,0)=(-\tfrac{2}{3}E^2+\tfrac{4}{3}E+1)G(0,0,0)+2G(2,0,0)-\tfrac{4}{3}(1+E)+\frac{8(1+E)}{\pi^2}\int_0^{\pi/2}dx\,E(k_0). \tag{3.16}$$

The last integral in (3.16) is calculated in an analogous way as we have derived (3.9) from (3.3). Thus,

$$\int_0^{\pi/2}dx\,E(k_0)=K(k_+)K(k_-)+2E(k_+)E(k_-)-K(k_+)E(k_-)-K(k_-)E(k_+). \tag{3.17}$$

Expressions of $G(0,0,0)$ and $G(2,0,0)$ are given from (3.9) and (3.10):

$$G(0,0,0)=\frac{4}{\pi^2(1+E)}K(k_+)K(k_-), \tag{3.18a}$$

$$G(2,0,0)=\frac{4}{\pi^2(1+E)^3k_+^2k_-^2}\left(K(k_+)K(k_-)+\frac{1}{(1-k_+^2)(1-k_-^2)}E(k_+)E(k_-)-\frac{1}{(1-k_-^2)}K(k_+)E(k_-)-\frac{1}{(1-k_+^2)}K(k_-)E(k_+)\right), \tag{3.18b}$$

where the moduli $k_\pm$ are given by (3.11).

So far we have tacitly assumed $E>3$, for which the moduli $k_\pm$ defined by (3.11) are real numbers. When $E<-1$, $k_\pm$ are pure imaginary. In this case we shall use a transformation

$${}_2F_1(\tfrac{1}{2},\pm\tfrac{1}{2};1;z)=(1-z)^{\mp1/2}\,{}_2F_1\left(\tfrac{1}{2},\pm\tfrac{1}{2};1;\frac{z}{z-1}\right), \tag{3.19}$$

so that the general expressions for $G(l,m,n)$ is valid for $E<-1$ if we replace $K(k_\pm)$ and $E(k_\pm)$ respectively by $K(q_\pm)/a_\pm$ and $a_\pm E(q_\pm)$, where

$$a_\pm=\frac{1}{2^{1/2}}\left(1+\frac{(1-E)(-1-E)^{1/2}(3-E)^{1/2}}{(-1-E)^2}\mp\frac{4(-E)^{1/2}(-1-E)^{1/2}}{(-1-E)^2}\right)^{1/2}, \tag{3.20a}$$

$$q_\pm=2[(-E)^{1/2}\mp(-1-E)^{1/2}]/[1-E+(3-E)^{1/2}(-1-E)^{1/2}]. \tag{3.20b}$$

4. CONCLUDING REMARKS

We have proved that Green's function $G(l,m,n)$ at an arbitrary lattice point (l,m,n) in fcc lattice can be expressed in an exact form as a real function of products of the complete elliptic integrals of the first and the second kinds for $E>3$ and $E<-1$. As is easily seen from (1.1), the function $G(l,m,n)$ is analytic in the complex E plane with a branch cut along the real axis between -1 and 3, so that the theory of analytic continuation makes it possible to obtain the expression of $G(l,m,n)$ for the entire E plane. In our case, the analytical continuation of $G(l,m,n)$ is equivalent to those of $K(k)$ and $E(k)$ in the complex k plane which has branch cuts on the real axis from 1 to ∞ and from -1 to $-\infty$. Since $K(k)$ and $E(k)$ on any branch on the Rieman surface is expressible in terms of $K(k)$, $K'(k)$, $E(k)$, and $E(k')$ on the principal branch, it is sufficient to calculate $K(k)$ and $E(k)$ with complex k on the latter branch, which is numerically evaluated on computer. Actual procedure of calculations for $K(k)$ with complex k is described in detail by Morita and Horiguchi in Ref. 2 for the case of $G(0,0,0)$. The method is easily extended to the present case to compute $G(2,0,0)$, $G(2,2,0)$ to a good accuracy.[5]

For other cubic lattices, we can also derive a general formula to calculate $G(l,m,n)$ in terms of complete elliptic integrals of the first and second kinds.[5]

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. Hirota for his critical reading of the manuscript.

[1]G. Iwata, Nat. Sci. Rep., Ochanomizu Univ. 20, 13 (1969).
[2]T. Morita and T. Horiguchi, J. Math. Phys. 12, 986 (1971).
[3]T. Morita, J. Math. Phys. 12, 1744 (1971).
[4]*Higher Transcendental Functions*, edited by A. Erdélyi, *et al.* (McGraw-Hill, New York, 1953), Vols. I–III.
[5]M. Inoue (unpublished).

Geometrodynamics regained: A Lagrangian approach*

Karel Kuchař

Department of Physics, University of Utah, Salt Lake City, Utah 84112
(Received 19 September 1973)

The closing relation between two super-Hamiltonians is cast into a condition on the super-Lagrangian by a functional Legendre transformation. It is shown that the ADM super-Lagrangian provides the unique representation of the "group" of deformations of a spacelike hypersurface embedded in a Riemannian space–time when the intrinsic geometry g_{ij} of the hypersurface is allowed as the sole configuration variable. No such uniqueness exists for the super-Lagrangians of source fields. As an illustration, the most general super-Lagrangian for a scalar field with nonderivative gravitational coupling is recovered from the closing relation.

1. INTRODUCTION

Geometrodynamics, pure or driven by sources, may be viewed as representing the "group" of deformations of a spacelike hypersurface embedded in a Riemannian space–time. In this paper, we show that pure geometrodynamics has a privileged position—it is the unique representation of the "group" of deformations using the intrinsic geometry g_{ij} of the hypersurface as the sole configuration variable. This generalizes the proof given in Ref. 1 that pure geometrodynamics is the only *time-reversible* representation. Therefore, no irreversible geometrodynamics exists!

While the representation requirement determines pure geometrodynamics uniquely, it still leaves a considerable freedom to the sources of geometry. We illustrate this in the simplest case of a scalar field with nonderivative gravitational coupling. There are infinitely many ways in which such a field may evolve and we shall show how to recover all of them directly from the representation requirement.

What do we mean by saying that a field dynamics represents the "group" of deformations? In the Hamiltonian formalism, the field on a spacelike hypersurface $X^\mu = X^\mu(x)$ is described by a set of canonical coordinates $\phi^A(x)$ and conjugate momenta $\pi_A(x)$.[2] In pure geometrodynamics, the field $\phi^A(x)$ is the intrinsic geometry $g_{ij}(x)$ of the hypersurface itself. In driven geometrodynamics, other fields (scalar, electromagnetic, etc.) enter as sources of the geometry g_{ij} and are included among the variables ϕ^A. We ask how the field changes when we deform the hypersurface $X^\mu = X^\mu(x)$ into a neighboring hypersurface $X^\mu = X^\mu(x) + \delta X^\mu(x)$. The displacement δX^μ is decomposed into normal and tangential components, δN and δN^i,

$$\delta X^\mu = n^\mu \delta N + X^\mu_{,i} \delta N^i. \tag{1.1}$$

Here, n^μ is the unit normal to the hypersurface and $X^\mu_{,i} \equiv X^\mu_{,i}$ are the tangent vectors in the direction of intrinsic coordinate lines x^i. The field dynamics is governed by a super-Hamiltonian $\mathcal{H}(x)[\phi^A, \pi_A]$ and a supermomentum $\mathcal{H}_i(x)[\phi^A, \pi_A]$ constructed from the canonical variables ϕ^A, π_A of the theory. An arbitrary functional F of these variables changes under the deformation δN, δN^i by the amount

$$\delta F = [F, \mathcal{H}_x]\delta N^x + [F, \mathcal{H}_{ix}]\delta N^{ix}. \tag{1.2}$$

The super-Hamiltonian $\mathcal{H}$ and supermomentum $\mathcal{H}_i$ push the field by means of the Poisson brackets (1.2) when the hypersurface is pushed according to Eq. (1.1). In order that the dynamics of the field develops consistently with the kinematics of slice deformations, the expressions $\mathcal{H}$ and $\mathcal{H}_i$ must close in the way which is characteristic for the generators of the "group" of deformations[3,1]:

$$[\mathcal{H}_i(x), \mathcal{H}_j(x')] = \mathcal{H}_i(x')\delta_{,j}(x, x') - (x \rightarrow x'), \tag{1.3}$$

$$[\mathcal{H}_i(x), \mathcal{H}(x')] = \mathcal{H}(x)\delta_{,i}(x, x'), \tag{1.4}$$

$$[\mathcal{H}(x), \mathcal{H}(x')] = \mathcal{H}^i(x)\delta_{,i}(x, x') - (x \rightarrow x'). \tag{1.5}$$

This poses the representation requirement.

The supermomentum may be found directly, not using Eqs. (1.3)–(1.5) at all, when realizing that the tangential deformation δN^i is equivalent to the relabeling $x^i \rightarrow \bar{x}^i = x^i + \delta N^i$ of the hypersurface. The change of the functional F under the tangential deformation should thus equal to its Lie derivative with respect to δN^i,

$$[F, \mathcal{H}_{ix'}]\delta N^{ix'} = \mathcal{L}_{\delta N^i} F. \tag{1.6}$$

Consider the simplest case of a scalar field ϕ. Taking for F the scalar field ϕ itself and then its conjugate momentum π (which ought to be a scalar density), we get from (1.6) two equations:

$$[\phi(x), \mathcal{H}_{ix'}]\delta N^{ix'} = \mathcal{L}_{\delta N^i}\phi(x) = \phi_{,i}(x)\delta N^i(x),$$

$$[\pi(x), \mathcal{H}_{ix'}]\delta N^{ix'} = \mathcal{L}_{\delta N^i}\pi(x) = (\pi(x)\delta N^i(x))_{,i}.$$

Because δN^i is arbitrary, these are actually two functional differential equations:

$$\frac{\delta \mathcal{H}_i(x')}{\delta \pi(x)} = \phi_{,i}(x)\delta(x, x'), \tag{1.7}$$

$$\frac{\delta \mathcal{H}_i(x')}{\delta \phi(x)} = -\pi_{,i}(x)\delta(x, x') - \pi(x)\delta_{,i}(x, x').$$

The system (1.7) has the unique solution

$$\mathcal{H}_i(x) = \pi(x)\phi_{,i}(x). \tag{1.8}$$

In the same way, one may find[1] the supermomentum of the metric field,

$$\mathcal{H}_i(x) = -2\pi_i{}^j{}_{|j}(x). \tag{1.9}$$

Because the Lie derivative of a tensor density is always linear in this density, the supermomentum is always linear in the field momentum. Each of the two supermomenta (1.8), (1.9), as well as their sum, satisfies the closing relation (1.3). The supermomentum (1.9) belongs to pure geometrodynamics, the total super-

momentum (1.8) + (1.9) to geometrodynamics driven by a scalar field.

The closing relation (1.3) is thus satisfied automatically. The next closing relation is also easy to take care of. Long time ago, Dirac showed[4] that Eq. (1.4) tells us merely that the super-Hamiltonian $\mathcal{H}(x)$ is a scalar density. The task of representing the "group" of deformations thus reduces to the task of constructing a functional $\mathcal{H}(x)$ from the canonical variables, which is a scalar density and satisfies the closing relation (1.5).

We say that the field ϕ^A, π_A has a nonderivative gravitational coupling, if the total super-Hamiltonian $\mathcal{H}^{(T)}$ of this field interacting with geometry falls into two parts, $\mathcal{H}$ and $\mathcal{H}^{(M)}$, the first of which depends only on the geometrodynamical variables g_{ij}, π^{ij}, and the second of which depends on the field variables ϕ^A, π_A and the metric g_{ij}, but not on the geometrodynamic momentum π^{ij}:

$$\mathcal{H}^{(T)}=\mathcal{H}[g_{ij},\pi^{ij}]+\mathcal{H}^{(M)}[\phi^A,\pi_A,g_{ij}]. \tag{1.10}$$

Under such circumstances one may show[1] that the geometrodynamical super-Hamiltonians $\mathcal{H}$ close into the geometrodynamical supermomentum (1.9), and the matter-field super-Hamiltonians $\mathcal{H}^{(M)}$ close to the matter–field super-momentum $\mathcal{H}^{(M)}{}_i$ according to the same relation (1.5). Moreover, the Poisson brackets may be restricted in the first case only to the geometrodynamical variables, and in the second case only to the matter–field variables. The problem of finding the geometrodynamical super-Hamiltonian thus nicely decouples from the problem of finding the super-Hamiltonian $\mathcal{H}^{(M)}$ of the sources.

Finally, it is easy to show that the geometrodynamical super-Hamiltonian must be purely local in the momentum π^{ij}, i.e., a function of the momentum $\pi^{ij}(x)$ taken at the same point x at which the super-Hamiltonian is evaluated. This follows from the fact that the intrinsic metric $g_{ij}(x)$ of a hypersurface $X^L(x)$ embedded in an arbitrary Riemannian space–time changes under the normal deformation $\delta N(x)$ according to the formula[1]

$$\delta_\perp g_{ij}(x)=K_{ij}(x)\delta N(x), \tag{1.11}$$

with $K_{ij}(x)$ characterizing the extrinsic curvature of the hypersurface,[5]

$$K_{ij}=2\frac{\nabla n_L}{\partial x^i}X^L_j. \tag{1.12}$$

In Hamiltonian geometrodynamics, $\delta_\perp g_{ij}$ is given by the dynamical rule (1.2) with $\delta N^i=0$, and the comparison with Eq. (1.11) yields the condition

$$\frac{\delta\mathcal{H}(x')}{\delta\pi^{ij}(x)}=[g_{ij}(x),\mathcal{H}(x')]=K_{ij}(x)\delta(x,x'). \tag{1.13}$$

Because of the δ- function on the right-hand side of Eq. (1.13), $\mathcal{H}(x)$ must be local in the momentum $\pi^{ij}(x)$.

A similar reasoning shows that the scalar field super-Hamiltonian $\mathcal{H}^{(\phi)}(x)$ must be local in the scalar field momentum $\pi(x)$. The change of an arbitrary space–time scalar $\phi(X^L)$ under the normal deformation $\delta N(x)$ is

$$\delta\phi(x)=K(x)\delta N(x), \tag{1.14}$$

with the normal scalar field velocity $K(x)$ introduced as

$$K(x)\equiv n^L(x)\phi_{,L}(X^K(x)). \tag{1.15}$$

Because the same change may be determined from the dynamical rule (1.2),

$$\delta_\perp\phi(x)=[\phi(x),\mathcal{H}^{(T)}{}_{x'}]\delta N^{x'}$$

$$=[\phi(x),\mathcal{H}^{(\phi)}{}_{x'}]\delta N^{x'}=\frac{\delta\mathcal{H}^{(\phi)}{}_{x'}}{\delta\pi(x)}\delta N^{x'},$$

we get by comparison

$$\frac{\delta\mathcal{H}^{(\phi)}(x')}{\delta\pi(x)}=K(x)\delta(x,x'). \tag{1.16}$$

Therefore, $\mathcal{H}^{(\phi)}(x)$ must be a function of $\pi(x)$ rather than a functional.

The scalar field ϕ and the metric field ${}^4g_{lk}$ are in this respect exceptional. The super-Hamiltonian of any other tensor field is necessarily nonlocal. Fortunately, this nonlocality is of a simple kind. The super-Hamiltonian additively splits into a local and a nonlocal parts. The nonlocal part is fixed once for all by purely kinematical considerations and turns out to be a spatial divergence of a vector density linear in the field momenta. It determines how the normal and tangential projections of the field behave under a "rotation" of the hypersurface. The true dynamical meaning is carried only by the local part of the super-Hamiltonian, which determines how the field changes under a "translation" of the hypersurface. We shall investigate the dynamics of general tensor fields in another paper. Here, we pay attention to the only two fields with local super-Hamiltonians: the scalar field ϕ and the metric field g_{ij}.

2. LEGENDRE TRANSFORMATION

Rather than trying to determine the super-Hamiltonian $\mathcal{H}(x)[\phi^A,\pi_A]$ from the closing relation (1.5) directly, it is easier to pass first to the super-Lagrangian $\mathcal{L}(x)[\phi^A,K^A]$ by the Legendre dual transformation and determine $\mathcal{L}(x)[\phi^A,K^A]$ from the transformed closing relation. The Legendre transformation takes on a simple form once we know that $\mathcal{H}(x)$ is local in the momenta $\pi_A(x)$,

$$\phi^A(x),\pi_A(x)\to\phi^A(x),K^A(x), \tag{2.1}$$

$$\pi_A(x)=\frac{\partial\mathcal{L}(x)}{\partial K^A(x)},K^A(x)=\frac{\partial\mathcal{H}(x)}{\partial\pi_A(x)}, \tag{2.2}$$

$$\mathcal{H}(x)=\pi_A(x)K^A(x)-\mathcal{L}(x), \tag{2.3}$$

$$\mathcal{L}(x)=\pi_A(x)K^A(x)-\mathcal{H}(x). \tag{2.4}$$

The Legendre formula

$$\left.\frac{\delta\mathcal{H}(x)}{\delta\phi^A(x')}\right|_{\pi_B\text{ fixed}}=-\left.\frac{\delta\mathcal{L}(x)}{\delta\phi^A(x')}\right|_{K^B\text{ fixed}} \tag{2.5}$$

helps us to transform the closing relation (1.5). Starting from the definition of the Poisson bracket

$$[\mathcal{H}(x),\mathcal{H}(x')]\equiv\frac{\delta\mathcal{H}(x)}{\delta\phi^{Ax''}}\frac{\delta\mathcal{H}(x')}{\delta\pi_{Ax''}}-(x\rightleftarrows x')$$

and recalling the locality of $\mathcal{H}$,

$$\frac{\delta\mathcal{H}(x')}{\delta\pi_A(x'')}=\frac{\partial\mathcal{H}(x')}{\partial\pi_A(x')}\delta(x',x''),$$

we get

$$\frac{\delta \mathcal{L}(x)}{\delta \phi^A(x')} K^A(x') + \mathcal{H}^i(x)\left[\phi^A, \frac{\partial \mathcal{L}}{\partial K^A}\right] \delta_{,i}(x,x') - (x \rightarrow\!\!\!\!\!- \; x') = 0. \tag{2.6}$$

The supermomentum $\mathcal{H}^i$ is always linear in the field momentum $\pi_A = \partial \mathcal{L}/\partial K^A$. This has two important consequences. First, Eq. (2.6) is linear in the super-Lagrangian. Second, it does not change its form under the velocity inversion

$$K^A \rightarrow -K^A. \tag{2.7}$$

Therefore, if $\mathcal{L}[\phi^A, K^A]$ satisfies Eq. (2.6), $\mathcal{L}[\phi^A, -K^A]$ also satisfies it. Furthermore, because Eq. (2.6) is a linear equation for $\mathcal{L}$, the even and odd velocity parts of $\mathcal{L}$,

$$\mathcal{L}^{\pm} \equiv \tfrac{1}{2}(\mathcal{L}[\phi^A, K^A] \pm \mathcal{L}[\phi^A, -K^A]), \tag{2.8}$$

satisfy Eq. (2.6) separately. This is the main simplification achieved by the Legendre transformation: the original closing relation (1.5) was quadratic in the super-Hamiltonian and it therefore mixed the even and odd momentum parts of $\mathcal{H}$.

3. PURE GEOMETRODYNAMICS

In Ref. 1, the complications associated with the mixing of $\mathcal{H}^+$ with $\mathcal{H}^-$ were circumvented by assuming that geometrodynamics is time-reversible, so that the super-Hamiltonian may be chosen since the beginning as an even function $\mathcal{H}^+$ of the momentum. Within the Lagrangian approach, we are able to remove this additional assumption and *prove* that geometrodynamics is time-reversible from the closing relation (2.6) itself.[6] This we do by proving that $\mathcal{L}^-(x)$ must have the form

$$\mathcal{L}^-(x) = \frac{\delta \Lambda[{}^3\mathcal{G}]}{\delta g_{ij}(x)} K_{ij}(x), \tag{3.1}$$

where $\Lambda[{}^3\mathcal{G}]$ is an arbitrary functional of the three-geometry ${}^3\mathcal{G} \equiv g_{ij}/\mathrm{Diff}(\mathcal{M}^3)$. The super-Langrangian (3.1) is dynamically irrelevant, because it satisfies the Lagrange equations

$$\frac{\delta_\perp}{\delta N(x')}\left(\frac{\partial \mathcal{L}(x)}{\partial K_{ij}(x)}\right) - \frac{\delta \mathcal{L}(x')}{\delta g_{ij}(x)} = 0 \tag{3.2}$$

identically. The evolution of geometry is then governed by the even part $\mathcal{L}^+$ of $\mathcal{L}$. The Lagrange equations (3.2) with $\mathcal{L} = \mathcal{L}^+$ remain invariant under the time-reversal

$$\delta N(x) \rightarrow -\delta N(x), \quad K_{ij}(x) \rightarrow -K_{ij}(x), \tag{3.3}$$

which means that pure geometrodynamics is reversible. The actual structure of $\mathcal{L}^+[g_{ij}, K_{ij}]$ is then determined from the closing relation (2.6).

Let us first cast the closing relation (2.6) specialized to pure geometrodynamics into an advantageous form

$$\frac{\delta \mathcal{L}(x)}{\delta g_{ij}(x')} K_{ij}(x') + 2\frac{\partial \mathcal{L}(x)}{\partial K_{ij}(x)} \delta_{|ij}(x,x') - (x \rightarrow\!\!\!\!\!- \; x') = 0. \tag{3.4}$$

We have arrived at Eq. (3.4) by using the identity

$$\mathcal{H}^i(x)\delta_{,i}(x,x') - (x \rightarrow\!\!\!\!\!- \; x') = 2\pi^{ij}(x)\delta_{|ij}(x,x') - (x \rightarrow\!\!\!\!\!- \; x') \tag{3.5}$$

which is satisfied by the geometrodynamical supermomentum (1.9).

We solve Eq. (3.4) by expanding $\mathcal{L}$ into a power series

$$\mathcal{L} = \sum_{n=0}^{\infty} {}^{(n)}G^{i_1 j_1 \cdots i_n j_n}[g^{kl}] K_{i_1 j_1} \cdots K_{i_n j_n} \tag{3.6}$$

and comparing the coefficients belonging to different powers of K_{ij}. The comparison is easy once all spatial derivatives are shifted to the δ-functions, as we have done in Eq. (3.4) by means of the identity (3.5). One can also see directly that Eq. (3.4) does not mix the even ($n = 0, 2, 4, \cdots$) and odd ($n = 1, 3, 5, \cdots$) coefficients ${}^{(n)}G^{i_1 j_1 \cdots i_n j_n}$, corresponding to the fact that $\mathcal{L}^+$ and $\mathcal{L}^-$ satisfy Eq. (3.4) separately.

The coefficients ${}^{(n)}G^{i_1 j_1 \cdots i_n j_n}$ are some functionals ${}^{(n)}G^{i_1 j_1 \cdots i_n j_n}[g_{kl}]$ of the metric g_{kl}. Because $\mathcal{H}$ is a scalar density, K_{ij} is a tensor, and π^{ij} is a tensor density, $\mathcal{L}$ must be a scalar density and all coefficients ${}^{(n)}G^{i_1 j_1 \cdots i_n j_n}$ must be tensor densities. Because K_{ij} is a symmetrical tensor, the coefficients ${}^{(n)}G^{i_1 j_1 \cdots i_n j_n}$ are assumed to be symmetrical in each pair $i_a j_a$ of indices and also symmetrical with respect to an interchange $i_a j_a \rightarrow\!\!\!\!\!- \; i_b j_b$ of pairs. Finally, we shall often omit the superscript ${}^{(n)}$, because the order of the coefficient ${}^{(n)}G^{i_1 j_1 \cdots i_n j_n}$ is indicated by the number of its indices.

A. $\mathcal{L}^-$ regained

Put $K_{ij} = 0$ in Eq. (3.4), collecting thus all terms which do not contain K_{ij},

$$G^{ij}(x)\delta_{|ij}(x,x') - G^{ij}(x')\delta_{|i'j'}(x',x) = 0. \tag{3.7}$$

The consequences of the distribution equation (3.7) are best extracted when we multiply it by two scalar test functions, $a(x)$ and $b(x')$, and integrate over x and x'. When integrating by parts, remember that $\delta(x,x')$ is a scalar in the first and a scalar density in the second argument, and that $G^{ij}(x)$ is a tensor density. After discarding some divergences, we obtain the equation

$$\int d^3x\, a(2b_{,i}G^{ij}{}_{|j} + bG^{ij}{}_{|ij}) = 0. \tag{3.8}$$

Because $a(x)$ is arbitrary,

$$2b_{,i}G^{ij}{}_{|j} + bG^{ij}{}_{|ij} = 0. \tag{3.9}$$

But $b(x)$ and $b_{,i}(x)$ at a given point x are also arbitrary, so that

$$G^{ij}{}_{|j} = 0. \tag{3.10}$$

Going backwards, Eq. (3.10) ensures that Eq. (3.8) and therefore the distribution equation (3.7) are satisfied. Equation (3.10) is thus the only conclusion we may draw from Eq. (3.7).

Next we collect those terms in Eq. (3.4) which are quadratic in K_{ij}, getting

$$\frac{\delta G^{ij}(x)}{\delta g_{kl}(x')} K_{ij}(x) K_{kl}(x') + 6G^{ijklmn}(x)K_{ij}(x)K_{kl}(x)\delta_{|mn}(x,x') - (x \rightarrow\!\!\!\!\!- \; x') = 0. \tag{3.11}$$

Equation (3.11) must hold for an arbitrary $K_{ij}(x)$. To follow its consequences step by step, write $K_{ij}(x)$ formally as a product of a scalar $k(x)$ and a tensor $\bar{K}_{ij}(x)$,

$$K_{ij}(x) = k(x)\bar{K}_{ij}(x). \tag{3.12}$$

Keep first the tensor $\bar{K}_{ij}(x)$ fixed and consider the

scalar $k(x)$ to be arbitrary. Later on, allow the tensor $\bar{K}_{ij}(x)$ to vary again.

We introduce the abbreviations

$$A(x,x') \equiv 6H^{mn}(x)\delta_{|mn}(x,x'), \tag{3.13}$$

$$H^{mn}(x) \equiv G^{ij\,kl\,mn}(x)\,\bar{K}_{ij}(x)\bar{K}_{kl}(x), \tag{3.14}$$

$$B(x,x') \equiv \left(\frac{\delta G^{ij}(x)}{\delta g_{kl}(x')} - \frac{\delta G^{kl}(x')}{\delta g_{ij}(x)}\right)\bar{K}_{ij}(x)\bar{K}_{kl}(x'). \tag{3.15}$$

Equation (3.11) then assumes the form

$$A(x,x')k(x)k(x) - A(x',x)k(x')k(x') + B(x,x')k(x)k(x') = 0. \tag{3.16}$$

Notice that H^{mn} is symmetrical, and

$$B(x',x) = -B(x,x'). \tag{3.17}$$

Equation (3.16) holds for an arbitrary scalar $k(x)$ and some distribution coefficients $A(x,x')$, $B(x,x')$ with the symmetry (3.17). According to the Lemma 1 proved in the Appendix, the coefficients $A(x,x')$, $B(x,x')$ must satisfy the relations

$$A^{(+)}(x,x') = A(x)\delta(x,x'), \tag{3.18}$$

$$-2A^{(-)}(x,x') = B(x,x'), \tag{3.19}$$

with

$$A^{(+)}(x,x') \equiv \tfrac{1}{2}(A(x,x') \pm A(x',x)),$$

$$A(x) \equiv \int d^3x'\, A^{(+)}(x,x').$$

For $A(x,x')$ given by Eq. (3.13), the relation (3.18) gives

$$H^{mn}(x)\delta_{|mn}(x,x') + H^{mn}(x',x)\delta_{|m'n'}(x',x) = H^{mn}{}_{|mn}(x)\delta(x,x'). \tag{3.20}$$

Multiply Eq. (3.20) by two scalar test functions, $a(x)$ and $b(x')$, and integrate it over x and x'. Recalling that H^{mn} is a tensor density, we again discard some divergences and obtain

$$\int d^3x\; 2a(b_{|mn}H^{mn} + b_{|m}H^{mn}{}_{|n}) = 0.$$

Because $a(x)$ is arbitrary, and also $b_{|m}(x)$ and $b_{|\underline{mn}}(x) \equiv \frac{1}{2}(b_{|mn}(x) + b_{|nm}(x))$ at a given point x are arbitrary, we conclude that

$$H^{mn} = 0. \tag{3.21}$$

Returning to the definition (3.14) of H^{mn} and using now the arbitrariness of $\bar{K}_{ij}$, we see that Eq. (3.21) really means that

$$G^{ij\,kl\,mn}(x) = 0. \tag{3.22}$$

Equation (3.21) implies that the coefficient $A(x,x')$ given by Eq. (3.13) vanishes. According to Eq. (3.19), the coefficient $B(x,x')$ must then also vanish. Because $B(x,x')$ has the form (3.15) and $\bar{K}_{ij}(x)$ may again be considered as an arbitrary tensor, we apply the Lemma 2 proved in the Appendix and conclude that

$$\frac{\delta G^{ij}(x)}{\delta g_{kl}(x')} - \frac{\delta G^{kl}(x')}{\delta g_{ij}(x)} = F^{ij\,kl}(x)\,\delta(x,x'), \tag{3.23}$$

where $F^{ij\,kl}$ is symmetrical in ij and kl and antisymmetrical in the interchange $ij \longleftrightarrow kl$ of pairs. One can check that Eqs. (3.22) and (3.23) are the only two consequences one may draw from Eq. (3.11).

The closing relation (3.4) has such a structure that once an odd (even) coefficient ${}^{(n)}G$ vanishes,

$${}^{(n)}G^{i_1j_1\cdots i_nj_n} = 0, \tag{3.24}$$

all subsequent odd (even) coefficients ${}^{(n+2)}G$, ${}^{(n+4)}G, \cdots$ are forced to vanish as well. Indeed, when Eq. (3.24) holds, the only term of the order $n+1$ remaining in Eq. (3.4) is

$$H^{ij}(x)\delta_{|ij}(x,x') - H^{ij}(x')\delta_{|i'j'}(x',x) = 0, \tag{3.25}$$

where

$$H^{ij} \equiv {}^{(n+2)}G^{i_1j_1\cdots i_{n+1}j_{n+1}\,ij}K_{i_1j_1}\cdots K_{i_{n+1}j_{n+1}}. \tag{3.26}$$

Equation (3.25) has the same form as Eq. (3.7) and it thus leads to the same consequence,

$$H^{ij}{}_{|j} = 0. \tag{3.27}$$

The term H^{ij} in Eq. (3.27) depends, however, on an arbitrary tensor K_{ij}, whereas the term G^{ij} in Eq. (3.10) was K_{ij}-independent. This allows us to conclude from Eq. (3.27) that the coefficient ${}^{(n+2)}G$ must vanish,

$${}^{(n+2)}G^{i_1j_1\cdots i_{n+2}j_{n+2}} = 0.$$

We can see that by substituting the expression (3.26) into Eq. (3.27),

$${}^{(n+2)}G^{i_1j_1\cdots i_{n+1}j_{n+1}\,ij}K_{i_1j_1}\cdots K_{i_{n+1}j_{n+1}}$$

$$+ (n+1)\,{}^{(n+2)}G^{i_1j_1\cdots i_nj_n\,i_{n+1}j_{n+1}\,ij}K_{i_1j_1}\cdots K_{i_nj_n}K_{i_{n+1}j_{n+1}|j} = 0,$$

and realizing that K_{ij} and $K_{ij|k}$ are both arbitrary at a given point x.

Because the coefficient ${}^{(3)}G$ vanishes according to Eq. (3.22), all subsequent odd coefficients must vanish and $\mathcal{L}^-$ thereby reduces to the term linear in K_{ij}. The coefficient G^{ij} of this term is subject to the conditions (3.10) and (3.23). Equation (3.10) tells us that G^{ij} is divergence-free. Equation (3.23) prescribes a definite form to the "functional curl" $\delta G^{ij}(x)/\delta g_{kl}(x') - \delta G^{kl}(x')/\delta g_{ij}(x)$ of $G^{ij}[g_{kl}]$.

In relativity, one often generates a divergence-free functional $G^{ij}(x)[g_{kl}]$ by taking a labeling-independent functional $\Lambda[{}^3\mathcal{G}]$ and varying it with respect to the metric g_{ij}; this is how the Einstein tensor density arises from the Hilbert action. Also, if

$$G^{ij}(x) = \frac{\delta\Lambda[g_{kl}]}{\delta g_{ij}(x)} \tag{3.28'}$$

is divergence-free (the divergence being taken as if G^{ij} were a tensor density), the functional $\Lambda[g_{kl}]$ must be a labeling-independent scalar density, and therefore a functional of the 3-geometry ${}^3\mathcal{G}$ rather than a functional of the metric g_{kl}.[7] It is not true, however, that all divergence-free tensor densities $G^{ij}(x)[g_{kl}]$ constructed invariantly from the metric tensor g_{kl} are variational derivatives of labeling-independent functionals $\Lambda[{}^3\mathcal{G}]$. A counterexample is provided by the (symmetrical) tensor density

$$\beta^{ij} \equiv \epsilon^{ikl}R^j{}_{kl|l} \tag{3.29}$$

introduced by York.[8] If G^{ij} is a functional gradient, Eq. (3.28), its functional curl must vanish,

$$\frac{\delta G^{ij}(x)}{\delta g_{kl}(x')} - \frac{\delta G^{kl}(x')}{\delta g_{ij}(x)} = 0. \tag{3.30}$$

It is easy to check that the functional curl of β^{ij} does not vanish.

To prove that geometrodynamics is time-reversible, we must show, according to Eq. (3.1), that the coefficient G^{ij} is a functional gradient. This is not clear at this stage, because Eq. (3.23) taken by itself does not say that the functional curl of G^{ij} vanishes, but only that it has a certain structure. Fortunately, as shown by Teitelboim,[9] Eqs. (3.10) and (3.23) taken together imply that an undetermined functional $F^{ij\,kl}(x)[g_{mn}]$ on the right-hand side of Eq. (3.23) must vanish, which leads to the desired result.

One arrives to Teitelboim's conclusion by proving first the following:

Lemma: When (1) $G^{ij}(x)[g_{kl}]$ is a symmetrical tensor density constructed invariantly from the metric tensor and (2) $G^{ij}{}_{|j} = 0$, then

$$I \equiv \int d^3x' \left(\frac{\delta G^{ij}(x)}{\delta g_{kl}(x')} - \frac{\delta G^{kl}(x')}{\delta g_{ij}(x)} \right) \pounds_{\delta N^m} g_{kl}(x') = 0. \tag{3.31}$$

The assumption (1) leads to the transformation formula

$$\int d^3x' \frac{\delta G^{ij}(x)}{\delta g_{kl}(x')} \pounds_{\delta N^m} g_{kl}(x') = G^{ij}(x)[g_{kl} + \pounds_{\delta N^m} g_{kl}] - G^{ij}(x)[g_{kl}]$$
$$= \pounds_{\delta N^m} G^{ij}(x), \tag{3.32}$$

which takes care of the first term of I. The second term may be rearranged into

$$-\int d^3x' \frac{\delta G^{kl}(x')}{\delta g_{ij}(x)} \pounds_{\delta N^m} g_{kl}(x') = -\frac{\delta \int d^3x'\, G^{kl}(x') \pounds_{\delta N^m} g_{kl}(x')}{\delta g_{ij}(x)}$$
$$+ \int d^3x' G^{kl}(x') \frac{\delta \pounds_{\delta N^m} g_{kl}(x')}{\delta g_{ij}(x)}. \tag{3.33}$$

Varying the Lie derivative of g_{kl}, we get

$$\frac{\delta \pounds_{\delta N^m} g_{kl}(x')}{\delta g_{ij}(x)} = \delta^{ij}_{kl} \delta N^m(x')\, \delta_{,m'}(x'x)$$
$$+ \delta^{ij}_{km} \delta N^m{}_{,l}(x')\, \delta(x',x) + \delta^{ij}_{lm} \delta N^m{}_{,k}(x')\, \delta(x',x). \tag{3.34}$$

The last integral in Eq. (3.33) is then evaluated, yielding $-\pounds_{\delta N^m} G^{ij}(x)$. The first integral on the right-hand side of Eq. (3.34) vanishes,

$$\int d^3x\, G^{kl}(x) \pounds_{\delta N^m} g_{kl}(x) = \int d^3x\, G^{kl} (\delta N_{k|l} + \delta N_{l|k})$$
$$= 2\int d^3x\, G^{kl} \delta N_{k|l} = 2\int d^3x\, (G^{kl} \delta N_k)_{|l} - 2\int d^3x\, G^{kl}{}_{|l} \delta N^k = 0,$$

because $(G^{kl}\delta N_k)_{|l}$ is an ordinary divergence and $G^{kl}{}_{|l}$ vanishes by assumption (2). Collecting all terms of I together, we see that I vanishes.

When we substitute into Eq. (3.31) the form which Eq. (3.23) prescribes for the functional curl of G^{ij}, we get

$$F^{ij\,kl} \delta N_{k|l} = 0. \tag{3.35}$$

Because $\delta N_{k|l}$ at a given point x is arbitrary, $F^{ij\,kl}$ must vanish. Equation (3.23) then ensures (disregarding the global problems which might arise if the space of functions was multiply connected) that G^{ij} is a functional gradient (3.28). Because of Eq. (3.10), Λ depends only on ${}^3\mathcal{G}$. This finally shows that $\mathcal{L}^-$ has the desired form (3.1).

B. $\mathcal{L}^+$ regained

Turning to the even part of the super-Lagrangian, we collect in Eq. (3.4) all terms linear in K_{ij}, getting

$$A^{ij}(x,x')K_{ij}(x') - A^{ij}(x',x)K_{ij}(x) = 0, \tag{3.36}$$

with

$$A^{ij}(x,x') = \frac{\delta G(x)}{\delta g_{ij}(x')} - 4G^{ij\,kl}(x')\delta_{|k'l'}(x',x). \tag{3.37}$$

Varying Eq. (3.36) with respect to $K_{ij}(x'')$,

$$A^{ij}(x,x')\delta(x',x'') - A^{ij}(x',x)\delta(x,x'') = 0,$$

and integrating the resulting equation over x', we learn that $A^{ij}(x,x'')$ must be proportional to the δ-function,

$$A^{ij}(x,x'') = F^{ij}(x)\, \delta(x,x''),$$
$$F^{ij}(x) \equiv \int d^3x'\, A^{ij}(x',x). \tag{3.38}$$

For the coefficient (3.37), Eq. (3.38) gives

$$\frac{\delta G(x)}{\delta g_{ij}(x')} = F^{ij}(x)\, \delta(x,x') + 4G^{ij\,kl}(x')\, \delta_{|k'l'}(x',x). \tag{3.39}$$

Equation (3.39) was analyzed in detail in Ref. 1. We shall repeat here only the main results.

First, we see that $G(x)$ must be a function of the metric tensor $g_{ij}(x)$ and its first and second partial derivatives, because otherwise the variation of $G(x)$ with respect to $g_{ij}(x')$ would yield higher derivatives of the δ-function than the second ones. Because $G(x)$ is a form-invariant scalar density, we know[10] that it may depend on $g_{ij}(x)$ and its derivatives only through the metric tensor $g_{ij}(x)$ and the Riemann tensor $R_{ijkl}(x)$. In a three-dimensional space, the Riemann tensor is expressible by means of the Ricci tensor $R_{ij}(x)$ and the metric tensor $g_{ij}(x)$, so that finally

$$G(x) = G(g_{ij}(x), R_{ij}(x)). \tag{3.40}$$

Varying Eq. (3.40) with respect to g_{ij}, we get

$$\delta G = (\phi^{ij} + \tfrac{1}{2}R^i{}_{kl}{}^j \phi^{kl} + \tfrac{1}{4}R^i_k \phi^{kj} + \tfrac{1}{4}R^j_k \phi^{ki}) \delta g_{ij}$$
$$+ \tfrac{1}{4}(\phi^{ik}g^{jl} + \phi^{il}g^{jk} + \phi^{jk}g^{il} + \phi^{jl}g^{ik}$$
$$- 2\phi^{ij}g^{kl} - 2\phi^{kl}g^{ij}) \delta g_{ij|kl}, \tag{3.41}$$

with

$$\phi^{ij} = \frac{\partial G(g_{kl}, R_{kl})}{\partial g_{ij}}, \quad \phi^{ij} = \frac{\partial G(g_{kl}, R_{kl})}{\partial R_{ij}}. \tag{3.42}$$

The variation of G may be also determined from Eq. (3.39), with the result

$$\delta G = (F^{ij} + 4G^{ij\,kl}{}_{|kl}) \delta g_{ij}$$
$$+ 8G^{ij\,kl}{}_{|l} \delta g_{ij|k} + 4G^{ij\,kl} \delta g_{ij|kl}. \tag{3.43}$$

Because δg_{ij}, $\delta g_{ij|k}$, and $\delta g_{ij|kl} = \frac{1}{2}(\delta g_{ij|kl} + \delta g_{ij|lk})$ may be chosen at a given point as arbitrary and mutually independent quantities, the comparison of expressions (3.41) and (3.43) yields three sets of equations. The first set,

$$G^{ij\,kl} = \tfrac{1}{16}(\phi^{ik}g^{jl} + \phi^{il}g^{jk} + \phi^{jk}g^{il} + \phi^{jl}g^{ik} - 2\phi^{ij}g^{kl} - 2\phi^{kl}g^{ij}), \tag{3.44}$$

obtained by comparing the coefficients of $\delta g_{ij|kl}$, determines the "supermetric" $G^{ij\,kl}$ in terms of the "potential" G. The second set,

$$G^{ij\,kl}{}_{|l} = 0, \tag{3.45}$$

obtained by comparing the coefficients of $\delta g_{ij|k}$, imposes then a condition on the potential G through the supermetric $G^{ij\,kl}$. This condition fixes the potential and through it the supermetric up to two arbitrary constants. The third set of equations, obtained by comparing the coefficients of δg_{ij}, then turns into an identity.

The condition for G is obtained when substituting the supermetric (3.44) into Eq. (3.45),

$$\phi^{ij|k} - \phi^{ik|j} + g^{ij}\phi^{kl}{}_{|l} - g^{ik}\phi^{jl}{}_{|l} = 0. \tag{3.46}$$

Contracting Eq. (3.46) in the indices ij, we get

$$\bar{\phi}^{kl}{}_{|l} = 0, \quad \bar{\phi}^{kl} \equiv \phi^{kl} + \phi g^{kl}. \tag{3.47}$$

According to its definition (3.42), ϕ^{ij} is a tensor density constructed invariantly from the metric tensor and its first and second derivatives. So is $\bar{\phi}^{kl}$, which is in addition divergence-free. Thanks to a theorem due to Lovelock[11], $\bar{\phi}^{kl}$ must be a linear combination of the Einstein tensor density $g^{1/2}(R^{kl} - \frac{1}{2}Rg^{kl})$ and the metric tensor density $g^{1/2}g^{kl}$. Using the full Eq. (3.46) again, we learn that ϕ^{ij} must be actually proportional to the metric density,

$$\phi^{ij} = (2\kappa)^{-1}g^{1/2}g^{ij}, \quad \kappa = \text{const.} \tag{3.48}$$

Substituting Eq. (3.48) into Eq. (3.44), we get the supermetric

$$G^{ij\,kl} = (16\kappa)^{-1}g^{1/2}(g^{ik}g^{jl} + g^{il}g^{jk} - 2g^{ij}g^{kl}). \tag{3.49}$$

Returning to the definition (3.42) of ϕ^{ij}, we may integrate Eq. (3.48) for G, picking up a second constant λ,

$$G = (2\kappa)^{-1}g^{1/2}(R - 2\lambda). \tag{3.50}$$

Linear terms in Eq. (3.4) thus determine the super-Lagrangian up to the terms quadratic in K_{ij}. Collect now the terms containing the product of three K_{ij} in Eq. (3.4). Because the supermetric (3.49) is purely local in the metric g_{kl}, $\delta\mathcal{L}(x)/\delta g_{ij}(x')$ is proportional in the third order to the δ-function, and the first expression in Eq. (3.4) gives no contribution after the commutation $(x \leftrightarrow x')$. We thus get

$$H^{ij}(x)\delta_{|i'j'}(x, x') - H^{ij}(x')\,\delta_{|i'j'}(x', x) = 0,$$

with

$$H^{ij} \equiv G^{i_1j_1\,i_2j_2\,i_3j_3\,ij}K_{i_1j_1}K_{i_2j_2}K_{i_3j_3}.$$

The argument given in Sec. 3A then shows that the coefficient ${}^{(4)}G$ and all further even coefficients ${}^{(6)}G$, ${}^{(8)}G, \ldots$ must vanish. The even part of the gravitational super-Lagrangian is therefore quadratic in the velocity K_{ij}, being equal to

$$\mathcal{L}^{+} = G^{ij\,kl}K_{ij}K_{kl} + (2\kappa)^{-1}g^{1/2}(R - 2\lambda), \tag{3.51}$$

with the supermetric (3.49). Remembering that K_{ij} is the -2 multiple of the extrinsic curvature, one recognizes in the expression (3.51) the ADM super-Lagrangian.[12] This completes the proof of the statement that the Einsteinian geometrodynamics is the only representation of the "group" of deformations of spacelike hypersurface embedded in a Riemannian space—time using the metric g_{ij} as the sole configuration variable.

4. SCALAR FIELD DYNAMICS

The closing relation (3.4) determines the gravitational super-Lagrangian completely—up to the constants κ and λ which are interpreted as the gravitational constant and the cosmological constant, and up to the trivial part $\mathcal{L}^{-}$ which has no influence on dynamics and may be thus safely discarded. This shows the uniqueness of Einsteinian geometrodynamics. The uniqueness breaks down when we pass from the metric field to its sources. The closing relation still ensures that the source-field dynamics may be consistently interpreted as taking place in a single Riemannian space—time, but there are many alternative dynamics compatiable with this requirement. We illustrate the situation on the scalar field (which is, as we have seen, the only field besides the metric field possessing a super-Hamiltonian local in the field momentum). We recover from the closing relation (2.6) itself all different ways in which the dynamics of the scalar field with nonderivative gravitational coupling may procede.

The field-part $\mathcal{H}^{(\phi)}$ of the total super-Hamiltonian satisfies the closing relation (1.5) separately, with the Poisson brackets being confined to the scalar field variables ϕ and π. Pass to the super-Lagrangian $\mathcal{L}^{(\phi)}$, performing the Legendre dual transformation (2.1)–(2.4) in the variables ϕ, π:

$$\phi(x), \pi(x) \to \phi(x), K(x),$$
$$K(x) = \frac{\partial\mathcal{H}^{(\phi)}(x)}{\partial\pi(x)}, \quad \mathcal{L}^{(\phi)}(x) = \pi(x)K(x) - \mathcal{H}^{(\phi)}(x). \tag{4.1}$$

The scalar field super-Lagrangian $\mathcal{L}^{(\phi)}$ may depend, besides the scalar field amplitude ϕ and the associated normal velocity K, only on the undifferentiated metric tensor g_{ij}. We know that $\mathcal{L}^{(\phi)}$ must be a scalar density with respect to the spatial transformations. This severely limits the manner in which the derivatives of ϕ may enter into $\mathcal{L}^{(\phi)}$. Indeed, $\mathcal{L}^{(\phi)}$ must have the form

$$\mathcal{L}^{(\phi)} = g^{1/2}L(\phi, \psi, K), \tag{4.2}$$

where L is an arbitrary function of the three scalars ϕ, K, and

$$\psi \equiv g^{ij}\phi_{,i}\phi_{,j}. \tag{4.3}$$

Higher derivatives of ϕ than the first ones cannot enter the super-Lagrangian $\mathcal{L}^{(\phi)}$, because the formation of a scalar density would require the use of covariant derivatives. This would introduce the differentiated metric tensor into $\mathcal{L}^{(\phi)}$, contrary to the assumption of the nonderivative coupling. The scalar ψ is the only scalar which can be formed from the first derivatives of ϕ.

The scalar field supermomentum is given by Eq. (1.8) and the transformed closing relation (2.6) takes the form

$$\frac{\delta L(x)}{\delta\phi(x')}K(x')+\frac{\partial L(x)}{\partial K(x)}\phi^{,i}(x)\delta_{,i}(x,x')-(x\leftrightarrow x')=0. \quad (4.4)$$

The variational derivative of L is

$$\frac{\delta L(x)}{\delta\phi(x')}=\frac{\partial L(x)}{\partial\phi(x)}\frac{\delta\phi(x)}{\delta\phi(x')}+\frac{\partial L(x)}{\delta\psi(x)}\frac{\delta\psi(x)}{\delta\phi(x')}$$

$$=\frac{\partial L(x)}{\partial\phi(x)}\delta(x,x')+2\frac{\partial L(x)}{\partial\psi(x)}\phi^{,i}(x)\delta_{,i}(x,x'). \quad (4.5)$$

Substituting it into Eq. (4.4), we get

$$A^i(x)\delta_{,i}(x,x')-A^i(x')\delta_{,i}(x',x)=0, \quad (4.6)$$

with

$$A^i\equiv\phi^{,i}\left(\frac{\partial L}{\partial K}+2\frac{\partial L}{\partial\psi}\right). \quad (4.7)$$

Multiplying the distribution equation (4.6) by a test function $a(x')$ and integrating it over x',

$$aA^i{}_{,i}+2a_{,i}A^i=0,$$

we learn that A^i must vanish, because a and $a_{,i}$ at a given point x are arbitrary. From Eq. (4.7) we get the equation

$$\frac{\partial L}{\partial K}+2\frac{\partial L}{\partial\psi}=0, \quad (4.8)$$

which tells us that L must have the form

$$L=L(\phi,\psi\text{-}K^2). \quad (4.9)$$

An arbitrary function L of two variables, ϕ and ψ, thus generates a permissable dynamics through the super-Lagrangian (4.2), (4.9), with ψ given by Eq. (4.3). We see that the scalar field super-Lagrangian is highly ambiguous.

From the super-Lagrangian (4.9), we can pass to the standard space–time Lagrangian ${}^4L$. The action contained in the coordinate cell with the edges $d_{(0)}X^l$, $d_{(1)}X^l$, $d_{(2)}X^l$, $d_{(3)}X^l$ may be expressed through the space–time Lagrangian ${}^4L$ as

$${}^4L\,{}^4\epsilon_{l\kappa\lambda\mu}d_{(0)}X^l d_{(1)}X^\kappa d_{(2)}X^\lambda d_{(3)}X^\mu, \quad (4.10)$$

where ${}^4\epsilon_{l\kappa\lambda\mu}$ is the space–time Levi–Civita pseudotensor. Adapt the cell to a given hypersurface, generating $d_{(i)}X^l$ by the edges $d_{(i)}x^j$ of a coordinate cell on the hypersurface,

$$d_{(i)}X^l=X^l_j d_{(i)}x^j, \quad (4.11)$$

and taking $d_{(0)}X^l$ to be perpendicular to the hypersurface,

$$d_{(0)}X^l=\delta N\,n^l. \quad (4.12)$$

The action contained in such a cell may be expressed through the super-Lagrangian $\mathcal{L}^{(\phi)}$ as

$$\mathcal{L}^{(\phi)}(x)\delta N(x)\,d^3x=L(x)\delta N(x)\epsilon_{klm}d_{(1)}x^k d_{(2)}x^l d_{(3)}x^m, \quad (4.13)$$

where ϵ_{klm} is the Levi–Civita pseudotensor on the hypersurface. We have

$${}^4\epsilon_{l\kappa\lambda\mu}n^l X^\kappa_k X^\lambda_l X^\mu_m=\epsilon_{klm}. \quad (4.14)$$

Comparing the two expressions for the action, (4.10) and (4.13), and taking into account the formulas (4.11)–(4.12), we see that

$${}^4L=L. \quad (4.15)$$

Knowing that $K(x)$ is the normal scalar field velocity (1.15) and $\phi_{,i}$ is the derivative of ϕ along the hypersurface, $\phi_{,i}=\phi_{,l}X^l_i$, we express $\psi-K^2$ in terms of space–time tensors,

$$\psi-K^2=(g^{ik}X^l_i X^\kappa_k-n^l n^\kappa)\phi_{,l}\phi_{,\kappa}={}^4g^{l\kappa}\phi_{,l}\phi_{,\kappa}. \quad (4.16)$$

In this way, we recover the space–time Lagrangian

$${}^4L=L(\phi,{}^4g^{l\kappa}\phi_{,l}\phi_{,\kappa}). \quad (4.17)$$

The closing relation (4.4) ensured that L depends on K and ψ only through the combination $\psi-K^2$, which is a space–time scalar. This illustrates the connection between the geometrodynamical and space–time approaches. The ambiguity of the scalar field dynamics, given by an arbitrary function L of two arguments, is easily understood from the space–time viewpoint. The space–time Lagrangian (4.17) is the most general space–time scalar which may be formed from the scalar field ϕ and its derivatives without involving the derivatives of the metric tensor.

ACKNOWLEDGMENT

A lively correspondence with Claudio Teitelboim on the subject of this paper is gratefully acknowledged.

APPENDIX

We prove here two lemmas which were needed in Sec. 3A.

Lemma 1: Let the equation

$$A(x,x')k(x)k(x)-A(x',x)k(x')k(x')+B(x,x')k(x)k(x')=0 \quad (A1)$$

hold for an arbitrary field $k(x)$ and some distribution coefficients $A(x,x')$ and

$$B(x,x')=-B(x',x). \quad (A2)$$

Then

$$A^{(+)}(x,x')=A(x)\delta(x,x'), \quad (A3)$$

$$-2A^{(-)}(x,x')=B(x,x'), \quad (A4)$$

with

$$A^{(\pm)}(x,x')\equiv\tfrac{1}{2}(A(x,x')\pm A(x',x)), \quad (A5)$$

$$A(x)\equiv\int d^3x'\,A^{(+)}(x,x').$$

Proof: Put $k(x)=1$ in Eq. (A1) and get Eq. (A4). Substitute $B(x,x')$ given by Eq. (A4) back into Eq. (A1), take the variational derivative of Eq. (A1) with respect to $k(x'')$, and put $k(x)=1$. Obtain the relation

$$A^{(+)}(x',x)\delta(x,x'')-A^{(+)}(x,x')\delta(x',x'')=0.$$

Integrate it over x and arrive thus to Eq. (A3).

Lemma 2: Let the equation

$$B^{AB}(x,x')k_A(x)k_B(x')=0 \quad (A6)$$

hold for an arbitrary field $k_A(x)$ and some distribution coefficients

$$B^{AB}(x,x') = -B^{BA}(x',x). \tag{A7}$$

Then

$$B^{AB}(x,x') = F^{AB}(x)\delta(x,x'), \tag{A8}$$

with

$$F^{AB}(x) = -F^{BA}(x) = \int d^3x'\ B^{BA}(x,x'). \tag{A9}$$

Proof: Take the second variational derivative of Eq. (A6) with respect to $k_C(x')$ and $k_D(x'')$. Get

$$B^{CD}(x,x')\delta(x,x'')\delta(x',x''') - B^{DC}(x',x)\delta(x',x'')\delta(x,x'''). \tag{A10}$$

Integrate Eq. (A10) over x''' and x. Arrive thus to Eq. (A9).

*Work supported in part by the National Science Foundation under Grant No. GP-30799X to the Princeton University.

[1]S. A. Hojman, K. Kuchař, and C. Teitelboim (unpublished).

[2]The notation of this paper follows that of Ref. 1. The Greek indices range from 0 to 3, the lower case italic indices from 1 to 3, the range of the upper case italic indices is left unspecified. The Greek indices are raised by the space–time metric ${}^4g^{\iota\kappa}$, the italic indices by the intrinsic metric g^{ij} of a hypersurface. In general, space–time tensors carry an index[4] indicating the dimension, hypersurface tensors are written without such an index (e.g., ${}^4R$ is the space–time curvature scalar, R is the curvature scalar of a hypersurface). Capital X^ι is used for the space–time coordinates, lower case x^i (the index ^{i} being often omitted) for the intrinsic coordinates of a hypersurface. Comma , denotes the partial differentiation, vertical stroke | the covariant differentiation on a hypersurface, the nabla symbol ∇ the space–time covariant differentiation. The summation convention is extended to continuous spatial labels: whenever x is written as an index, the integration over a repeated x is implied. For example, $\mathcal{H}_{ix}\delta N^{ix} \equiv \int d^3x \mathcal{H}_i(x^j)\delta N^i(x^j)$. All tensor densities used in the paper are tensor densities of weight 1. The δ-functions $\delta(x,x')$ are considered to be scalars in the first and scalar densities in the second argument. The Levi–Civita pseudotensors are written as ${}^4\epsilon_{\iota\kappa\lambda\mu}$ and ϵ_{ijk}. The symbol $(x \longleftrightarrow x')$ used in the formulas means "the same term with the spatial labels x and x' interchanged." The pound symbol $\pounds_{\delta N^i}$ denotes the Lie derivative with respect to δN^i. The square brackets indicate the functional dependence; *e.g.*, $\mathcal{H}(x)[g_{ij}, \pi^{ij}]$ means that the super-Hamiltonian at the point x of a hypersurface is a functional of the metric g_{ij} and of the geometrodynamical momentum π^{ij} on that hypersurface. The Poisson brackets are denoted by the square brackets [,].

[3]K. Kuchař, "Canonical Quantization of Gravity," in *Relativity, Astrophysics and Cosmology*, edited by W. Israel (Reidel, Dordrecht, 1973).

[4]P. A. M. Dirac, *Lectures on Quantum Mechanics* (Academic, New York, 1965).

[5]Note that K_{ij} is not the extrinsic curvature under the standard convention, but rather -2 times the extrinsic curvature. This enables us to identify K_{ij} directly with the normal geometrodynamical velocity.

[6]A preliminary account of this proof is given in S. A. Hojman, K. Kuchař, and C. Teitelboim, Nature (to be published).

[7]C. W. Misner, Rev. Mod. Phys. **29**, 497 (1957); P. W. Higgs, Phys. Rev. Lett. **1**, 373 (1958); **3**, 66 (1959).

[8]J. W. York, Phys. Rev. Lett. **26**, 1656 (1971); J. Math. Phys. **13**, 125 (1972).

[9]C. Teitelboim, "The Hamiltonian Structure of Spacetime," Ph. D. Thesis, Princeton University Preprint, May 1973 (unpublished).

[10]See, e.g., M. von Laue, *Die Relativitätstheorie II, Die Allgemeine Relativitätstheorie and Einsteins Lehre von der Schwerkraft* (Vieweg, Braunschweig, 1922), 1st ed.

[11]D. Lovelock, J. Math. Phys. **12**, 498 (1971).

[12]See, e.g., B. S. DeWitt, Phys. Rev. **160**, 113 (1967).

Closed-form expressions of matrix elements and eigenfunctions from ladder-operator considerations

G. Hadinger, N. Bessis, and G. Bessis

Laboratoire de Spectroscopie et de Luminescence, Université Claude Bernard, Lyon I, 69621 Villeurbanne, France
(Received 1 February 1974)

Within the Schrödinger–Infeld–Hull factorization scheme, it is shown that, by suitable transformations, the "accelerated" or "ν-step" ladder operator can always be brought to a simple canonical form, i.e., the νth derivative operation. Thus, one obtains a closed form expression of the eigenfunctions involving a Rodrigues' formula. The necessary and sufficient condition that this Rodrigues' formula generates classical orthogonal polynomials is found to be equivalent to the factorizability condition. Consequently, a closed form expression of any matrix element (diagonal or off-diagonal) on the basis of the eigenfunctions of any factorizable equation is easily derived from the calculation of one unique particular integral. In most cases, this last integral is known analytically. The Kepler problem is reinvestigated as an example. As a concluding remark, further applications of the method are considered.

I. INTRODUCTION

Recently, using an "accelerated" ladder operator procedure, we have been able to obtain closed form expressions of the vibration[1] and rotation–vibration[2] intensities of diatomic molecules for a Morse–Pekeris potential and, further, as a particular case, explicit expressions of the off-diagonal hydrogenic r^k integrals.[3,4,5] Nevertheless, the application of our procedure to the calculation of the intensities for other nuclear diatomic potentials (Manning–Rosen,[6] Rosen–Morse[7]) results in intricate expressions of the "accelerated" ladder operators. When trying to overcome these difficulties, we have found that, by suitable transformations, it is possible to reduce the "one-step" ladder operator, when it is a linear function of the quantum number, to a simple canonical form, i.e., the derivative operation acting on a new function. Such transformations had been previously investigated by Duff[8] when he considered the Truesdell's F equation.[9] Furthermore, as outlined by Duff,[8] the treatment of equations leading to ladder operators which are not linear in the quantum number, can be amenable to the resolution of equations which correspond to linear "one step" ladder operators. Thus, a unified treatment of all Infeld–Hull[10] factorizable cases (types A to F) can be undertaken.

In the present paper, we focus our attention on the calculation of closed form expressions of the matrix elements of a Hermitian operator in terms of a simple matrix element. After briefly recalling the theory and introducing the notations (Sec. II), the transformation leading to the canonical form of the ladder operator is explained. Then, once the "one step" ladder operator is transformed into the derivative operation, the corresponding "ν step" or "accelerated" ladder operator, is of course, merely the νth derivative and, consequently, leads to an expression of the eigenfunction involving a Rodrigues' formula. (Sec. III). While investigating the conditions under which this Rodrigues' formula generates orthogonal polynomials, we found that there is an equivalence between these conditions and the factorizability condition of the original second-order differential equation. From this last analysis, closed form expressions of the eigenfunctions of general Infeld–Hull type A to F factorizable equations in terms of the orthogonal polynomials are derived (Sec. IV), and the corresponding closed form of matrix elements of a Hermitian operator Q are obtained in terms of one unique matrix element (Sec. V). In other words, the full matrix, i.e., the set of all diagonal and off-diagonal matrix elements of an operator Q on the basis of the eigenfunctions of a factorizable equation can be built up by calculating one unique particular integral in which is concentrated the specificity of the Q operator: The change of the operator only changes the expression of the integral. The Kepler problem is reinvestigated as a type F illustrative example (Sec. VI).

II. GENERAL CONSIDERATIONS

Many problems of fundamental interest in quantum mechanics lead to the resolution of differential equations of the Sturm–Liouville type. Without restricting the generality of the problem, by an adequate transformation of variable and functions, these equations can be reduced to the standard form[10] (see Appendix A)

$$\left(\frac{d^2}{dx^2} + r(x,m) + \lambda\right) U(x) = 0 \tag{1}$$

with associated boundary conditions ($x_1 \leqslant x \leqslant x_2$)

$$\begin{aligned} &U(x_1) = U(x_2) = 0 \\ &\int_{x_1}^{x_2} |U(x)|^2 \, dx = 1, \end{aligned} \tag{2}$$

where $m = m_0,\ m_0+1,\ m_0+2, \ldots$ is assumed to take successive discrete values labeling the eigenfunctions.

When such an equation (1) subjected to the boundary conditions (2) is factorizable, it can be replaced by each of the following two differential equations[8,10]:

$$\begin{aligned} &H_m^+ H_m^- U_S^m = [\lambda - L(m)]\, U_S^m, \\ &H_{m+1}^- H_{m+1}^+ U_S^m = [\lambda - L(m+1)] U_S^m, \end{aligned} \tag{3}$$

where[11]

$$H_m^{\pm} = k(x,m) \mp \frac{d}{dx}. \tag{4}$$

S is the quantum number associated with the eigenvalue λ and $L(m)$ is a function which does not depend on x.

From the comparison of Eqs. (1) and (3) one gets the necessary and also sufficient condition[10] to be satisfied by $k(x,m)$ and $L(m)$ allowing the factorization of Eq. (1),

i.e.,

$$k^2(x,m+1)-k^2(x,m)+\frac{dk}{dx}(x,m+1)+\frac{dk}{dx}(x,m)$$
$$=L(m)-L(m+1). \tag{5}$$

Particularly, when the ladder operator $H_m^{\pm}$ [Eq. (4)] is a linear function in m, i.e.,

$$k(x,m)=k_1(x)\,m+k_0(x), \tag{6}$$

the necessary and sufficient condition [Eq. (5)] of factorizability of the original equation (1) becomes

$$(2m+1)\left(\frac{dk_1}{dx}+k_1^2\right)+2\left(\frac{dk_0}{dx}+k_0k_1\right)$$
$$=L(m)-L(m+1). \tag{7}$$

As it will appear later, this particular linear case (6) is of fundamental importance.

As stated by Schrödinger,[12] Infeld and Hull,[10] when the condition (5) is fulfilled, the eigenfunctions U_S^m are solutions of the following pair of difference–differential equations:

$$H_m^-\,U_S^m=N_m\,U_S^{m-1},$$
$$H_{m+1}^+\,U_S^m=N_{m+1}\,U_S^{m+1}, \tag{8}$$
$$N_m=[\lambda-L(m)]^{1/2}.$$

The operators $H_m^{\pm}$ in the equation (8) may be considered as "one step" ladder operators which generate the eigenfunctions, step by step, downward or upward, and allow the determination of any solution U_S^m from the knowledge of the top or bottom eigenfunction U_S^S, i.e., the "key" function which is the solution of a first order differential equation. One has to distinguish two cases depending on the problem under consideration.

A. Class I problems

$L(m)$ is an increasing function of m. The eigenvalues λ_S are

$$\lambda_S=L(S+1). \tag{9}$$

The "key" eigenfunction U_S^S is solution of the first order differential equation

$$H_{S+1}^+\,U_S^S=0. \tag{10}$$

B. Class II problems

$L(m)$ is a decreasing function of m. The eigenvalues λ_S are

$$\lambda_S=L(S). \tag{11}$$

The "key" eigenfunction U_S^S is solution of the first order differential equation

$$H_S^-\,U_S^S=0. \tag{12}$$

In both cases, the necessary condition for the existence of quadratically integrable solutions is

$$|m-S|=v=\text{integer}. \tag{13}$$

The mutually adjoint ladder operators [Eq. (4)] are defined so that they preserve not only the quadratic integrability but also the normalization of the eigenfunctions U_S^m $(m=S\pm v)$. Each function of the whole discrete set is completely characterized by the integer value of v which fixes its rank starting from the key function U_S^S $(v=0)$.

Infeld and Hull[10] have considered six factorization types, namely A, B, C, D, E, F. In surveying these six cases, one finds that cases A, B, C, and D correspond to ladder operators $H_m^{\pm}$ which are linear functions of m while for the two last cases, E and F, $H_m^{\pm}$ are nonlinear in m. Nevertheless, following Duff's suggestion,[8] it is shown (see Appendix B) that a connection can be established respectively between types E and A and between types F and B, and that, by a suitable transformation of variable and function, we can always deal with ladder operators $H_m^{\pm}$ linear in m.

This is why, in order to obtain the explicit expressions of the eigenfunctions, it is sufficient to consider only the fundamental case where $H_m^{\pm}$ is linear in m. In the next section, it is shown how, for this fundamental case, one can obtain a Rodrigues' formula in the expression of the eigenfunctions. Thus, the closed form expressions of the eigenfunctions, and then of the corresponding closed form expressions of matrix elements, will be derived, in a straightforward manner, from this Rodrigues' formula.

III. EIGENFUNCTIONS AND RODRIGUES' FORMULA

As previously shown[1–5], starting from the one step ladder operators equation (4), one can define the corresponding "accelerated" or "v step" operators $\mathcal{H}_v^{\pm}$ which directly generates any $U_S^{S\pm v}$ function from the key function U_S^S

$$U_S^{S\mp v}(x)=\frac{1}{\mathcal{N}_v}\,\mathcal{H}_v^{\mp}\,U_S^S(x). \tag{14}$$

For a class I problem

$$\mathcal{H}_v^-=\prod_{i=1}^{v}H_{m+i}^-=\prod_{i=1}^{v}\left(k(x,m+i)+\frac{d}{dx}\right). \tag{15}$$

For a class II problem

$$\mathcal{H}_v^+=\prod_{i=1}^{v}H_{m+1-i}^+=\prod_{i=1}^{v}\left(k(x,m+1-i)-\frac{d}{dx}\right). \tag{16}$$

$\mathcal{N}_v$ is a constant which depends on the class considered.

Now, one can question if, by a suitable transformation, it is possible to introduce, instead of the "accelerated" ladder $\mathcal{H}_v^{\mp}$ [Eqs. (15) and (16)], the "canonical accelerated" operator which is the vth derivative operation acting on a new function so that the relation (14) becomes the Rodrigues' formula of this function.

First, in order to force the "one step" ladder operator [Eq. (4)] to be a derivative operator on a new function, one has to determine a function which satisfies the following proportionnality relationship:

$$g_m(x)\,U_S^m(x)\sim\frac{d}{dx}\,[g_{m\pm1}(x)\,U_S^{m\pm1}(x)], \tag{17}$$

where $g_m(x)$ may depend on the class under consideration and the signs (+) and (−) stand respectively for class I and class II problems.

Let us consider a class I problem. The corresponding one step equation is [see Eq. (8)]

$$U_S^m(x)=\frac{1}{N_{m+1}}\,H_{m+1}^-\,U_S^{m+1}(x). \tag{18}$$

When introducing a function $g(x)$ defined on the same subspace ($x_1 \leq x \leq x_2$) as $U(x)$, from Eq. (17) and the expression (4) of the ladder operator, one can write

$$g(x)\,U_S^m(x)=\frac{1}{N_{m+1}}\,g(x)\left(k(x,m+1)+\frac{d}{dx}\right)U_S^{m+1}(x), \tag{19}$$

and in order to obtain the relationship (17), the function $g(x)$ has to satisfy

$$\left(k(x,\ m+1)-\frac{d}{dx}\right)g(x)=H_{m+1}^+\,g(x)=0. \tag{20}$$

That is, $g(x)$ must be a solution of the first order differential equation generating the key function U_S^S [Eq. (10)] and corresponds to the value $S=m$, i.e., one can write

$$g_{m+1}(x)=(U_S^S)_{S=m}. \tag{21}$$

Then, Eq. (19) becomes

$$g_{m+1}(x)\,U_S^m(x)=\frac{1}{N_{m+1}}\,\frac{d}{dx}\,[g_{m+1}(x)\,U_S^{m+1}(x)]. \tag{22}$$

It will be of the required form (17) if one introduces in the left member the function $g_m(x)$ instead of $g_{m+1}(x)$.

Let us assume that the ladder operators are linear in m, i.e.,

$$H_m^\pm=mk_1(x)+k_0(x)\mp\frac{d}{dx}. \tag{23}$$

The corresponding key function is

$$U_S^S(x)=\exp[(S+1)K_1(x)+K_0(x)], \tag{24}$$

where

$$K_1(x)=\int k_1(x)\,dx, \quad K_0(x)=\int k_0(x)\,dx, \tag{25}$$

and from (21) and (24)

$$g_{m+1}(x)=\exp[(m+1)K_1(x)+K_0(x)]. \tag{26}$$

Consequently,

$$[g_{m+1}(x)/g_m(x)]=\exp[K_1(x)]$$

Then, when introducing the new variable[13] y which is defined by

$$dy=\exp[K_1(x)]\,dx, \tag{27}$$

one obtains the required "canonical one step" relationship

$$g_m(x)\,U_S^m(x)=\frac{1}{N_{m+1}}\,\frac{d}{dy}\,[g_{m+1}(x)\,U_S^{m+1}(x)], \tag{28}$$

where x has to be considered as a function of y [Eq. (27)].

Finally, as wanted, by introducing instead of $U(x)$ the product function $g(x)\,U(x)$ of the new variable y which is defined by (27), the one step ladder operator $H_m^\pm$ reduces to the simple derivation operator d/dy. Consequently, the "accelerated" or "v step" ladder operator H_v^- [Eq. (15)] which directly generates any product $g_m U_S^m$ from the new pseudokey function $g_s\,U_S^S$ reduces to d^v/dy^v and, since $m=S-v$, one gets

$$U_S^{S-v}=\frac{1}{N_v}\,\frac{1}{g_{S-v}}\,\frac{d^v}{dy^v}\,[g_S\,U_S^S], \tag{29}$$

where

$$N_v=\prod_{i=1}^{v}N_{m+i} \tag{30}$$

and N_{m+i} is defined by (8).

The expression (29) can be considered as a Rodrigues' form of the eigenfunction U. Actually, it can be transformed to the usual Rodrigues' formula involving a weight function $w(x)$ and the vth derivative of a product $w(x)\,[h(x)]^v$. Indeed, using (24) and (26), one can write

$$U_S^m(x)=\frac{1}{N_v}\,U_S^S(x)\exp[-vk_1(x)]$$
$$\times\left(\frac{1}{w(x)}\,\frac{d^v}{dy^v}\,\{w(x)\,[h(x)]^v\}\right), \tag{31}$$

where dy is defined by (27) and

$$h(x)=\exp[2K_1(x)]. \tag{32}$$

For class II problems, with only slight modifications, the same procedure still applies.

Finally, class I and class II eigenfunctions are represented by the same expression (31) involving formally the same Rodrigues' formula with identical function $h(x)$ [Eq. (32)] but with respectively inverse weight functions $w(x)$, i.e.,

$$\text{class I}\quad\begin{cases} v=S-m, \\ w(x)=\exp[(2m+1)K_1(x)+2K_0(x)], \\ U_S^S(x)=\exp[(S+1)K_1(x)+K_0(x)], \\ N_v=\prod_{i=1}^{v}\,[L(S+1)-L(S+1-i)]^{1/2}, \end{cases} \tag{33}$$

$$\text{class II}\quad\begin{cases} v=m-S, \\ w(x)=\exp[-(2m+1)K_1(x)-2K_0(x)], \\ U_S^S(x)=\exp[-SK_1(x)-K_0(x)], \\ N_v=\prod_{i=1}^{v}\,[L(S)-L(S+i)]^{1/2}, \end{cases} \tag{34}$$

where the function $L(m)$ is defined by the factorizability condition (5). It should be noted that, when calculating the $K_1(x)$ and $K_0(x)$ functions, one can neglect to take into account the integration constants: Indeed, it is easily seen that their introduction only changes the value of the normalization constant[14] of the eigenfunction U_S^m.

Briefly stated, when the ladder operators $H_m^\pm$ are linear in m (and even nonlinear in m, since one can use the transformation of the Appendix B) the formula (31) constitutes a straightforward finite algorithm to obtain any eigenfunction from the only knowledge of the ladder operator (23) and one elementary quadrature [Eq. (25)]. Nevertheless, for calculating matrix elements in closed form, one has rather to work out the Rodrigues' formula in terms of orthogonal polynomials as it is shown in the

following section.

IV. CLOSED FORM EXPRESSION OF THE EIGENFUNCTIONS

As has been shown in Sec. III, the expression (31) of the eigenfunctions of Eq. (1), when it is factorizable, involves a Rodrigues' formula. Now, one could question what kind of function is generated by this Rodrigues' formula and if it does not correspond to an orthogonal polynomial.

When writing x as a function of y [see Eq. (27)], i.e., $w(x) \to W(y)$, $h(x) \to H(y)$, the Rodrigues' formula in (31) becomes

$$F_v(y) = \frac{1}{W(y)} \frac{d^v}{dy^v} \{W(y)\,[H(y)]^v\}, \tag{35}$$

and, as is shown in the Appendix C, the necessary and sufficient condition for $F_v(y)$ to be the Rodrigues' representation of an orthogonal polynomial is

$$\frac{d}{dy}\left(\frac{H}{W}\frac{dW}{dy}\right) = c_1 = \text{const}, \quad \frac{d^2H}{dy^2} = a_2 = \text{const}. \tag{36}$$

Then, it is interesting to deduce and examine the corresponding condition to be satisfied by the functions $k_0(x)$ and $k_1(x)$ of the linear ladder operator (23) and, hence, by the original equation (1).

Let us first consider a class I problem. Replacing in (35) $H(y)$, $W(y)$, and y by their expressions [see Eqs. (27), (32), and (33)], one readily gets

$$(2m+1)\left[\frac{d^2K_1}{dx^2} + \left(\frac{dK_1}{dx}\right)^2\right] + 2\left(\frac{d^2K_0}{dx^2} + \frac{dK_0}{dx}\frac{dK_1}{dx}\right) = c_1,$$
$$2\left[\frac{d^2K_1}{dx^2} + \left(\frac{dK_1}{dx}\right)^2\right] = a_2. \tag{37}$$

Hence, using (25), one can write the required conditions to be satisfied by $k_0(x)$ and $k_1(x)$, i.e.,

$$(2m+1)\left(\frac{dk_1}{dx} + k_1^2\right) + 2\left(\frac{dk_0}{dx} + k_0k_1\right) = c_1,$$
$$2\left(\frac{dk_1}{dx} + k_1^2\right) = a_2 \tag{38}$$

These conditions (38) are actually the necessary and sufficient conditions (7) allowing the factorization of the original equation (1).

It is easy to show that this result holds for a class II problem. Indeed, when passing from class I to class II, the only change occuring in the Rodrigues' formula (35) concerns the weight function $W(y)$ which is now replaced by $1/W(y)$ [see Eqs. (33) and (34)]. Consequently, the only corresponding change occuring in the condition (36) and then in the final result (38) is the change $c_1 \to -c_1$.

Finally, its results that the eigenfunctions of any factorizable equation (1) can be expressed in terms of the key function and well-known[15] orthogonal polynomials.

Now, let us explicit this general result for all Infeld–Hull factorizable types.[16]

A. General type A

The required eigenfunctions satisfy the following factorizable differential equation

$$\left(\frac{d^2}{dx^2} - \frac{\mathrm{a}^2(m+\mathrm{c})(m+\mathrm{c}+1) + \mathrm{d}^2 + 2\mathrm{ad}(m+\mathrm{c}+\frac{1}{2})\cos\mathrm{a}(x+p)}{\sin^2\mathrm{a}(x+p)} + \lambda\right) U(x) = 0 \tag{39}$$

The corresponding "one step" ladder operators are

$$H_m^{\pm} = (m+\mathrm{c})\,\mathrm{a}\cot\mathrm{a}(x+p) + \frac{\mathrm{d}}{\sin\mathrm{a}(x+p)} \mp \frac{d}{dx}$$

and

$$L(m) = \mathrm{a}^2(m+\mathrm{c})^2.$$

Then,

$$k_1(x) = \mathrm{a}\cot\mathrm{a}(x+p),$$
$$k_0(x) = \mathrm{ac}\cot\mathrm{a}(x+p) + \frac{\mathrm{d}}{\sin\mathrm{a}(x+p)};$$

one obtains

$$K_1(x) = \int \mathrm{a}\cot\mathrm{a}(x+p)\,dx = \ln[\sin\mathrm{a}(x+p)],$$
$$K_0(x) = \mathrm{c}\ln[\sin\mathrm{a}(x+p)] + \mathrm{d/a}\ln\{\tan[\mathrm{a}(x+p)/2]\}.$$

Consequently,

$$dy = \sin\mathrm{a}(x+p)\,dx$$

and

$$y = -(1/\mathrm{a})\cos\mathrm{a}(x+p),$$
$$h(x) = \sin^2\mathrm{a}(x+p) = 1 - \mathrm{a}^2y^2.$$

For a class I problem the weight function is

$$w(x) = [\sin\mathrm{a}(x+p)]^{2m+2\mathrm{c}+1}\,\{\tan[\mathrm{a}(x+p)/2]\}^{2\mathrm{d/a}}$$
$$= (1-\mathrm{a}y)^{m+\mathrm{c}+1/2-\mathrm{d/a}}\,(1+\mathrm{a}y)^{m+\mathrm{c}+1/2+\mathrm{d/a}}$$

Keeping in mind that, for a class II problem, the weight function is $1/w(x)$ and using the general expression (31) together with the results of Sec. 1(a) of the Appendix C, one gets the eigenfunction[14] U_S^m expressed in terms of a Jacobi polynomial, i.e.,

$$U_S^m(x) = \frac{(2\mathrm{a})^v v!}{N_v}\,U_S^S(x)\,[\sin\mathrm{a}(x+p)]^{-v} \times P_v^{(\alpha-v,\beta-v)}[\cos\mathrm{a}(x+p)], \tag{40}$$

where

$$U_S^S(x) = \{\sin[\mathrm{a}(x+p)/2]\}^{\alpha+1/2}\,\{\cos[\mathrm{a}(x+p)/2]\}^{\beta+1/2},$$

$$\text{class I}\quad v=S-m \quad \begin{cases} \alpha=S+c+\tfrac{1}{2}+d/a, \\ \beta=S+c+\tfrac{1}{2}-d/a, \\ \mathcal{N}_v=a^v\left(v!\prod_{u=1}^{v}(2S+2c+2-u)\right)^{1/2}, \end{cases}$$

$$\text{class II}\quad v=m-S \quad \begin{cases} \alpha=-(S+c+\tfrac{1}{2}+d/a), \\ \beta=-(S+c+\tfrac{1}{2}-d/a), \\ \mathcal{N}_v=a^v\left(v!\prod_{u=1}^{v}(-2S-2c-u)\right)^{1/2}. \end{cases}$$

B. General type B

The required eigenfunctions satify the following factorizable differential equation:

$$\left(\frac{d^2}{dx^2}-d^2\exp(2ax)+2\,ad(m+c+\tfrac{1}{2})\exp(ax)+\lambda\right)U(x)=0. \tag{41}$$

The corresponding "one step" ladder operators are

$$H_m^{\pm}=d\exp(ax)-a(m+c)\mp\frac{d}{dx}$$

and

$$L(m)=-a^2(m+c)^2,$$

Then

$$k_1(x)=-a,\quad k_0(x)=d\exp(ax)-ac;$$

one obtains

$$K_1(x)=-ax,\quad K_0(x)=(d/a)\exp(ax)-acx$$

Consequently,

$$dy=\exp(-ax)dx \quad\text{and}\quad y=-(1/a)\exp(-ax),$$

$$h(x)=\exp(-2ax).$$

For a class I problem the weight function is

$$w(x)=\exp[-(2m+1+2ac)ax+(2d/a)\exp(ax)]$$

$$=(-ay)^{2m+1+2ac}\exp(-2d/a^2y).$$

Keeping in mind that for a class II problem, the weight function is $1/w(x)$, and using the general expression (31) together with the results of Sec. 1(b) of the Appendix C, one gets the eigenfunction[14] U_S^m expressed in terms of a generalized Laguerre polynomial, i.e.,

$$U_S^m(x)=(a^v v!/\mathcal{N}_v)\,U_S^S(x)\,L_v^{\alpha}(\beta\exp(ax)), \tag{42}$$

where

$$U_S^S(x)=\exp[\tfrac{1}{2}\alpha ax-\tfrac{1}{2}\beta\exp(ax)],$$

$$\text{class I}\quad v=S-m \quad \begin{cases} \alpha=-2S-2c-2,\quad \beta=-2d/a, \\ \mathcal{N}_v=a^v\left(v!\prod_{u=1}^{v}(-2S-2c-2+u)\right)^{1/2}, \end{cases}$$

$$\text{class II}\quad v=m-S \quad \begin{cases} \alpha=2S+2c,\quad \beta=2d/a, \\ \mathcal{N}_v=a^v\left(v!\prod_{u=1}^{v}(2S+2c+u)\right)^{1/2} \end{cases}$$

C. General type C

The required eigenfunctions satisfy the following differential factorizable equation:

$$\left(\frac{d^2}{dx^2}-\frac{(m+c)(m+c+1)}{x^2}-\frac{b^2x^2}{4}+b(m-c)+\lambda\right)U(x)=0. \tag{43}$$

The corresponding "one step" ladder operators are

$$H_m^{\pm}=\frac{m+c}{x}+\frac{bx}{2}\mp\frac{d}{dx}$$

and

$$L(m)=-2bm+b/2.$$

Then

$$k_1(x)=1/x,\quad k_0(x)=c/x+bx/2;$$

one obtains

$$K_1(x)=\ln x,$$

$$K_0(x)=c\,\ln x+(b/4)x^2.$$

Consequently,

$$dy=x\,dx \quad\text{and}\quad y=\tfrac{1}{2}x^2,$$

$$h(x)=x^2=2y$$

For a class I problem the weight function is

$$w(x)=x^{2m+1+2c}\exp(bx^2/2)$$

$$=(2y)^{m+c+1/2}\exp(by).$$

Keeping in mind that for a class II problem, the weight function is $1/w(x)$, and using the general expression (31) together with the results of the Sec. 2(a) of Appendix C, one gets the eigenfunctions[14] U_S^m expressed in terms of a Laguerre polynomial:

$$U_S^m(x)=\frac{1}{\mathcal{N}_v}\,2^v v!\,U_S^S(x)\;x^{-v}\,L_v^{(\alpha-v)}(\beta x^2), \tag{44}$$

where

$$U_S^S(x)=x^{\alpha+1/2}\exp(-\beta x^2/2),$$

$$\text{class I}\quad v=S-m \quad \begin{cases} \alpha=S+c+\tfrac{1}{2},\quad \beta=-b/2, \\ \mathcal{N}_v=[(-2b)^v v!\,]^{1/2}, \end{cases}$$

$$\text{class II}\quad v=m-S \quad \begin{cases} \alpha=-(S+c+1/2),\quad \beta=b/2, \\ \mathcal{N}_v=[(2b)^v v!\,]^{1/2}. \end{cases}$$

It should be noted that the particular case $b=0$ leads to $L(m)=0$, i.e., although the ladder operators still exist, the class as well as the key functions are no longer defined. Nevertheless, the set of the eigenfunctions can be generated from any available solution used as a pseudokey function.[17] Then, when formally applying our procedure, it is seen from the results of the Sec. 2(b) of Appendix C that the orthogonal polynomial generated by the Rodrigues' formula in (31) reduces to a constant. In fact, in that case, since the original equation (1) reduces to the well-known differential equation of Bessel functions, explicit expressions of the eigenfunctions are obtainable elsewhere[15]: The use of the ladder operators may enable to find again known properties of Bessel functions.

D. General type D

The required eigenfunctions satisfy the following differential factorizable equation:

$$\left(\frac{d^2}{dx^2} - (bx+d)^2 + b(2m+1) + \lambda\right) U(x) = 0. \tag{45}$$

The corresponding "one step" ladder operators are

$$H_m^{\pm} = bx + d \mp \frac{d}{dx}$$

and

$$L(m) = -2bm.$$

Then

$$k_1(x) = 0, \quad k_0(x) = bx + d;$$

one obtains

$$K_1(x) = 0, \quad K_0(x) = \tfrac{1}{2}bx^2 + dx$$

Consequently,

$$dy = dx \text{ and } y = x, \quad h(x) = 1.$$

For a class I problem, the weight function is

$$w(x) = \exp[(bx^2 + 2dx)].$$

Keeping in mind that, for a class II problem, the weight function is $1/w(x)$ and using the general expression (31) together with the results of the Sec. 3(a) of Appendix C, one gets the eigenfunction[14] U_S^m expressed in terms of an Hermite polynomial, i.e.,

$$U_S^m(x) = \frac{1}{N_v}(-\epsilon b)^{v/2} \; U_S^S(x) H_v[(-\epsilon b)^{1/2}(x + d/b)], \tag{46}$$

where

$$U_S^S(x) = \exp[\epsilon(bx^2/2 + dx)],$$

$$N_v = [(-\epsilon 2b)^v v!]^{1/2},$$

and the sign $\epsilon = +1$ stands for class I problems ($v = S - m$) while $\epsilon = -1$ stands for class II problems ($v = m - S$).

It should be noted that, for the particular case $b = 0$, as it can be shown from the results of the Sec. 3(b) of Appendix C, the Rodrigues' formula in (31) generates a constant and U_S^m becomes the exponential function. Indeed, in that case, the original equation (1) merely reduces to the differential equation of this function.

E. General type E

The required eigenfunctions satisfy the following factorizable differential equation:

$$\left(\frac{d^2}{dx^2} - \frac{S(S+1)}{[(1/a)\sin a(x+p)]^2} - 2aq\cot a(x+p) + \lambda(m)\right) U_S^m = 0. \tag{47}$$

The corresponding "one step" ladder operators are

$$H_S^{\pm} = S\,a\cot a(x+p) + \frac{q}{S} \mp \frac{d}{dx}$$

and

$$L(S) = a^2S^2 - q^2/S^2.$$

The ladder operators $H_S^{\pm}$ are not linear in the quantum number S. Nevertheless, since there is a connection between factorizable type E and type A equations, as it has been suggested by Infeld and Hull,[10] we first introduce the transformation of function and variable which is given in Appendix B, i.e.,

$$dX = a\,dx/\sin a(x+p),$$

$$U(x) = [(1/a)\sin a(x+p)]^{1/2}\,\mathcal{U}(X).$$

We take

$$X = \ln\{\tan[a(x+p)/2]\}.$$

One obtains

$$\left(\frac{d^2}{dx^2} - \frac{1}{a^2\cosh^2 X} \times [\tfrac{1}{4}a^2 - \lambda(m) - 2aq\sinh X] - (S+\tfrac{1}{2})^2\right) \mathcal{U}_S^m(X) = 0$$

For a class I (type E) problem,

$$\lambda(m) = L(m+1) = a^2(m+1)^2 - q^2/(m+1)^2,$$

For a class II (type E) problem,

$$\lambda(m) = L(m) = a^2m^2 - q^2/m^2.$$

Then, when introducing artificial factorization,[10] i.e., an artificial shift of the eigenvalue, $\mathcal{U}(X)$ becomes a solution of the factorizable type A equation and one can make use of the results of Sec. IVA. It should be noted that when $U(x)$ is solution of a (type E) class I problem (or a class II problem) the corresponding eigenfunction $\mathcal{U}(X)$ is a solution of a (type A) class II problem (or a class I problem).

Finally, one gets the required closed form expression of the eigenfunctions[14] of the general type E factorizable equation (47):

$$U_S^m(x) = \frac{2^v v!}{N_v} \; U_S^S(x)[\sin a(x+p)]^v \times P_v^{(\alpha-v,\beta-v)}[\coth ia(x+p)], \tag{48}$$

where

$$U_S^S(x) = [\sin a(x+p)]^{-(\alpha+\beta)/2} \times \exp\{-i[(\alpha-\beta)/2]a(x+p)\},$$

and, following the class of the original type E problem,

$$\text{class I } v = m - S \begin{cases} \alpha = -S - 1 + iq/a(m+1), \\ \beta = -S - 1 - iq/a(m+1), \\ N_v = \left(v! \prod_{u=1}^{v}(2S+1+u)\right)^{1/2}, \end{cases}$$

$$\text{class II } v = S - m \begin{cases} \alpha = S - iq/am, \quad \beta = S + iq/am, \\ N_v = \left(v! \prod_{u=1}^{v}(2S+1-u)\right)^{1/2}. \end{cases}$$

U_S^S is a pseudokey function which depends on the value of the "artificial parameter" which is $iq/a(m+1)$ for a class I problem and $-iq/am$ for a class II problem.

F. General type F

The required eigenfunctions satisfy the following

TABLE I. Explicit expressions of the eigenfunctions[14] $U_S^m(x) = (1/\mathcal{N}_v)\sum_{j=0}^{v} A_j \phi_j(x)$.

Type	A_j	$\phi_j(x)$	Reference for parameters
A	$\binom{v}{j}\frac{(-)^j\Gamma(\alpha+1)\Gamma(\alpha+\beta-v+1+j)}{v!\,\Gamma(\alpha+1-v+j)\Gamma(\alpha+\beta-v+1)}$	$(a)^v v!\,\{\sin[a(x+p)/2]\}^{\alpha+1/2-v+2j}\{\cos[a(x+p)/2]\}^{\beta+1/2-v}$	Sec. IV A
B	$\binom{v}{j}\frac{(-)^j\Gamma(\alpha+1+v)}{v\Gamma(\alpha+1+j)}\,(\beta)^j$	$(a)^v v!\,\exp[(\alpha/2+j)ax-(\beta/2)e^{ax}]$	Sec. IV B
C	$\binom{v}{j}\frac{(-)^j\Gamma(\alpha+1)}{v!\,\Gamma(\alpha+1-v+j)}\,(\beta)^j$	$(2)^v v!\,(x)^{\alpha+1/2-v+2j}\,e^{-\beta x^2/2}$	Sec. IV C
D	$\frac{(\epsilon b)^{j/2}(2)^j v!}{j!\,[(v-j)/2]!}$ if $v-j$ is even; otherwise is zero	$(-\epsilon b)^{v/2}(x+d/2)^j\exp[\epsilon(bx^2/2+dx)]$	Sec. IV D
E	$\binom{v}{j}\frac{(-i)^j\Gamma(\alpha+1)\Gamma(\alpha+\beta-v+1+j)}{v!\,(2)^j\Gamma(\alpha+1-v+j)\Gamma(\alpha+\beta-v+1)}$	$(2)^v v!\,[\sin a(x+p)]^{-[(\alpha+\beta)/2]+v-j}\exp\{-ia(x+p)[(\alpha-\beta)/2+j]\}$	Sec. IV E
F	$\binom{v}{j}\frac{\Gamma(2S+2+v)}{v!\,\Gamma(2S+2+j)}\left(\frac{2q}{m+1}\right)^j$	$v!\,(x)^{S+1+j}e^{qx/(m+1)}$	Sec. IV F

differential factorizable equation:

$$\left(\frac{d^2}{dx^2}-\frac{S(S+1)}{x^2}-\frac{2q}{x}+\lambda(m)\right)U_S^m(x)=0. \tag{49}$$

The corresponding "one step" ladder operators are

$$H_S^{\pm}=\frac{S}{x}+\frac{q}{S}\mp\frac{d}{dx}$$

and

$$L(S)=-q^2/S^2.$$

Since the ladder operators are nonlinear in S, we first introduce the transformation given in Appendix B, i.e.,

$$dX=\frac{dx}{x},\quad U_S^m(x)=x^{1/2}\mathcal{U}(X).$$

We take

$$X=\ln x;$$

one obtains

$$\left(\frac{d^2}{dX^2}+\lambda(m)\exp(2X)-2q\exp(X)-(S+\tfrac{1}{2})^2\right)\mathcal{U}(X)=0,$$

where, for a class I (type F) problem (q is supposed to be real)

$$\lambda(m)=L(m+1)=-q^2/(m+1)^2.$$

Then, when introducing artificial factorization,[10] $\mathcal{U}(X)$ becomes a solution of a factorizable type B equation, and one can make use of the general results of Sec. IVB.

Finally, one gets[14]

$$U_S^m(x)=(v!/\mathcal{N}_v)\;U_S^S(x)\,L_v^{(2S+1)}[-2qx/(m+1)], \tag{50}$$

where

$$U_S^S(x)=x^{S+1}\exp[qx/(m+1)],\quad v=m-S,$$

$$\mathcal{N}_v=\left(v!\prod_{u=1}^{v}(2S+1+u)\right)^{1/2}.$$

V. CLOSED FORM EXPRESSIONS OF MATRIX ELEMENTS

Finally, we have now at our disposal (Sec. IV) closed form expressions of the eigenfunctions $U_S^m(x)$ of any factorizable equation (1) in terms of orthogonal polynomials. When introducing the classical expressions of these polynomials [see Sec. (II) of Appendix C], one can write

$$U_s^m(x)=\frac{1}{\mathcal{N}_v}\sum_{j=0}^{v}A_j\,\phi_j(x). \tag{51}$$

The expressions of the functions $\phi_j(x)$ as well as of the coefficients A_j are gathered up in Table I for type A to type F factorizable cases.

Now, let us consider the general matrix element of a Hermitian operator $Q(x)$, i.e.,[18]

$$\langle Q\rangle = C\;C'\;\mathcal{M}_{v,v'}\;(S,S'), \tag{52}$$

where

$$\mathcal{M}_{v,v'}(S,S')=\int_{x_1}^{x_2}(U_S^m(x))^*\,Q(x)\;U_{S'}^{m'}(x)\,dx.$$

C is the normalization constant of the eigenfunction U_S^m. Using the expression (51), one obtains the closed form expression

$$\mathcal{M}_{v,v'}(S,S')=\frac{1}{\mathcal{N}_v\mathcal{N}'_v}\sum_{j=0}^{v}\sum_{k=0}^{v'}A_jA'_k\,\mathcal{J}_{jk}, \tag{53}$$

$$\mathcal{J}_{jk}=\int_{x_1}^{x_2}\phi_j^*(x)\;Q(x)\;\phi'_k(x)\,dx. \tag{54}$$

Owing to the particularly simple dependance upon j of all the $\phi_j(x)$ functions (see Table I), i.e., j appears always as a power index, the $\mathcal{J}_{jk}$ integral (54) will have a simple dependance upon the j and k index. Furthermore, in most cases of interest in quantum theory, we experienced that the analytical expression of $\mathcal{J}_{jk}$ can be found in *Tables*.[19] Subsequently, formula (53) defines an analytical closed form expression of any matrix element, hence, an easy and quick algorithm of computation.

Finally, to calculate any matrix element of an operator $Q(x)$ between eigenfunctions of an factorizable equa-

tion, one can use the following recipe:

(1) Find in Sec. IV to which type the reduced equation (1) belongs.

(2) Look, in Table I, at the adequate type entry and pick up the corresponding expressions of ϕ_j and A_j.

(3) Find in *Table*[19] or calculate, the fundamental integral $\mathcal{J}_{jk}$ [Eq. (54)].

(4) Use the contraction formula (53) and the expression (52), this is, introduce the adequate normalization constants of the eigenfunctions.

VI. ILLUSTRATIVE APPLICATION

Let us consider, for instance, the determination of a closed form expression of hydrogenic radial r^K matrix elements.

After setting $\psi_{nl}(r)=r^{-1}R_{nl}(r)$, the radial Schrödinger equation is

$$\left(\frac{d^2}{dr^2}+\frac{2Z}{r}-\frac{l(l+1)}{r^2}+\lambda_n\right)R_n^l(r)=0. \tag{55}$$

It is easily seen (Sec. IV) that Eq. (55) is a type F (class I) factorizable case which corresponds to the following values,

$$q=-Z,\quad S=l,\quad \lambda_n=\lambda(m)=-Z^2/(m+1)^2,$$

and, when introducing the usual radial quantum number $n=m+1$, one obtains the quantification condition

$$v=m-S=n-l-1.$$

Now, let us consider the determination of the r^K matrix elements.

Taking out from Table I (F entry) the $\phi_j(r)$ function, the fundamental integral $\mathcal{J}_{jk}$ [Eq. (54)] is immediately written as

$$\mathcal{J}_{jk}=(n-l-1)!(n'-l'-1)!$$
$$\times\int_0^\infty \exp[-Zr(1/n+1/n')]\,r^{(l+l'+2+K+j+k)}\,dr.$$

One recognizes namely a Γ function and gets

$$\mathcal{J}_{jk}=(n-l-1)!(n'-l'-1)!\left(\frac{nn'}{Z(n+n')}\right)^{l+l'+K+3+j+k}$$
$$\times(l+l'+K+2+j+k)!.$$

Using formula (53) together with the expressions of the A_j coefficients (Table I), one obtains

$$\langle nl|r^K|n'l'\rangle=\mathcal{A}$$
$$\times\sum_{j=0}^{n-l-1}\binom{n-l-1}{j}\left(-\frac{1}{n}\right)^j$$
$$\times\sum_{k=0}^{n'-l'-1}\binom{n'-l'-1}{k}\left(-\frac{1}{n'}\right)^k$$
$$\times\left(\frac{2nn'}{n+n'}\right)^{j+k}\frac{(l+l'+K+2+j+k)!}{(2l+1+j)!(2l'+1+k)!},$$

where

$$\mathcal{A}=\frac{CC'}{N_vN_v'}\left(\frac{nn'}{Z(n+n')}\right)^{l+l'+K+3}(l+n)!(l'+n')!.$$

When explaining the normalization constant C and N_v (see Sec. IVF), one finds again our previous results.[4]

Furthermore, if needed, from Eq. (50), the hydrogenic function is seen to be, as expected,

$$\psi_{nl}(r)=r^{-1}R_n^l(r)$$
$$=(C/N_v)(n-l-1)!\,r^l\exp(-Zr/n)L_{n-l-1}^{2l+1}(2Zr/n).$$

VII. CONCLUSION

Summarizing our approach to the resolution of factorizable Sturm—Liouville equations, we have introduced a suitable change of variable and function which reduces the one step ladder operator to a canonical form that is the derivation operation. Consequently, this results in a very simple expression of the "accelerated" or "v step" ladder operator which becomes the vth derivative and, when applied to the key or any avalaible pseudokey function, generates any eigenfunction U_S^m ($v=|S-m|$) in a standard canonical closed form involving a Rodrigues' formula. Furthermore, we have pointed out the equivalence between the factorizability condition of the original equation and the condition under which this Rodrigues' formula generates orthogonal polynomials. Consequently, the calculation of eigenfunctions or/and matrix elements becomes straightforward. From a practical point of view, the use of our results gathered in Table I, together with the preceeding manufacturing recipe (Sec. V) gives a closed form expression of any matrix element of an operator Q between eigenfunctions of any factorizable equation in terms of one unique integral which, in most cases, is obtainable from *Tables*.[19]

Such a procedure works nicely, for instance, for calculating rotation—vibration intensities of diatomic molecules, since, for several nuclear potential functions (Morse—Pekeris, Rosen—Morse, Manning—Rosen, Tietz potentials ···) the diatomic nuclear equation is still factorizable if an adequate expansion technique is used to include the rotation—vibration coupling.[20] Results of this study will be given in a forthcoming paper. The same treatment is indicated when considering screened Coulombic potentials (Hulthen, long-range Yukawa potential) and does not involve more complication than the hydrogenic case. Moreover, directly from the results of Table I, it appears that the integral of the product of three eigenfunctions of a factorizable equation is easily obtainable in a closed form. As a particular example, one finds an easy to compute explicit expression, involving factorials, of the well-known integral of the product of three spherical harmonics, i.e., of the Slater—Condon coefficients which play an important role in atomic calculations. The same treatment is adequate for other functions: The determination of the associated integrals and of their selection rules is hopefully undertaken.

As a last point, in order to overstep the bounds of applicability of this powerful and elegant method which is the Schrödinger—Infeld—Hull factorization method, it is worthwhile to question how our method works within the perturbation scheme. This aspect, as well as the connection of our method with other theories, is to be considered.

APPENDIX A

Consider a one-dimensional differential equation of the Sturm—Liouville type:

$$\left[\frac{d}{du}\left(f(u)\frac{d}{du}\right)+q(u,m)+\lambda\rho(u)\right]P(u)=0, \tag{A1}$$

It can be transformed into the standard form[10]

$$\left(\frac{d^2}{dx^2}+r(x,m)+\lambda\right)U(x)=0. \tag{A2}$$

The transformation connecting Eqs. (A1) and (A2) is

$$U=(f\rho)^{1/4}P, \quad dx=(\rho/f)^{1/2}\,du. \tag{A3}$$

Indeed, the possibility of such a transformation implies that the functions $f(u)$ and $q(u,m)$ are never negative and $\rho(u)/f(u)$ exists everywhere.

APPENDIX B

The general form of type E and type F factorizable equations is

$$\left(\frac{d^2}{dx^2}+\mathrm{b}\frac{S(S+1)}{[f(x)]^2}-g(x)+\lambda(m)\right)U_S^m(x)=0. \tag{B1}$$

The corresponding ladder operators $H_S^{\pm}$ are not linear in S.

One has to search for a suitable change of variable and function in order to transform (B1) into the required standard form:

$$\left(\frac{d^2}{dX^2}+G(X)+\lambda(m)\,F(X)+\mathrm{b}S(S+1)\right)\mathcal{U}_S^m(X)=0; \tag{B2}$$

when setting

$$x=\varphi(X), \quad U(x)=\Psi(X)\,\mathcal{U}(X), \tag{B3}$$

Eq. (B1) becomes

$$\Psi\left\{\frac{1}{\varphi'^2}\left[\mathcal{U}''+\mathcal{U}'\left(2\frac{\Psi'}{\Psi}-\frac{\varphi''}{\varphi'}\right)\right]\right.$$

$$\left.+\,\mathcal{U}\left(\frac{\Psi''}{\Psi}-\frac{\varphi''}{\varphi'}\frac{\Psi'}{\Psi}+\mathrm{b}\frac{S(S+1)}{f^2}-g+\lambda(m)\right)\right\}=0. \tag{B4}$$

In Eq. (B4) one has introduced the shortened notation

$$\mathcal{U}''=\frac{d^2\mathcal{U}}{dX^2}, \quad \mathcal{U}'=\frac{d\mathcal{U}}{dX}, \ldots.$$

In order to obtain a differential equation in $\mathcal{U}(X)$ not containing the first derivative $\mathcal{U}'$, one has first to choose the function $\Psi(X)$ and $\varphi(X)$ so that

$$2\,\Psi'/\Psi=\varphi''/\varphi'.$$

Subsequently

$$[\Psi(X)]^2=\varphi'=\frac{d\varphi}{dX}. \tag{B5}$$

Hence, when taking into account the relationship (B5), Eq. (B4) becomes

$$\frac{d^2\mathcal{U}}{dX^2}+\mathcal{U}(X)\left[\Psi\frac{d}{dX}(\Psi'\Psi^{-2})\right.$$

$$\left.+\left(\frac{\mathrm{b}S(S+1)}{f^2}+\lambda-g\right)\Psi^4\right]=0. \tag{B6}$$

Therefore, if one chooses

$$[\Psi(X)]^2=f(x) \tag{B7}$$

One obtains

$$\frac{d^2\mathcal{U}}{dX^2}+\mathcal{U}\left(\Psi\frac{d}{dX}(\Psi'\Psi^{-2})\right.$$

$$\left.+\,\Psi^4[\lambda-g(x)]+\mathrm{b}\,S(S+1)\right)=0, \tag{B8}$$

where $g(x)$ is to be considered as a function of X and $\Psi(X)$ is given by (B7).

Finally, the suitable change of variable and function is defined by

$$dX=\frac{dx}{f(x)}, \quad U(x)=[f(x)]^{1/2}\,\mathcal{U}(X). \tag{B9}$$

APPENDIX C

(I) Let us consider the Rodrigues' formula

$$F_v(y)=\frac{1}{W(y)}\frac{d^v}{dy^v}\{W(y)[H(y)]^v\} \tag{C1}$$

Our purpose is to show that the following condition

$$\frac{d}{dy}\left(\frac{H(y)}{W(y)}\frac{dW(y)}{dy}\right)=c_1=\text{const},$$

$$\frac{d^2H(y)}{dy^2}=a_2=\text{const} \tag{C2}$$

is the necessary and sufficient condition for $F_v(y)$ to be the Rodrigues' representation of a polynomial.

Indeed, it is easily shown that the condition (C2) is satisfied when (C1) is the Rodrigues' formula of the "classical" polynomials.

Reciprocally, let us assume that the condition (C2) is fulfilled. One can write

$$\frac{H}{W}\frac{dW}{dy}=c_1y+c_0,$$

$$H=a_2y^2+a_1y+a_0, \tag{C3}$$

and from (C3) one gets

$$\frac{1}{W}\frac{dW}{dy}=\frac{c_1y+c_0}{a_2y^2+a_1y+a_0}. \tag{C4}$$

Different cases have to be considered depending upon the effective degree in y of the H polynomial, i.e., upon the values of the constants.

1. First case $a_2 \neq 0$

(a) When $a_1^2-4a_0a_2\neq 0$, one can introduce the linear change of variable

$$u=\frac{2a_2}{[a_1^2-4a_0a_2]^{1/2}}\left(y+\frac{a_1}{2a_2}\right). \tag{C5}$$

The first order differential equation (C4) becomes

$$\frac{1}{W}\frac{dW}{du}=-\frac{c_1}{a_2}\left(\frac{u}{1-u^2}\right)$$

$$-\frac{(2c_0a_2-c_1a_1)}{a_2(a_1^2-4a_0a_2)^{1/2}}\left(\frac{1}{1-u^2}\right),$$

and one gets

$$W=(1-u)^{\alpha}(1+u)^{\beta}, \qquad H=-[(a_1^2-4a_0a_2)/4a_2](1-u^2), \tag{C6}$$

where

$$\alpha=\frac{c_1}{2a_2}+\frac{1}{2a_2}\left(\frac{2c_0a_2-c_1a_1}{[a_1^2-4a_0a_2]^{1/2}}\right), \qquad \beta=\frac{c_1}{2a_2}-\frac{1}{2a_2}\left(\frac{2c_0a_2-c_1a_1}{[a_1^2-4a_0a_2]^{1/2}}\right). \tag{C7}$$

Hence, from the expression (C6) of W and H, one finds that the Rodrigues' formula (C1) generates a Jacobi polynomial $P_v^{(\alpha,\beta)}(u)$, i.e.,

$$F_v=v!\,(a_1^2-4a_0a_2)^{v/2}\,P_v^{(\alpha,\beta)}(u), \tag{C8}$$

where u is defined as a function of y by (C5) and the parameters α and β are given by (C7).

(b) When $a_1^2-4a_0a_2=0$, if one introduces the following change of variable:

$$u=y+a_1/2a_2. \tag{C9}$$

The first order differential equation (C4) becomes

$$\frac{1}{W}\frac{dW}{du}=\frac{c_1}{a_2}\frac{1}{u}+\left(\frac{c_0}{a_2}-\frac{c_1a_1}{2a_2^2}\right)\frac{1}{u^2}$$

and one gets

$$W=u^{c_1/a_2}\ \exp\left[-\left(\frac{c_0}{a_2}-\frac{c_1a_1}{2a_2^2}\right)\frac{1}{u}\right], \qquad H=a_2u^2. \tag{C10}$$

Since (cf. for instance, the Appendix of Ref. 8)

$$\frac{d^n}{dt^n}[t^{-k}e^{1/t}]=(-)^n n!\ t^{-n-k}\,e^{1/t}L_n^{k-1}\left(-\frac{1}{t}\right),$$

where $L_n^{\alpha}(x)$ is a Laguerre polynomial.[14]

Finally, one gets

$$F_v=(-)^v v!\ a_2^v u^v L_v^{\alpha}\left[-\left(\frac{c_0}{a_2}-\frac{c_1a_1}{2a_2^2}\right)\frac{1}{u}\right], \tag{C11}$$

where

$$\alpha=-c_1/a_2-2v-1 \tag{C12}$$

and u is defined as a function of y by (C9)

2. Second case, $a_2 = 0$ and $a_1 \neq 0$

(a) When $c_1\neq 0$, one can introduce the change of variable

$$u=-(c_1/a_1^2)\,(a_1y+a_0). \tag{C13}$$

The equation (C4) becomes

$$\frac{1}{W}\frac{dW}{du}=-1+\left(\frac{a_0c_1}{a_1^2}+\frac{c_0}{a_1}\right)\frac{1}{u},$$

and one gets

$$W=u^{\alpha}e^{-u}, \qquad H=-(a_1^2/c_1)\,u, \tag{C14}$$

where

$$\alpha=a_0c_1/a_1^2+c_0/a_1. \tag{C15}$$

Hence, the Rodrigues' formula generates Laguerre polynomial $L_v^{\alpha}(u)$ and

$$F_v=v!\ a_1^v L_v^{\alpha}(u) \tag{C16}$$

(b) When $c_1=0$, if one introduces the change of variable

$$u=a_1y+a_0, \tag{C17}$$

Eq. (C4) becomes

$$\frac{1}{W}\frac{dW}{du}=\frac{c_0}{a_1}\left(\frac{1}{u}\right),$$

and one gets

$$W=u^{c_0/a_1}, \qquad H=u, \tag{C18}$$

and, in fact, the function F_v reduces to a constant value, i.e.,

$$F_v=a_1^v\,\frac{\Gamma(v+c_0/a_1+1)}{\Gamma(c_0/a_1+1)}=v!\ a_1^v L_v^{\alpha}(0), \tag{C19}$$

where $\alpha=c_0/a_1$.

3. Third case, $a_2 = a_1 = 0$

(a) When $c_1\neq 0$ and introducing the change of variable

$$u=(-2a_0c_1)^{-1/2}(c_1y+c_0), \tag{C20}$$

one gets

$$W=\exp(-u^2), \qquad H=a_0 \tag{C21}$$

and the Rodrigues' formula (C1) generates Hermite polynomials $H_v(u)$, i.e.,

$$F_v=(-)^v\,(-c_1a_0/2)^{v/2}H_v(u) \tag{C22}$$

(b) When $c_1=0$, one gets

$$W=\exp[(c_0/a_0)y], \qquad H=a_0, \tag{C23}$$

and the function F_v reduces to a constant value, i.e.,

$$F_v=c_0^v. \tag{C24}$$

Finally, when the condition (C2) is fulfilled, it has been shown that the Rodrigues' formula (C1) generates classical orthogonal polynomials and, a straightforward way, explicit expressions of F_v can be obtained in terms of the well-known explicit expressions of the polynomials.

(II) Explicit expressions of the orthogonal polynomials[15]:

1. Jacobi polynomial:

$$P_v^{(\alpha,\beta)}(u)=\frac{\Gamma(\alpha+1+v)}{v!(v+\alpha+\beta+1)}\times\sum_{j=0}^{v}\binom{v}{j}\frac{\Gamma(v+\alpha+\beta+1+j)}{2^j\,\Gamma(\alpha+1+j)}(u-1)^j.$$

2. Laguerre polynomial:

$$L_v^{\alpha}(u)=\sum_{j=0}^{v}(-)^j\,\frac{\Gamma(\alpha+1+v)}{(v-j)!\,\Gamma(\alpha+1+j)}\,u^j.$$

3. Hermite polynomial:

$$H_v(u)=v!\sum_{j=0}^{[v/2]}(-)^j\,\frac{1}{j!(v-2j)!}\,(2u)^{v-2j}.$$

[1]M. Badawi, N. Bessis, and G. Bessis, Can. J. Phys. **51**, 2075 (1973).
[2]M. Badawi, N. Bessis, G. Bessis, and G. Hadinger, Can. J. Phys. **52**, 110 (1974).
[3]M. Badawi, N. Bessis, G. Bessis, and G. Hadinger, Phys. Rev. A **8**, 727 (1973).
[4]N. Bessis, G. Bessis, and G. Hadinger, Phys. Rev. A **8**, 2246 (1973).
[5]G. Hadinger, N. Bessis, and G. Bessis, Phys. Rev. A **8**, 2781 (1973).
[6]M. F. Manning and N. Rosen, Phys. Rev. **44**, 953 (1933).
[7]N. Rosen and P. M. Morse, Phys. Rev. **42**, 210 (1932).
[8]G. F. D. Duff, Can. J. Math. **1**, 379 (1949).
[9]C. Truesdell, *A Unified Theory of Special Functions* (Princeton U. P., Princeton, N. J., 1948).
[10]L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).
[11]Hereafter our $H_m^{\pm}$ operators differ from Hull's operators by a sign.
[12]E. Schrödinger, Proc. R. Irish Acad. A **46**, 9, 183 (1940); **47**, 53 (1941).
[13]When $H_m^{\pm}$ is nonlinear in m, the ratio $g_{m+1}(x)/g_m(x)$ depends on m, and then the change of variable and function which leads to the Rodrigues' formula is no longer feasible.
[14]Hereafter, purposely, the normalization constant of the eigenfunctions U_S^m is not made explicit. For type A to type D cases, in order to obtain normalized eigenfunctions U_S^m, it is sufficient to normalize the key function U_S^S. Nevertheless, this is not true for type E and type F cases and it is better to use the results of Sec. VI to calculate the normalization constant from the explicit particular expression of the matrix element of the operator $Q(x) \equiv 1$.
[15]M. Abramowitz and I. A. Segun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[16]Hereafter roman letters a, b, c, d are coefficients (cf. Infeld—Hull notation) while the italic d stands for the derivative symbol.
[17]Such a situation was encountered when considered r^k hydrogenic discrete continuous integrals, and a pseudokey function was introduced.
[18]Primed notation corresponds to quantum numbers S', m', and v'.
[19]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York and London, 1965).
[20]M. Badawi, N. Bessis, and G. Bessis, J. Phys. B: Atom. Molec. Phys. **5**, L157 (1972).

A fluid sphere in general relativity

Ronald J. Adler

The Institute for Advanced Study, Princeton, New Jersey 08540
(Received 3 December 1973; revised manuscript received 4 February 1974)

We solve the Einstein field equations for the interior of a static fluid sphere in closed analytic form. The model sphere obtained has a physically reasonable equation of state, and a maximum mass of 2/5 the fluid radius (in geometric units). As the maximum mass is approached the central density and pressure become infinite, while for masses greater than about 0.35 times the fluid radius the velocity of sound in portions of the fluid exceeds the velocity of light, indicating that the fluid is noncausal in this mass range. In the low mass limit the solution becomes identical to the Schwarzschild interior solution.

1. INTRODUCTION

Exact solutions to the Einstein field equations in closed analytic form are difficult to obtain due to the nonlinearity of the equations. In particular the problem of constructing a static model sphere of perfect fluid (e.g., a neutron star model) is usually solved by numerical methods using the Tolman–Oppenheimer–Volkoff (TOV) equation[1,2,3] with an equation of state specified; this is a straightforward procedure, but yields results expressed in cumbersome numerical or graphic form.[3]

The small number of analytic solutions which have been obtained are valuable and interesting because one may study their properties in complete detail and with comparative ease, especially their behavior at high field intensity or high pressure and density. The analytic solutions are thus complementary to the numerical solutions obtained with realistic equations of state. Indeed in the pioneering work on neutron stars by Oppenheimer and Volkoff[2] appeal was made to qualitatively similar closed analytic solutions obtained by Tolman for an understanding of such features as a maximum mass and infinite central densities. Similarly the classic interior solution of Schwarzschild for an incompressible fluid has provided insight into the effects of relativity on qualitative features and order of magnitude quantitative features of white dwarfs and neutron stars.[4] Indeed the density of the heaviest neutron star models is roughly constant throughout most of the star, making the Schwarzschild interior solution a surprisingly good approximation considering its extreme simplicity.

The following questions are particularly interesting regarding analytic solutions for a static fluid sphere:

(1) Does a maximum mass occur?

(2) Can the central density and pressure become infinite?

(3) Can the velocity of sound exceed c, thereby violating causality?

(4) Can the average polytropic index γ become less than 4/3, making the solution unstable to radial oscillation?

In the present work we generate a closed analytic solution to the Einstein equations for a fluid sphere by specifying the metric term $g_{00}=e^{\nu}$. All of the functions involved in the solution are algebraic, and a physically reasonable equation of state results. The solution has the following properties, corresponding to the questions noted above:

(1) A maximum mass occurs, and is equal in geometric units to 2/5 the fluid radius. This may be compared to the Schwarzschild interior solution, in which the demand that the Schwarzschild radius not be exterior to the fluid leads to a maximum mass of 1/2 the fluid radius.[3,4]

(2) The central pressure and density both become infinite as the maximum mass is approached. This may also be compared with the behavior of the Schwarzschild interior solution, in which the central pressure becomes infinite when the mass reaches 4/9 the fluid radius, the density, of course, remaining constant.[3,4]

(3) The velocity of sound exceeds c in portions of the fluid when the geometrical mass is about 0.35 times the fluid radius. Thus the demand of causality provides a more stringent upper limit on the mass than occurs in (1) or (2) above.

(4) The average polytropic index γ remains greater than 4/3 for all masses less than about 0.35 times the fluid radius.

A further remarkable property of the present solution is that in the low mass limit ($\ll$ the fluid radius) it is identical to the Schwarzschild interior solution and also to one of Tolman's analytic solutions. This is despite the fact that the Schwarzschild interior solution has no equation of state and, for example, cannot support sound waves. It thus appears that the present solution is in several respects superior to the Schwarzschild solution for illustrating some of the peculiarly relativistic qualitative features of a fluid sphere.

2. SOLUTION BY QUADRATURES OF THE EINSTEIN EQUATIONS

The Einstein equations for an ideal fluid are[4]

$$G_{\mu\nu}=-8\pi[\rho u_\mu u_\nu - p(g_{\mu\nu}-u_\mu u_\nu)], \tag{2.1}$$

where $G_{\mu\nu}$ is the Einstein tensor, u_μ is the 4-velocity of a fluid element, and $g_{\mu\nu}$ is the metric. (We set c and the gravitational constant κ equal to 1 by choice of units and specify a zero cosmological constant.) For a static spherically symmetric system an appropriate metric is

$$ds^2=e^{\nu(r)}dt^2-e^{\lambda(r)}dr^2-r^2(d\theta^2+\sin^2\theta\,d\varphi^2). \tag{2.2}$$

It then follows that the field equations may be written

as[1,4]

$$\frac{e^{\lambda}}{r^2} = \frac{1}{r^2} - \frac{\nu'^2}{4} - \frac{\nu''}{2} + \frac{\nu'\lambda'}{4} + \frac{\nu'+\lambda'}{2r}, \tag{2.3a}$$

$$8\pi p = -\frac{1}{r^2} - e^{-\lambda}\left(\frac{1}{r^2} + \frac{\nu'}{r}\right), \quad p = p(r), \tag{2.3b}$$

$$8\pi\rho = \frac{1}{r^2} - e^{-\lambda}\left(\frac{1}{r^2} - \frac{\lambda'}{r}\right), \quad \rho = \rho(r), \tag{2.3c}$$

where a prime denotes differentiation with respect to r. It is well known that (2.3a) may be solved by quadratures in a number of ways; e.g., Tolman[1] specifies various conditions on the functions ν and λ that simplify the equation and allow immediate integration, while Adler, Bazin, and Schiffer[4] note that λ may be obtained if ν is given. Once ν and λ are obtained, p and ρ follow directly from (2.3b) and (2.3c). It is rather remarkable how simple the explicit solution of (2.3a) can be made. We define

$$\gamma(r) = e^{\nu/2} = \sqrt{g_{00}}, \tag{2.4a}$$

$$\tau(r) = e^{-\lambda} = -1/g_{11}. \tag{2.4b}$$

Then (2.3a) may be written as a linear first order equation for γ

$$\gamma' - \frac{2(\gamma + r\gamma' - r^2\gamma'')}{r(\gamma + r\gamma')}\gamma = \frac{-2\gamma}{r(\gamma + r\gamma')}. \tag{2.5}$$

This has the solution

$$\gamma(r) = \exp[-F(r)]\left\{\int^r \exp[F(r')]\, g(r')\, dr' + C\right\},$$
$$F(r) = \int^r f(r')\, dr', \quad f(r) = \frac{-2(\gamma + r\gamma' - r^2\gamma'')}{r(\gamma + r\gamma')}, \tag{2.6}$$
$$g(r) = \frac{-2\gamma}{r(\gamma + r\gamma')}, \quad C = \text{const.}$$

It is clear that (2.6) combined with (2.3b) and (2.3c) represents all solutions for static spherically symmetric fluid bodies. There is of course no reason to expect that all such solutions will be physically reasonable and have, for example, a positive ρ and p distribution. Only a subclass of these solutions, corresponding to certain functions $\gamma(r)$, will be physically reasonable in this sense, and a still smaller subclass will correspond to physically reasonable equations of state. A judicious choice of $\gamma(r)$ is thus necessary for a physically interesting solution.

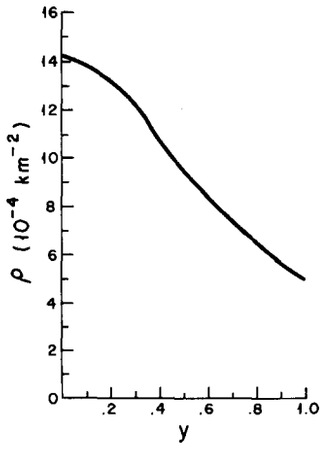


FIG. 1. Density distribution ρ for $m_0 = 3$ km and $r_0 = 10$ km.

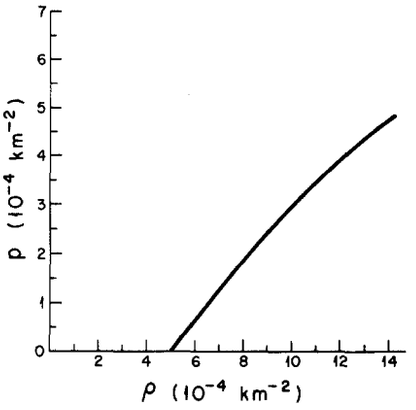


FIG. 2. Equation of state $p(\rho)$ for $m_0 = 3$ km and $r_0 = 10$ km.

The solution (2.6) may be rewritten in many forms by transforming the independent variables γ and τ defined in (2.4); for example, rewriting (2.5) in terms of γ^2 leads to a solution analogous to (2.6), from which Tolman's solution number 4 may be most easily obtained. Alternatively the dependent variable r may be transformed; the choice of $\rho = \log r$ leads to an analogue of (2.6) from which Tolman's solution number 5 may be obtained.

3. SPECIFIC ANALYTIC SOLUTION

We now wish to find a solution of (2.6) that is both simple and yields physically interesting results. The solution (2.6) for $\tau(r)$ will be particularly simple if $f = g$, since in this case the integral in (2.6) will be $\exp[F(r)]$. This is accomplished by demanding that $r\gamma' - r^2\gamma''$ vanish or

$$\gamma(r) = A + Br^2, \tag{3.1}$$

in which case

$$\tau(r) = 1 + Cr^2/(A + 3Br^2)^{2/3} \tag{3.2}$$

The constants A, B, and C are specified by matching the solution to the exterior Schwarzschild solution for a mass m_0, at radius r_0. We then obtain as the solution the following algebraic functions:

$$\gamma = \exp(\nu/2) = (1 - \tfrac{5}{2}\epsilon + \tfrac{1}{2}\epsilon y^2)(1 - 2\epsilon)^{-1/2}, \tag{3.3a}$$

$$\tau = \exp(-\lambda) = 1 - 2\epsilon y^2 (1 - \epsilon)^{2/3}(1 - \tfrac{5}{2}\epsilon + \tfrac{3}{2}\epsilon y^2)^{-2/3}, \tag{3.3b}$$

$$\rho = (\epsilon/4\pi r_0^2)(1 - \epsilon)^{2/3}(1 - \tfrac{5}{2}\epsilon + \tfrac{3}{2}\epsilon y^2)^{-2/3} \times [3 - 2\epsilon y^2/(1 - \tfrac{5}{2}\epsilon + \tfrac{3}{2}\epsilon y^2)], \tag{3.3c}$$

$$p = (\epsilon/4\pi r_0^2) \times [\exp(-\lambda)(1 - \tfrac{5}{2}\epsilon + \tfrac{1}{2}\epsilon y^2)^{-1} - (1 - \tfrac{5}{2}\epsilon + \tfrac{3}{2}\epsilon y^2)^{-2/3}], \tag{3.3d}$$

where $\epsilon \equiv m_0/r_0$ and $y \equiv r/r_0$. The mass distribution, defined as $m(r) = [1 - \exp(-\lambda)]r/2$, is

$$m(r) = m_0 y^3 (1 - \epsilon)^{2/3}(1 - \tfrac{5}{2}\epsilon + \tfrac{3}{2}\epsilon y^2)^{-2/3}. \tag{3.4}$$

In Figs. 1 and 2 we have plotted $\rho(r)$ as well as the equation of state p vs ρ for this solution for the values $m_0 = 3$ km and $r_0 = 10$ km, characteristic of a heavy neutron star. Unlike Schwarzschild's solution with $\rho = \text{const}$, the density peaks sharply at $r = 0$, and the equation of state is physically reasonable. It is evident, moreover, that ρ and p will in general be positive functions.

4. PROPERTIES OF THE SOLUTION

For a fixed radius the sphere described by (3.3) has a maximum mass. We observe that the central density $\rho(0)$ is a function of $\epsilon = m_0/r_0$.

$$\rho(0) = \frac{3\epsilon}{4\pi r_0^2}\left(\frac{1-\epsilon}{1-\frac{5}{2}\epsilon}\right)^{2/3} \tag{4.1}$$

so that for fixed r_0 the mass m_0 is a function of $\rho(0)$. This function has a maximum value when $d\rho(0)/d\epsilon$ is infinite, which occurs for

$$\epsilon = \tfrac{2}{5}, \quad m_0 = \tfrac{2}{5} r_0. \tag{4.2}$$

This may be compared with the larger value of $\frac{1}{2}r_0$ obtained with the Schwarzschild interior solution by demanding that the Schwarzschild radius not be exterior to the fluid.[4]

From (3.2c) and (3.3d) it is evident that both $\rho(0)$ and $p(0)$ become infinite at the same value of $\epsilon = 2/5$. The Schwarzschild interior solution has the similar property that the central pressure becomes infinite when $\epsilon = 4/9$, a less stringent mass limit than given by the present model.[4]

A further interesting property of the present solution is that for $\epsilon \gtrsim 0.35$ the derivative of p with respect to ρ exceeds unity in parts of the fluid. This may be shown by calculating $dp/d\rho = (dp/dy^2)/(d\rho/dy^2)$ and evaluating it as a function of ϵ and y. We may interpret this to mean that the speed of sound, given by $dp/d\rho$, exceeds the velocity of light and the fluid becomes noncausal. We thereby infer a yet more stringent maximum mass of about $0.35 r_0$. This is in fact comparable to the value obtained for realistic models of neutron stars.[5]

Let us now procede to the low mass limit $m_0 \ll r_0$. The constant density solution of Schwarzschild[4] has

$$\exp(\nu/2) = [\tfrac{3}{2}(1-2\epsilon) - \tfrac{1}{2}(1-2\epsilon y^2)] \tag{4.3}$$

and Tolman's solution (number four) has[1]

$$\exp(\nu) = (1-3\epsilon) + \epsilon y^2. \tag{4.4}$$

Both of these solutions and the present solution have the following common limit functions for small ϵ:

$$\exp(\nu) = 1 - 3\epsilon + \epsilon y^2, \quad \exp(-\gamma) = 1 - 2\epsilon y^2,$$

$$\rho = 3\epsilon/4\pi r_0^2, \quad p = \tfrac{1}{2}\rho\epsilon(1-y^2). \tag{4.5}$$

The Schwarzschild solution has $\rho \equiv \text{const}$ for *all* values of ϵ. For the present solution however we may expand ρ to second order in ϵ to obtain

$$\rho = (3\epsilon/4\pi r_0^2)(1 + \epsilon - \tfrac{5}{3}\epsilon y^2). \tag{4.6}$$

We then may use this and the expression for p above to eliminate y^2 and obtain an approximate equation of state for small ϵ

$$p = \tfrac{3}{10}(\rho - \epsilon/4\pi r_0^2). \tag{4.7}$$

Note that ϵ is contained in this relation explicitly. The same is true of the implicit exact relation contained in (3.3c) and (3.3d). The solution of Tolman has an equation of state that also depends explicitly on the mass and radius of the fluid.

5. CONCLUSIONS

Using a quadrature solution of the Einstein equations for a static fluid sphere we have obtained an interesting analytic solution with the following properties: (1) The structure of the fluid is expressible in terms of algebraic functions. (2) The fluid has a density that peaks in the center, and a maximum value of mass of about 0.35 the radius, dictated by causality. (3) The present solution, one obtained by Tolman, and the Schwarzschild interior solution are the same for asymptotically small values of mass/radius, but equations of state exist for our solution and Tolman's solution; these equations of state depend explicitly on the value of the mass and radius of the sphere.

ACKNOWLEDGMENTS

We would like to thank Charles Sheffield and J.A. Wheeler for valuable suggestions and acknowledge stimulating discussions with F. Dyson, P. Solomon, and J.M. Cohen.

[1] R. Tolman, Phys. Rev. 55, 364 (1939).
[2] J.R. Oppenheimer and G.M. Volkoff, Phys. Rev. 55, 374 (1939).
[3] C. Misner, K. Thorne, J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), see Chap. 23.
[4] R.J. Adler, M. Bazin, and M.M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965, 1974), see Chap. 9 of 1st ed. (1965) or Chap. 14 of 2nd ed. (1974).
[5] J.M. Cohen and A.G.W. Cameron, Astrophys. Space Sci. 10, 227 (1971).

A semi-Euclidean approach to boson–fermion model theories*

David Brydges†

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104

Paul Federbush‡§

Institute for Advanced Study, Princeton, New Jersey 08540
(Received 17 January 1974)

A formulation is presented for the study of semiboundedness of coupled boson–fermion model field theories. Euclidean–boson fields and ordinary fermion fields are employed. Expansion steps used to derive estimates are presented.

I. INTRODUCTION

There is a great deal of interest at present in discovering techniques for treating boson—fermion model field theories parallel to the use of Euclidean boson fields in studying purely boson models. We address ourselves here solely to the question of semiboundedness of the energy (as the first problem usually encountered for any model) although there is no reason to exclude further applications of the machinery discussed. There are three superrenormalizable models available, Y_2, Y_3, and the generalized Yukawa model in one space dimension (hereafter called GY_2). The treatment of Y_2 and GY_2 is in some sense just practice for the tackling of Y_3. Four-dimensional theories so far appear impregnable.

Glimm obtained semiboundedness of the energy for Y_2 in Ref. 1. Schrader extended this result to show the linear dependence of the bound on the volume.[2] One of the authors showed the semiboundedness of the GY_2 energy.[3] There are studies under way attempting to study boson—fermion field theory models by eliminating the fermi fields initially, using the closed form expression involving a Fredholm determinant, similar to the corresponding expression in the variational approach to field theory.[4,5] Here we continue the development initiated in Ref. 3. A unified treatment of Y_2 and GY_2 is obtained,[6] whose basic line is here presented. Whether these methods, or the methods in Ref. 5, will be successful in studying Y_3 must be decided in the future. Other paths of evolution, or unifications, cannot be excluded, such as the work of Gross.[7] We are enthusiastic about the usefulness of the present program since it captures for boson—fermion models analogs of all the techniques used by Glimm and Jaffe in obtaining semiboundedness for ϕ_3^4, including localization.[8]

II. FEYNMAN-KAC FORMULA

Any Hamiltonian we consider is of the form

$$H = H_{0B} + H_{0F} + G(\phi) + \int dx \int dy\, Q(x, y, \phi)\, \bar{\psi}(x)\, \psi(y) \quad (1)$$

$$= H_{0B} + K_F. \quad (2)$$

There are volume and momentum cutoffs in the interaction and renormalization terms in the $G(\phi)$. Subscripts F will often denote expressions in terms of Foch space operators. Using the Trotter product formula, we have

$$\langle 0 | \exp(-HT) | 0 \rangle_F = \lim_{n\to\infty} \langle 0 | [\exp(-H_{0B}T/n) \exp(-K_F T/n)]^n | 0 \rangle_F. \quad (3)$$

$|0\rangle_F$ denotes the Foch vacuum. We introduce a total Hilbert space $\mathcal{H}$, the tensor product of Euclidean boson Hilbert space $\mathcal{H}_{EB}$ and $\mathcal{H}_F$ the Foch fermi Hilbert space

$$\mathcal{H} = \mathcal{H}_{EB} \times \mathcal{H}_F \quad (4)$$

and Euclidean boson fields $\phi(x,t)$. We also introduce dummy variables into the Fermi fields

$$\psi(x,t) = \psi(x). \quad (5)$$

These dummy t variables will only be used to define a time ordering operation. (One may alternatively say we are developing a Euclidean Fermi field theory—translation invariant but not rotation invariant—with the zero operator generating time translations. The Fermi kinetic energy terms are included in the interaction; in form, they and the interaction terms do not appreciably differ. This may be contrasted with the boson situation where the energy contains π's and the interaction does not.) K_F is replaced by $K(t)$ by substituting the time dependent fields for the $(t=0)$ Foch field:

$$K_F = H_{0F}(\bar{\psi}, \psi) + G(\phi) + \int dx \int dy\, Q(x, y, \phi)\, \bar{\psi}(x)\, \psi(y)$$

$$K(t) = H_{0F}(\bar{\psi}(t), \psi(t)) + G(\phi(t)) + \int dx \int dy\, Q(x, y, \phi(t))\, \bar{\psi}(x,t)\, \psi(y,t). \quad (6)$$

Equation (3) becomes

$$\langle 0 | \exp(-HT) | 0 \rangle_F = T \langle 0 | \exp[-\textstyle\int_0^T K(t)\, dt] | 0 \rangle. \quad (7)$$

Here $|0\rangle$ is the vector in $\mathcal{H}$ that is the product of the boson Euclidean space vacuum with the Fermi Foch space vacuum. T indicates a time ordering in the t variables in the $\phi(x,t)$ and $\psi(x,t)$. All of our efforts are directed to finding techniques for estimating the right side of Eq. (7).

III. THE DUHAMEL EXPANSION

The process we have for removing parts of the exponent is the Duhamel expansion. We decompose $K(t)$ into two parts:

$$K(t) = K_0(t), \quad K(t) = K_i(t) + R_i(t), \quad i = 1, 2, \ldots, \quad (8)$$

where $K_i(t)$ and $R_i(t)$ are functions of $\psi(t)$, $\bar{\psi}(t)$, and $\phi(t)$ $\phi(t)$—all the fields at the fixed time t. Often in applications the $K_i(t)$ and $R_i(t)$ are picked to have no explicit time dependence. The Duhamel expansion assumes the form

$$T\langle 0 | \exp[-\textstyle\int_0^T K(t)\, dt] R | 0 \rangle = \sum_0^\infty (-1)^n \int_0^T dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1$$

$$\times T\langle 0|\prod_{i=1}^{n} R_i(t_i)\exp\left[-\sum_{j=0}^{n}\int_{t_j}^{t_{j+1}} K_j(s)\,ds\right]R|0\rangle, \tag{9}$$

where $t_0=0$ and $t_{n+1}=T$. [The $K(t)$ in (8) and (9) is not necessarily the same as in (6), but may be a similar expression such as one of the $K_i(t)$ arising in an inductive procedure.] An example of this expansion for $P(\phi)_2$ is found in Ref. 9 and for GY_2 in Ref. 3.

If space—time is divided into regions and a separate Duhamel expansion is developed for the interaction in each region, then the different Duhamel expansions can be combined into a sum of single Duhamel expansions such as (9). This is a primary device for localization.

IV. THE PULL THROUGH EXPANSION

The "pull through" operation was introduced in Ref. 10. Like the Duhamel expansion it is purely algebraic and applies alike to Fermi and boson fields. An operator in some $R_i(t_i)$ in (9) is decomposed into creation and annihilation operators which are "pulled through" until they either annihilate on the vacuum, contract on the exponent, contract on some other $R_j(t_j)$, or until the operator being pulled through has moved far enough to collect some desirable time factor, $\exp(-ut)$, and then stopped. This last operation is not used for bosons. It is possible to iteratively use pull through operations and Duhamel expansions, to generate an inductive procedure.

After any number of applications of the two operations above one has an expression

$$\langle 0|\exp(-HT)|0\rangle_F=\sum_\alpha T_\alpha$$

where a typical term T_α has the form

$$T_\alpha=\int_0^T dt_r\int_0^{t_r} dt_{r-1}\cdots\int_0^{t_2} dt_1 T\langle 0|R_\alpha\exp(-K_\alpha)|0\rangle \tag{10}$$

with

$$K_\alpha=\sum_{j=0}^{r}\int_{t_j}^{t_{j+1}} K_j(s)\,ds \tag{11}$$

and

$$R_\alpha=\int dx_1\cdots dx_s\int dy_1\cdots dy_s\,\bar\psi(x_1,t_{\alpha(1)})\cdots\bar\psi(x_s,t_{\alpha(s)})$$
$$\times\cdots\psi(y_s,t_{\alpha(2s)})\,Q_\alpha(x_1,\ldots,x_s,y_1,\ldots,y_s,\phi,t). \tag{12}$$

Summation over Fermion field indices is always implicit. As many of the time integrations as possible are included in Q_α. Time variables arising from contractions from the exponent in which both fermions are contracted fall in this category. Combination of terms in T_α is also advantageous; in the pull through procedure it is possible to construct time and space locally averaged boson fields as in Ref. 8 which would appear in Eq. (12).

V. ESTIMATES AND DEFERMIATION

When the algebraic operations of the last two sections are completed, estimates are required for $K_j(s)$ and R_α. Assume estimates of the form

$$K_j(s)\geq C_j(\phi,s) \tag{13}$$

and

$$|Q_\alpha(x,y,\ldots)|_1\leq d_\alpha(\phi,t_1,\ldots,t_r), \tag{14}$$

where $|f(x_1,\ldots,x_n)|_1$ is the inf of $\sum_1^\infty|a_i|$ over a_i satisfying

$$f=\sum_i a_i\prod_{j=1}^{n} g_{ij}(x_j)$$

with $|g_{ij}(x)|_2=1$. The "defermiation" step is then the estimate:

$$\langle 0|\exp(-HT)|0\rangle_F\leq\sum_\alpha|T_\alpha|, \tag{15}$$

$$|T_\alpha|\leq\int_0^T dt_r\cdots\int_0^{t_2} dt_1\langle 0|d_\alpha(\phi)$$
$$\times\exp-\sum_{j=0}^{r}\int_{t_j}^{t_{j+1}} C_j(\phi,s)\,ds\,|0\rangle. \tag{16}$$

In (16) only boson fields remain, and all the techniques for estimating such a purely boson expression are available. Unlike the algebraic operations discussed above the defermiation can be performed just once in the procedure, it is a decisive step.

In Ref. 3 is an illustrative use of a Duhamel expansion, pull throughs, the estimates of Eq. (13), defermiation, and estimation of Eq. (16). There is one important technical improvement here over Ref. 3, the use of $|\ |_1$ estimates for Q_α. The estimate procedure in Ref. 3 is adequate to obtain semiboundedness for Y_2 or GY_2 in a finite volume, but yields an incorrect volume dependence. The present procedure behaves correctly under localization and therefore is the correct one to use for obtaining the volume dependence and attempting Y_3.

The statement that the estimates behaves correctly under localization is easiest to explain in the case when all the K_i contain only the fermion kinetic energies (a heuristic example). Then in estimates (15) and (16) the terms involving only contractions between operators lying in the same space—time squares contribute to the sum $\sum_\alpha|T_\alpha|$ an expression of the form

$$\sum_{\text{nonoverlap}}|T_\alpha|\leq\langle 0|\prod_\Delta D_\Delta(\phi)|0\rangle, \tag{17}$$

where $D_\Delta(\phi)$ are corresponding estimates for the squares Δ. When the K_i contain other than just energies the localization property imposes conditions on the form of estimate (13)—the right side must be a sum of adequate estimates for the individual squares cut at $t=s$. Localization methods as used in Ref. 2 are valuable to achieve this.

VI. DISCUSSION

We say a few words about the treatment of Y_2 and GY_2. In these models in each unit space time block the Duhamel expansion may be performed just once—no induction is necessary. The interaction terms are included in the K_i with an upper momentum cutoff on the fermions increasing with i. (Alternate developments are possible.) The pull throughs are used to exhibit the renormalization cancellation. Additional pull throughs are required also; those for GY_2 are slightly different from those in Ref. 3 since $|\ |_1$ estimates are used. In particular each vertex (basic interaction term not in the exponent) must be connected to at least one other vertex by a fermion line, however the number of contractions is to be limited. In any expansion in which no fermion operators other than the kinetic energy appear in the

exponents, the whole procedure could have been performed using Osterwalder—Schrader fields.[4]

It is interesting to consider what special properties of fermions are used in the above program. One could have derived the same formulas for a boson ψ field, except Eq. (14). The fermion nature has been used so far in three ways (two of these ways only implicit in this paper):

(1) To derive Eq. (14) one has used that $|\psi(f)| \leq |f|_2$.

(2) To derive in Eq. (13) a useful estimate for the Y_2 or GY_2 scattering terms the free $N_{0F}^{-1/2}$ factor in N_τ estimates with fermions is useful.

(3) To derive in Eq. (13) a useful estimate for the Y_2 or GY_2 creation and annihilation terms, employing as in Ref. 1 a partial dressing for the fermions (see Ref. 3), the sign of a term arising from the anticommutativity is crucial. This sign is available in other models and other dressings.

Possibly to treat Eq. (13) for Y_3 more properties will be discovered, though this may not be necessary. In any case the exchange of boson commutativity for these three properties, a three for one deal, may not be a bad trade.

We feel that the approach of this paper provides sufficiently powerful machinery to consider an attack on the Y_3 problem and that it may be as close as one can come to realizing for fermions a Euclidean formulation for performing estimates.

ACKNOWLEDGMENTS

One of the authors (P. F.) would like to thank Professor J. Glimm for a helpful discussion and is grateful to Dr. Carl Kaysen for the hospitality at the Institute for Advanced Study.

*This work was supported in part by NSF Grant GP-17523.
†Michigan Junior Fellow.
‡John Simon Guggenheim Memorial Fellow.
§Permanent address: Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104.

[1]J. Glimm, Comm. Math. Phys. **5**, 343 (1967); **6**, 120 (1967).
[2]R. Schrader, Ann. Phys. (N.Y.) **70**, 412 (1972).
[3]P. Federbush, J. Math. Phys. **14**, 1532 (1973).
[4]K. Osterwalder and R. Schrader, "Euclidean Fermi Fields and a Feynman—Kac Formula for Boson—Fermion Models," Helv. Phys. Acta, to appear.
[5]K. Osterwalder and E. Seiler, private communication.
[6]D. Brydges, in preparation.
[7]L. Gross, J. Funct. Anal. **10**, 52 (1972).
[8]J. Glimm and A. Jaffe, "Positivity of the ϕ_3^4 Hamiltonian," to be published.
[9]J. Glimm and A. Jaffe, "Positivity and Self Adjointness of the $P(\phi)_2$ Hamiltonian," to be published.
[10]J. Glimm and A. Jaffe, J. Math. Phys. **13**, 1568 (1972).

A pair of coupled quantum anharmonic oscillators*

Francis R. Halpern

Department of Physics, University of California, San Diego, La Jolla, California 92037
(Received 23 October 1973)

The energy levels of a pair of coupled anharmonic oscillators are studied. The technique employed is to find two approximately canonical coordinate momentum pairs. Particular emphasis is placed on the qualitative dependence of the coordinate and momentum operators on the quantum numbers.

I. INTRODUCTION

Quantum mechanics deals with operators rather than numbers as classical mechanics does. In this sense it its a technically more difficult subject. Perhaps though in a deeper sense it is much harder because it is harder to have a good intuitive feeling about what an interaction means in quantum mechanics.

Newtonian mechanics is already somewhat removed from common sense. For thousands of years it seemed obvious that the exertion of effort produced velocity or displacement rather than acceleration. The quantum statement that when I exert effort I produce some collection of matrix elements is an abstraction that I really cannot make. Thus my intuitions in quantum mechanics are wholly objective. They never follow from direct personal involvement. I have tried with some success to train myself to think about matrix elements, but the only success I have is in treating them as abstract quantities. quantities.

I shall be primarily concerned with finding energy levels. A solution to this problem consists in diagonalizing the Hamiltonian. Less formally, the Hamiltonian is a function usually a polynomial of various coordinate and momentum operators. Somehow the matrix elements of these operators enter the off-diagonal elements of the Hamiltonian with different signs in such a way as to cancel each other out.

There are two sets of equations to be solved in a quantum mechanics problem. The commutator equations necessary to insure the operators x and p are a canonical pair and the equations necessary to make the off-diagonal Hamiltonian matrix elements vanish. The first equations to solve are the diagonal commutator equation and the equations that insure the vanishing of the Hamiltonian matrix elements adjacent to the principal diagonal.

I have found that in one-dimensional oscillator equations this type of solution can generate a step-by-step procedure that yields an accurate solution and offers a good insight into the dependences of the matrix elements on the parameters of the problem.[1] In the present work we explore the lowest order solution of this type for a coupled problem.

The most primitive understanding of a quantum problem is to see just which operators enter a particular Hamiltonian matrix elements with what signs to permit this cancellation without any regard for the magnitudes of the operators. At a more quantitative level one may attempt to determine the dominant behavior of the various coordinate and momentum operator matrix elements. A definitive solution to a problem results from a specification of how the more delicate dependences of the coordinate and momentum operators are to be systematically determined and finally what they are.

Here I have divided the quantum problem into two parts. The first discussed above is to operate with a mathematical intuition to determine the proper form for the coordinate and momentum operators. The second part, which I do not discuss here, would be to relate the calculated matrix elements to physically observable properties. The problems that I would like to solve—field theory—is sufficiently complicated so that I can only consider simple models at the present time and the physical interpretation of my current results is not of great interest to me.

I have tried to adopt this point of view and to instruct my intuition accordingly. I think the effort can be of some help in gaining a qualitative feeling for what happens in quantum mechanical problems. I would like to emphasize one feature of the quantum problem which seems of central importance to me. Most wavefunctions extend over all of configuration space. Thus a wavefunction responds to *all* of a potential not to its values in some region. A specific number may be related to values of the potential in a region of configuration space, but the qualitative character of a wavefunction and the solution to a quantum problem depends on an entire potential. It enforces a uniformity condition in the mathematician's sense of uniformity on arguments about quantum mechanics. Specifically a perturbation must be uniformly small or it is *not* small.

I would like to consider this notion in relation to perturbed harmonic oscillators from the problems of a single oscillator to the problems of interacting scalar fields. A typical theory of this type has a potential with an harmonic part together with so-called perturbing terms. The perturbing terms involve powers of the field or coordinate operators higher than the second. The first claim is that the anharmonic terms are not uniformly small. In fact, over most of configuration space the anharmonic part of the potential is larger than the harmonic part. Thus such theories are never weak coupling theories. Attempts to estimate the numerical values of some quantities, e.g., the ground state energy of an anharmonic oscillator, may be carried through successfully with a weak coupling technique, but no genuine insight into the character of the solution can be obtained this way.

A much more severe problem can occur in a problem with many degrees of freedom or a field theory with an infinite number of degrees of freedom. In these cases the theory becomes effectively a strong coupling theory no matter what number multiplies the potential. This follows because not only are most of the off-diagonal matrix element of the Hamiltonian larger than the diag-

onal ones but because they are also much more numerous than the diagonal ones. Thus any field theory of this type is automatically a strong coupling theory. The harmonic term is really the perturbation.

To gain a real understanding of such a theory, the anharmonic terms should be treated first. In the following exercise I consider the oscillator with Hamiltonian H

$$H=\tfrac{1}{2}(P_1^2+P_2^2+x_1^2+x_2^2)+Ax_1^4+Bx_1^2x_2^2+Cx_2^4.$$

The choice is the simplest one in which the two anharmonic oscillators interact. With just A or C different from zero the oscillator would decouple. The technique I shall employ is to assume that the various coordinate and momentum operators only have matrix elements that connect states in which one quantum number is changed by unity. This is certainly not a correct representation of the situation. The only justification for this procedure is that for a single uncoupled anharmonic oscillator it is apparently a good first approximation.[2]

In the next sections some elementary algebra of quartic polynomials is studied. In Sec. III the assumptions are used to formally determine energy levels and matrix elements. The real concern in the solution is to try and view the quantum problem as an infinite set of algebraic equations and to throw away most of them. In the ensuing set one tries to see how cancellations can be made to occur in off-diagonal positions of the Hamiltonian.

II. POSITIVE DEFINITENESS OF THE POTENTIAL

The fourth degree potential V is given by the homogeneous polynomial

$$V(x_1,x_2)=Ax_1^4+4Bx_1^3x_2+6Cx_1^2x_2^2+4Dx_1x_2^3+Ex_2^4.$$

In this section the condition that V be positive for all x_1 and x_2 are considered. It is clear that A and E must both be positive. If there is any value of (x_1,x_2) for which V is negative, then it is also negative and has an arbitrarily large negative value by considering the sequence of points $(\lambda x_1,\lambda x_2)$.

Let the ratio of (x_1/x_2) be ξ. The polynomial

$$P=\xi^4+(4B/A)\xi^3+(6C/A)\xi^2+(4D/A)\xi+E/A$$

must be positive for all real values of ξ. If α, β, γ, δ are the roots of P, then the relations

$$\alpha+\beta+\gamma+\delta=-(4B/A)$$
$$\alpha\beta+\alpha\gamma+\alpha\delta+\beta\gamma+\beta\delta+\gamma\delta=6C/A,$$
$$\alpha\beta\gamma+\alpha\beta\delta+\alpha\gamma\delta+\beta\gamma\delta=-4D/A,$$
$$\alpha\beta\gamma\delta=E/A$$

hold. The quantities

$$q=(B/A)^2-(C/A),$$
$$c=D/A)-3(BC/A^2)+2(B/A)^3,$$
$$b=(E/A)-4(BD/A^2)+3(C/A)^2$$

are functions of the differences of the roots; they are invariant under the translation $\xi\to\xi+\Delta$. They are called q, c, and b because q is quadratic, c cubic, and biquadratic in the roots. Explicitly they are given by

$$q=\tfrac{1}{16}(\alpha^2+\beta^2+\gamma^2+\delta^2)-\tfrac{1}{24}(\alpha\beta+\alpha\gamma+\alpha\delta+\beta\gamma+\beta\delta+\gamma\delta)$$
$$=\tfrac{1}{48}[(\alpha+\beta-\gamma-\delta)^2+(\alpha-\beta+\gamma-\delta)^2+(\alpha-\beta-\gamma+\delta)^2],$$
$$c=\tfrac{1}{32}(-\alpha^3-\beta^3-\gamma^3-\delta^3+\alpha^2\beta+\alpha^2\gamma+\alpha^2\delta+\alpha\beta^2+\beta^2\gamma+\beta^2\delta+\alpha\gamma^2+\beta\gamma^2+\gamma^2\delta+\alpha\delta^2+\beta\delta^2+\gamma\delta^2-2\alpha\beta\gamma-2\alpha\beta\delta-2\alpha\gamma\delta-2\beta\gamma\delta)$$
$$=-\tfrac{1}{32}(\alpha+\beta-\gamma-\delta)(\alpha-\beta+\gamma-\delta)(\alpha-\beta-\gamma+\delta),$$
$$b=\tfrac{1}{12}(\alpha^2\beta^2+\alpha^2\gamma^2+\alpha^2\delta^2+\beta^2\gamma^2+\beta^2\delta^2+\gamma^2\delta^2-\alpha^2\beta\gamma-\alpha^2\beta\delta-\alpha^2\gamma\delta-\alpha\beta^2\gamma-\alpha\beta^2\delta-\beta^2\gamma\delta-\alpha\beta\gamma^2-\alpha\gamma^2\delta-\beta\gamma^2\delta-\alpha\beta\delta^2-\alpha\gamma\delta^2-\beta\gamma\delta^2+6\alpha\beta\gamma\delta)$$
$$=\tfrac{1}{24}[(\alpha-\beta)^2(\gamma-\delta)^2+(\alpha-\gamma)^2(\beta-\delta)^2+(\alpha-\delta)^2(\beta-\gamma)^2].$$

If the polynomial P has a double root, the discriminant of P given by

$$d=(\alpha-\beta)^2(\alpha-\gamma)^2(\alpha-\delta)^2(\beta-\gamma)^2(\beta-\delta)^2(\gamma-\delta)^2$$

must vanish. The discriminant d can be expressed in terms of the coefficients of P or more compactly in terms of the invariants q, c, and b:

$$d=(24b)^3/54-\tfrac{1}{27}[27(4c)^2-(12q)^3+\tfrac{3}{2}(12q)(24b)]^2$$

There are three possible configurations for the roots of P:

(1) all real roots,
(2) two real roots and a complex conjugate pair,
(3) two complex conjugate pairs.

In order that P always be positive, it is necessary that all the roots be complex. The transition between the three cases occurs when $d=0$. This is because a complex conjugate pair becomes a real double root at the transition between the three cases. A simple calculation shows that

(1) $d(a,b,c,d)>0$ if a,b,c,d are real,
(2) $d(a,b,c+id,\ c-id)<0$,
(3) $d(a+ib,\ a-ib,\ c+id,\ c-id)>0$.

Thus in order that P be everywhere positive it is necessary but not sufficient that $d>0$. Cases (1) and (3) can be separated by considering the expression

$$\sqrt{12b}-12q,$$

In case (1) where the roots α, β, γ and δ are real, b can be written as

$$0\leq 12b$$
$$=\tfrac{1}{2}[(\alpha-\beta)^2(\gamma-\delta)^2+(\alpha-\gamma)^2(\beta-\delta)^2+(\alpha-\delta)^2(\beta-\gamma)^2]$$
$$=\tfrac{1}{32}\{[(\alpha-\beta+\gamma-\delta)+(\alpha-\beta-\gamma+\delta)]^2[(\alpha-\beta+\gamma-\delta)-(\alpha-\beta-\gamma+\delta)]^2+[(\alpha+\beta-\gamma-\delta)+(\alpha-\beta-\gamma+\delta)]^2[(\alpha+\beta-\gamma-\delta)-(\alpha-\beta-\gamma+\delta)]^2+[(\alpha+\beta-\gamma-\delta)+(\alpha-\beta+\gamma-\delta)]^2\times[(\alpha+\beta-\gamma-\delta)-(\alpha-\beta+\gamma-\delta)]^2\},$$
$$=\tfrac{1}{32}\{[(\alpha-\beta+\gamma-\delta)^2-(\alpha-\beta-\gamma+\delta)^2]^2+[\alpha+\beta-\gamma-\delta)^2$$

$$-(\alpha-\beta-\gamma+\delta)^2]^2+[(\alpha+\beta-\gamma-\delta)^2-(\alpha-\beta+\gamma-\delta)^2]^2\}$$
$$=\tfrac{1}{16}[(\alpha+\beta-\gamma-\delta)^4+(\alpha-\beta+\gamma-\delta)^4+(\alpha-\beta-\gamma+\delta)^4$$
$$-(\alpha+\beta-\gamma-\delta)^2(\alpha-\beta+\gamma-\delta)^2$$
$$-(\alpha-\beta+\gamma-\delta)^2(\alpha-\beta-\gamma+\delta)^2$$
$$-(\alpha-\beta-\gamma+\delta)^2(\alpha+\beta-\gamma-\delta)^2]$$
$$=\tfrac{1}{16}[(\alpha+\beta-\gamma-\delta)^2+(\alpha-\beta+\gamma-\delta)^2+(\alpha-\beta-\gamma+\delta)^2]^2$$
$$-\tfrac{3}{16}[(\alpha+\beta-\gamma-\delta)^2(\alpha-\beta+\gamma-\delta)^2+(\alpha-\beta+\gamma-\delta)^2$$
$$(\alpha-\beta-\gamma+\delta)^2+(\alpha-\beta-\gamma+\delta)^2(\alpha+\beta-\gamma-\delta)^2]$$
$$=(12q)^2-\tfrac{3}{16}[(\alpha+\beta-\gamma-\delta)^2$$
$$\times(\alpha-\beta+\gamma-\delta)^2+(\alpha-\beta+\gamma-\delta)^2(\alpha-\beta-\gamma+\delta)^2$$
$$+(\alpha-\beta-\gamma+\delta)^2(\alpha+\beta-\gamma-\delta)^2].$$

It follows that

$$(12q)^2\geq 12b\geq 0$$

so that

$$0\geq\sqrt{12b}-12q.$$

On the other hand, for two pairs of complex conjugate roots $\alpha+i\beta$, $\alpha-i\beta$, $\gamma+i\delta$, $\gamma-i\delta$, $12b$ is given by

$$12b=[(\alpha-\gamma)^2+(\beta^2+\delta^2)]^2+12\beta^2\delta^2$$

and $12q$ is given by

$$12q=(\alpha-\gamma)^2-2(\beta^2+\delta^2).$$

The term $12b$ can be rewritten

$$12b=[12q+3(\beta^2+\delta^2)]^2+12\beta^2\delta^2$$

so that if $12q\geq 0$, $\sqrt{12b}\geq 12q$. If, on the other hand, $12q\leq 0$, then $\sqrt{12b}\geq 0\geq 12q$; hence the inequality

$$\sqrt{12b}-12q\geq 0$$

is always valid and the sign of $\sqrt{12b}-12q$ distinguishes between cases (1) and (3). The necessary and sufficient conditions for a stable potential are

$$d>0,\quad \sqrt{12b}-12q>0.$$

III. APPROXIMATE SOLUTION

The oscillator under consideration has the Hamiltonian H given by

$$H=\tfrac{1}{2}(p_1^2+p_2^2+x_1^2+x_2^2)+Ax_1^4+2Bx_1^2x_2^2+Cx_2^4.$$

The invariants q, c, and b of Sec. II for this quartic polynomial are

$$q=-\tfrac{1}{3}(B/A),\quad c=0,\quad b=C/A+\tfrac{1}{3}(B/A)^2.$$

The discriminant d is given by

$$d=256(C/A)[(C/A)-(B/A)^2]^2\geq 0$$

and is indicated is nonnegative. The other quantity needed to test the potential, $\sqrt{12b}-12q$ is given by

$$\sqrt{12b}-12q=\sqrt{12(C/A)+4(B/A)^2}+4(B/A)$$

This quantity has its only root at

$$\sqrt{C/A}+B/A=0$$

so the requirements for a positive potential are $C>0$, $A>0$, and

$$B/A>-\sqrt{C/A}.$$

In this simple case this requirement could have been derived much more directly

The assumed form for the oscillator operators is

$$(p_1)_{n'm',nm}=ia_{nm}\delta_{n'-1,n}\delta_{m'm}-ia_{n-1\,m}\delta_{n'+1,n}\delta_{m'm},$$
$$(p_2)_{n'm'nm}=ic_{nm}\delta_{n'n}\delta_{m'-1\,m}-ic_{n\,m-1}\delta_{n'n}\delta_{m'+1\,m},$$
$$(x_1)_{n'm'nm}=b_{nm}\delta_{n'-1\,n}\delta_{m'm}+b_{n-1\,m}\delta_{n'+1\,n}\delta_{m'm},$$
$$(x_2)_{n'm'nm}=d_{nm}\delta_{n'n}\delta_{m'-1\,m}+d_{n\,m-1}\delta_{n'n}\delta_{m'+1\,m}.$$

The six commutators of these operators are calculated and listed in the Appendix.

For our purposes we consider only the diagonal elements of $[p_1,x_1]$ and $[p_2,x_2]$. These are

$$[p_1,x_1]_{n\,m\,nm}=2i(a_{n-1\,m}b_{n-1\,m}-a_{nm}b_{nm}),$$
$$[p_2,x_2]_{nm,nm}=2i(c_{n\,m-1}d_{n\,m-1}-c_{nm}d_{nm}).$$

In addition to these diagonal matrix elements the commutators have 20 off-diagonal matrix elements that are not explicitly equal to zero. There are 10 conjugate pairs. These matrix elements may not be set equal to zero in the approximation I am considering. The quantities a, b, c, and d would be overdetermined. The quality of the approximation depends in part on the size of these nonvanishing off-diagonal matrix elements of the commutator compared to unity the value of the commutator diagonal matrix elements.

The diagonal equation can easily be summed to give:

$$a_{nm}b_{nm}=\tfrac{1}{2}n,\quad c_{nm}d_{nm}=\tfrac{1}{2}m.$$

The failure of x_1 and x_2 to commute makes the Hamiltonian non-Hermitian. This can be remedied by choosing $3B(x_1^2x_2^2+x_2^2x_1^2)$ or more practically by ignoring the difficulty because the numerical results of the nonhermiticity of lower order than I will consider here.

The Hamiltonian may be expressed in terms of the function a_{nm}, b_{nm}, c_{nm}, and d_{nm}. This result is recorded in the Appendix. Within the limits of the approximation I am considering, it is adequate to neglect differences between the values of function a_{nm} whose indices n and m differ by less than 4. The same assumption will be made for the function b_{nm}, c_{nm}, and d_{nm}. This assumption leads to a greatly simplified Hamiltonian:

$$H_{n'm'nn}$$
$$=Ab_{nm}^4(\delta_{n'n+4}+\delta_{n'n-4})\delta_{m'm}+Cd_{nm}^4\delta_{n'n}(\delta_{m'm+4}+\delta_{m'm-4})$$
$$+2Bb_{nm}^2d_{nm}^2(\delta_{n'n+2}\delta_{m'm+2}+\delta_{n'n+2}\delta_{m'm-2}$$
$$+\delta_{n'n-2}\delta_{m'm+2}+\delta_{n'n-2}\delta_{m'm-2})$$
$$+(-\tfrac{1}{2}a_{nm}^2+\tfrac{1}{2}b_{nm}^2+4Ab_{nm}^4+4Bb_{nm}^2d_{nm}^2)$$
$$\times(\delta_{n'n+2}+\delta_{n'n-2})\delta_{m'm}+(-\tfrac{1}{2}c_{nm}^2+d_{nm}^2+4Cd_{nm}^4+4Bb_{nm}^2d_{nm}^2)$$
$$\times\delta_{n'n}(\delta_{m'm+2}+\delta_{m'm-2})$$
$$+(a_{nm}^2+b_{nm}^2+c_{nm}^2+d_{nm}^2+6Ab_{nm}^4$$
$$+8Bb_{nm}^2d_{nm}^2+6Cd_{nm}^4)\delta_{m'm}\delta_{n'n}.$$

In this Hamiltonian the term proportional to $\delta_{m'm}\delta_{n'n\pm4}\delta_{n'n}$, and $\delta_{m'm\pm2}\delta_{n'n\pm2}$ may not be set equated to

zero consistently with the requirements that A, B, and C are not zero and the implications of the diagonal commutator equations. The terms proportional to $\delta_{n'n\pm2}\delta_{m'm}$ and $\delta_{m'm\pm2}\delta_{n'n}$ may, however, be equated to zero in this approximation. If this is done, the pair of equations

$$-\tfrac{1}{2}a^2+\tfrac{1}{2}b^2+4Ab^4+4Bb^2d^2=0,$$

$$-\tfrac{1}{2}c^2+\tfrac{1}{2}d^2+4Cd^4+4Bb^2d^2=0$$

results. The diagonal commutator matrix elements can be used to eliminate a and c and give the equations

$$32Ab^6+32Bb^4d^2-n^2+4b^4=0,$$

$$32Cd^6+32Bb^2d^4-m^2+4d^4=0.$$

The situation is simplified if the harmonic terms $4b^4$ and $4d^6$ are dropped. If the variables $\beta=n^{-1/3}\,(32A)^{1/6}b$ and $\delta=m^{-1/3}\,(32C)^{1/6}d$ are introduced the equations become

$$\begin{aligned}&\beta^6+\sigma\beta^4\delta^2-1=0,\\&\delta^6+\tau\beta^2\delta^4-1=0,\end{aligned}\tag{1}$$

where $\sigma=BA^{-2/3}C^{-1/3}(m/n)^{2/3}$ and $\tau=BA^{-1/3}C^{-2/3}(n/m)^{2/3}$. The relation $\sigma\tau=B^2/AC$ is obvious.

It is not possible to give an exact and simple solution to this pair of equations. In order to bring out the qualitative features of the energy levels and the coordinate and momentum operator matrix elements, I will use an approximate solution to these equations. Since the exact β and δ would themselves give only an approximate solution to the eigenvalue problem, there are two levels of approximation. In the following paragraphs I deal with β_0 and δ_0, an approximate solution to the equations for β and δ. I can check numerically and determine how good or bad an approximation β_0 and δ_0 are to β and δ, and these figures are reported below.

An approximate solution β_0, δ_0 to these equations is given by

$$\beta_0=\frac{1+\sigma^{1/12}+\tau^{1/6}}{1+\tau^{1/6}+\sigma^{1/3}},\quad \delta_0=\frac{1+\tau^{1/12}+\sigma^{1/6}}{1+\sigma^{1/6}+\sigma^{1/3}}.$$

The approximation $\beta_0\delta_0$ is valid asymptotically in the regions σ and τ separately large or small to lowest nonvanishing order. The equations for β and δ may be solved by Newton's method if $\beta_0\delta_0$ is the first trial solution. The results are too complicated to carry out algebraically. Numerically work indicates that the maximum error $(\beta-\beta_0)/\beta_0$ or $(\delta-\delta_0)/\delta_0$ is about 35% where β and δ are the exact solutions of Eq. (1).

The energy levels can be given in terms of β and δ by

$$E_{nm}=2^{-1/3}\{(An^4)^{1/3}[(\beta^6+4)/4\beta^2]+(Cm^4)^{1/3}[(\delta^6+4)/4\delta^2]\}.$$

The equations satisfied by β and δ are used to reduce the energy to this form. If β_0 and δ_0 are substituted for β and δ, the energy is accurate to better than a factor of 2.

The quantities σ and τ are dimensionless since A, B, and C have the same dimensions. Only the ratio of (m/n) occurs in σ and τ. The energy depends on the quantities $(An^4)^{1/3}$ and $(Cm^4)^{1/3}$ multiplied by dimensionless functions; functions of σ and τ. I believe this is a feature of an exact solution to the eigenvalue problem.

The expression for E given above is only valid to lowest order in each of the four regions,

I: σ and τ small,

II: σ small τ large,

III: σ large τ small,

IV: σ and τ large.

By using Newton's method more accurate approximations of β and δ can be achieved in each of these asymptotic regions and the energy can be calculated to greater accuracy. This gives the following results for the energy:

$$\text{I: } E_{nm}=1.25[(\tfrac{1}{2}An^4)^{1/3}+(\tfrac{1}{2}Cm^4)^{1/3}]+\tfrac{1}{3}(2AC)^{-1/3}B(n^2m^2)^{1/3},$$

$$\text{II: } E_{nm}=1.25(\tfrac{1}{2}An^4)^{1/3}+\tfrac{7}{6}[(B^3/A)^{1/2}nm^3]^{1/3},$$

$$\text{III: } E_{nm}=1.25(\tfrac{1}{2}Cm^4)^{1/3}+\tfrac{7}{6}[(B^3/C)^{1/2}n^3m]^{1/3},$$

$$\text{IV: } E_{nm}=2(\tfrac{1}{2}Bm^2n^2)^{1/3}.$$

An examination of the off-diagonal portions of the commutator and the Hamiltonian indicates they are small compared to the diagonal portions so that the approximation furnished by the exact solution of (1) is a reasonable one.

IV. CONCLUSIONS

It is possible either by exactly solving the cubic equations (1) for β and δ or approximately by using the approximate solutions β_0 and δ_0 to give momentum and position operators that simultaneously satisfy the diagonal commutator equations and the first off-diagonal Hamiltonian equation. This is the first step in a step-by-step procedure to solve the two-coupled oscillator problem.

In higher orders of the approximation additional terms more remote from the diagonal would be added to x and p, and further commutator and Hamiltonian equations would be solved. I do not yet know the order in which this is to be done. My experience with the single oscillator suggests that the dependence on $m+m'$ and $n+n'$ should be the same for the additional terms, but they should rapidly decrease with $|m'-m|$ and $|n'-n'|$ leading to the convergence of the approximation.

APPENDIX

The operators p_1, p_2, x_1, and x_2 are given in III. The calculated values of the six commutators of these operators are listed below:

$$[p_1x_1]_{n'm'nm}=i\delta_{m'm}[(a_{n+1\,m}b_{nm}-a_{nm}b_{n+1\,m})\delta_{n'-2\,n}+2(a_{n-1\,m}b_{n-1\,m}-a_{nm}b_{nm})\delta_{n'n}+(a_{n-1\,m}b_{n-2\,m}-a_{n-2\,m}b_{n-1\,m})\delta_{n'+2\,m}],$$

$$[p_2,x_2]_{n'm'nm}=i\delta_{n'n}[(c_{n\,m+1}d_{nm}-c_{nm}d_{n\,m+1})\delta_{m'-2\,m}+2(c_{n\,m-1}d_{n\,m-1}d_{n\,m-1}-c_{nm}d_{nm})\delta_{m'm}+(c_{n\,m-1}d_{n\,m-2}-c_{n\,m-2}d_{n\,m-1})\delta_{m'+2\,m}],$$

$$[p_1,p_2]_{n'm'nm}=(a_{nm}c_{n+1\,m}-a_{n\,m+1}c_{nm})\delta_{m'm+1}\delta_{n'n+1}+(a_{n\,m-1}c_{n\,m-1}-a_{n\,m}c_{n+1\,m-1})$$
$$\times\delta_{m'm-1}\delta_{n'-1\,n}+(a_{n-1\,m+1}c_{nm}-a_{n-1\,m}c_{n-1\,m})\delta_{m'm+1}\delta_{n'+1\,n}+(a_{n-1\,m}c_{n-1\,m-1}-a_{n-1\,m-1}c_{n\,m-1})\delta_{m'm-1}\delta_{n'n-1},$$

$$[p_1,x_2]_{n'm'nm}=i(a_{n\,m+1}d_{nm}-a_{nm}d_{n+1\,m})\delta_{m'm+1}\delta_{n'n+1}+i(a_{n\,m-1}d_{nm-1}-a_{nm}d_{n+1\,m-1})$$
$$\times\delta_{m'm-1}\delta_{n'n+1}+i(a_{n-1\,m}d_{n-1\,m}-a_{n-1\,m+1}d_{nm})\delta_{m'm+1}\delta_{n'n-1}+i(a_{n-1\,m}d_{n-1\,m-1}-a_{n-1\,m-1}d_{n\,m-1})\delta_{m'm-1}\delta_{n'n-1},$$

$$[p_2,x_1]_{n'm'nm}=i(b_{nm}c_{n+1\,m}-b_{n\,m+1}c_{nm})\delta_{m'm+1}\delta_{n'n+1}+i(b_{n\,m-1}c_{n\,m-1}-b_{nm}c_{n+1\,m-1})$$
$$\times\delta_{m'm-1}\delta_{n'n+1}+i(b_{n-1\,m}c_{n-1\,m}-b_{n-1\,m+1}c_{nm})\delta_{m'm+1}\delta_{n'n-1}+i(b_{n-1\,m-1}c_{n\,m-1}-b_{n-1\,m}c_{n-1\,m-1})\delta_{m'm-1}\delta_{n'n-1},$$

$$[x_1,x_2]_{n'm'nm}=(b_{n\,m+1}d_{nm}-b_{nm}d_{n+1\,m})\delta_{m'm+1}\delta_{n'n+1}+(b_{n\,m-1}d_{n\,m-1}-b_{nm}d_{n+1\,m-1})$$
$$\times\delta_{m'm-1}\delta_{n'n+1}+(b_{n-1\,m+1}d_{nm}-b_{n-1\,m}d_{n-1\,m})\delta_{m'm+1}\delta_{n'n-1}+(b_{n-1\,m-1}d_{n\,m-1}-b_{n-1\,m}d_{n-1\,m-1})\delta_{m'm-1}\delta_{n'n-1},$$

$$H_{n'm'nm}=A(b_{n+3\,m}b_{n+2\,m}b_{n+1\,m}b_{nm}\delta_{n'n+4}\delta_{m'm}+b_{n-4\,m}b_{n-3\,m}b_{n-2\,m}b_{n-1\,m}\delta_{n'+4\,n}\delta_{m'm})$$
$$+C(d_{n\,m+3}d_{n\,m+2}d_{n\,m+1}d_{nm}\delta_{n'n}\delta_{m'm+4}+d_{n\,m-4}d_{n\,m-3}d_{n\,m-2}d_{n\,m-1}\delta_{n'n}\delta_{m'm-4})$$
$$+2B(b_{n+1\,m+2}b_{n\,m+2}d_{n\,m+1}d_{nm}\delta_{n'n+2}\delta_{m'm+2}+b_{n+1\,m-2}b_{n\,m-2}d_{n\,m-2}d_{n\,m-1}\delta_{n'n+2}\delta_{m'm-2}$$
$$+b_{n-2\,m+2}b_{n-1\,m+2}d_{n\,m+1}d_{nm}\delta_{n'n-2}\delta_{m'm+2}+b_{n-2\,m-2}b_{n-1\,m-2}d_{n\,m-2}d_{n\,m-1}\delta_{m'm-2}\delta_{n'n-2})+[-\tfrac{1}{2}a_{n+1\,m}a_{nm}+\tfrac{1}{2}b_{n+1\,m}b_{nm}$$
$$+A(b_{n+1\,m}b_{nm}b_{n-1\,m}^2+b_{n+1\,m}b_{nm}^3+b_{n+1\,m}^3b_{nm}+b_{n+2\,m}^2b_{n+1\,m}b_{nm})+2Bb_{n+1\,m}b_{nm}(d_{n\,m-1}^2+d_{nm}^2)]\delta_{n'n+2}\delta_{m'm}$$
$$+[-\tfrac{1}{2}a_{n-2\,m}a_{n-1\,m}+\tfrac{1}{2}b_{n-2\,m}b_{n-1\,m}+A(b_{n-3\,m}^2b_{n-2\,m}b_{n-1\,m}+b_{n-2\,m}^3b_{n-1\,m}+b_{n-2\,m}b_{n-1\,m}^3+b_{n-2\,m}b_{n-1\,m}b_{nm}^2)$$
$$+2Bb_{n-2\,m}b_{n-1\,m}(d_{n\,m-1}^2+d_{nm}^2)]\delta_{n'n-2}\delta_{m'm}+[-\tfrac{1}{2}c_{n\,m+1}c_{nm}+\tfrac{1}{2}d_{n\,m+1}d_{nm}+C(d_{n\,m+1}d_{nm}d_{n\,m-1}^2+d_{n\,m+1}d_{nm}^3+d_{n\,m+1}^3d_{nm}$$
$$+d_{n\,m+2}^2d_{n\,m+1}d_{nm})+2B(b_{n-1\,m+2}^2+b_{n\,m+2}^2)d_{n\,m+1}d_{nm}]\delta_{n'n}\delta_{m'm+2}+[-\tfrac{1}{2}c_{n\,m-2}c_{n\,m-1}+\tfrac{1}{2}d_{n\,m-2}d_{n\,m-1}+C(d_{n\,m-3}^2d_{n\,m-2}d_{n\,m-1}$$
$$+d_{n\,m-2}^3d_{n\,m-1}+d_{n\,m-2}d_{n\,m-1}^3+d_{n\,m-2}d_{n\,m-1}d_{nm}^2)+2B(b_{n-1\,m-2}^2+b_{n\,m-2}^2)d_{n\,m-2}d_{n\,m-1}]\delta_{n'n}\delta_{m'm-2}+[\tfrac{1}{2}(a_{n-1\,m}^2+a_{nm}^2)$$
$$+\tfrac{1}{2}(c_{n\,m-1}^2+c_{nm}^2)+\tfrac{1}{2}(b_{n-1\,m}^2+b_{nm}^2)+\tfrac{1}{2}(d_{n\,m-1}^2+d_{nm}^2)+A(b_{n-2\,m}^2b_{n-1\,m}^2+b_{n-1\,m}^4+2b_{n-1\,m}^2b_{nm}^2+b_{nm}^4+b_{nm}^2b_{n+1\,m}^2)$$
$$+2B(b_{n-1\,m}^2+b_{nm}^2)(d_{n\,m-1}^2+d_{nm}^2)+C(d_{n\,m-1}^2d_{n\,m-2}^2+d_{n\,m-1}^4+2d_{n\,m-1}^2d_{nm}^2+d_{nm}^4+d_{nm}^2d_{n\,m+1}^2)]\delta_{n'n}\delta_{m'm}.$$

*Work supported in part by the United Stated Atomic Energy Commission.

[1]F. Halpern and T.W. Yonkman, "Matrix mechanics approach to a nonlinear oscillator," submitted to J. Math. Phys.

[2]A. Z. Capri, Lett. Nuovo Cimento 3, 351 (1972); Francis R. Halpern, J. Math. Phys. **14**, 219 (1973).

Differential inequalities and stochastic functional differential equations

G. S. Ladde

Department of Mathematics, The State University of New York at Potsdam, Potsdam, New York 13676
(Received 27 November 1973)

Consider the system of stochastic functional differential equations

$$dx = f(t,x_t)dt + \sigma(t,x_t)dz(t),\ x_{t_0}=\phi_0, \tag{S}$$

where σ is a $n\times m$ matrix, column vectors of σ, f are continuous, and $z(t)$ is a normalized m-vector Wiener process with

$$E[(z(t)-z(s))\cdot(z(t)-z(s))^T]=I|t-s|.$$

By developing a comparison principle, sufficient conditions are given for stability and boundedness in the mean of solutions of (S). The main technique here is the theory of functional differential inequalities and Lyapunov-like functions.

1. INTRODUCTION

Stochastic differential systems provide a mathematical model for sophisticated dynamical systems in physical, biological, medical and social sciences. In many circumstances, the future state of a system depends, not only on the present state but also on its past history. Stochastic functional differential equations give a mathematical formulation for such systems in which the stochastic increment (in the sense of Itô) of the system may depend on the influence of its hereditary effects.

The problem of existence and uniqueness of stationary solutions of functional differential equations has been investigated by Itô and Nisio.[1] In a recent paper Kolmanovskii[2] has studied the problem of stability in the mean. In that paper, an attempt was made to formulate the asymptotic stability criteria by employing Lyapunov functionals.

The notion of Lyapunov function, together with the theory of functional differential inequalities provide a very general comparison principle by means of which a number of qualitative properties of solutions of functional differential equations may be studied in a unified way. For more details see the book of Lakshmikantham and Leela.[3] It is natural to expect such an extension to stochastic functional differential systems.

In this paper, we wish to extend this comparison principle to stochastic functional differential systems. In Sec. 2, we define various notions of stability and boundedness in the mean. In Sec. 3, we develop general comparison principle based on functional differential and integral inequalities. In Sec. 4, we give sufficient conditions for stability and boundedness in the mean. These results include some of the results of Ladde, Lakshmikantham, and Liu.[4] At the end, we provide examples to illustrate the applicability of our results.

2. NOTATIONS AND DEFINITIONS

Let R^n denote the n-dimensional Euclidean space with any convenient norm $\|\cdot\|$. We also denote by the same symbol $\|\cdot\|$ the norm of a matrix. R^+ and R stand for the nonnegative real and real line respectively. Let $(\Omega,\mathcal{J},P)$ be a complete probability space. By $E[x/\mathcal{K}]$ we shall mean the conditional mean of x, where $\mathcal{K}\subset\mathcal{J}$ is a sub σ-algebra of $\mathcal{J}$. Given any $\tau>0$, let $C^n=C[[-\tau,0],R^n]$ denote the space of continuous functions with domain $[-\tau, 0]$ and range in R^n. For $\phi\in C^n$, we define $\|\phi\|_0 = \sup_{-\tau\le s\le 0}\|\phi(s)\|$.

Let $(\cdot)^T$ denote the transpose of a vector or a matrix. Suppose $x\in C[[-\tau,\infty),R^n]$. For any $t\ge 0$, we shall let x_t denote a translation of the restriction of x to the interval $[t-\tau,t]$; more specifically, x_t is an element of C^n defined by $x_t(s)=x(t+s)$, $-\tau\le s\le 0$. Let $S[C^n]$ and $S[C^+]$ denote the system of all C^n and C^+ valued random variables, where $C^+=C[[-\tau,0],R^+]$.

Consider the system of stochastic functional differential equations of the type

$$dx=f(t,x_t)dt+\sigma(t,x_t)dz(t),\quad x_{t_0}=\phi_0, \tag{2.1}$$

where x, $f\in R^n$, $\sigma(t,\cdot)$ is a $n\times m$ matrix, and $z(t)$ is a normalized m-vector Wiener process with

$$E[(z(t)-z(s))\cdot(z(t)-z(s))^T]=I|t-s|,$$

where I is an identity matrix.

We assume that the functions f, σ, and ϕ_0 satisfy the following assumptions:

(a_1) the m column vectors of $\sigma(t,\phi)$ and $f(t,\phi)$ belong to $C[R^+\times C^n,R^n]$;

(a_2) for all (t,ϕ), $(t,\psi)\in R^+\times C^n$,

$$\|f(t,\phi)-f(t,\psi)\|\le\int_{-\tau}^{0}\|\phi(s)-\psi(s)\|dK_1(s)$$

and

$$\|\sigma(t,\phi)-\sigma(t,\psi)\|\le\int_{-\tau}^{0}\|\phi(s)-\psi(s)\|dK_2(s)$$

where dK_1 and dK_2 are bounded measures on $[-\tau,0]$;

(a_3) $\phi_0(s)$, $s\in[-\tau,0]$ is a sample continuous stochastic process, i.e., $\phi_0\in S[C^n]$ independent of $z(t)$, for $t\in R^+$ and satisfies the relation

$$\sup_{-\tau\le s\le 0}E[\|\phi_0(s)\|^4]\le C,$$

for some constant $C>0$.

Under these assumptions, it is known[1] that the solution $x(t)=x(t_0,\phi_0)(t)$ of (2.1) is

(I) sample continuous on $[t_0-\tau,\infty)$,

(II) strictly stationary process, i.e., $x(t)$ and $z(t)$ are strictly correlated,

(III) $E[\|x(t)\|^4] \leq C_1 \sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|^4] e^{C_2 t}$, for some C_1, $C_2 > 0$.

We shall now formulate the definitions of stability and boundedness in the mean of the trivial solution of (2.1).

Definition 2.1: The trivial solution of (2.1) is said to be:

(i) *equi-stable in the mean*, if for each $\epsilon > 0$, $t_0 \in R^+$, there exists a positive function $\delta = \delta(t_0, \epsilon)$ that is continuous in t_0 for each $\epsilon > 0$ such that the inequality

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \delta$$

implies

$$E[\|x(t_0, \phi_0)(t)\|] < \epsilon, \text{ for } t \geq t_0;$$

(ii) *uniformly stable in the mean*, if the δ in (i) is independent of t_0;

(iii) *quasi-equi-asymptotically stable in the mean*, if for each $\epsilon > 0$, $t_0 \in R^+$, there exist $\delta_0 = \delta(t_0) > 0$ and $T = T(t_0, \epsilon) > 0$ such that for $t \geq t_0 + T$ and

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \delta_0$$

implies

$$E[\|x(t_0, \phi_0)(t)\|] < \epsilon;$$

(iv) *quasi-uniformly asymptotically stable in the mean*, if δ_0 and T in (iii) are independent of t_0;

(v) *equi-asymptotically stable in the mean*, if (i) and (iii) hold simultaneously;

(vi) *uniformly asymptotically stable in the mean*, if (ii) and (iv) hold together.

Definition 2.2: The stochastic system (2.1) is said to be:

(i) *equi-bounded in the mean*, if each $\alpha \geq 0$, $t_0 \in R^+$, there exists a positive function $\beta = \beta(t_0, \alpha)$ that is continuous in t_0 for each α such that

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \alpha$$

implies

$$E[\|x(t_0, \phi_0)(t)\|] < \beta, \quad t \geq t_0;$$

(ii) *uniformly bounded in the mean*, if the α in (i) is independent of t_0;

(iii) *quasi-equi-ultimately bounded in the mean*, if given $\alpha \geq 0$, $t_0 \in R^+$, there exist numbers N and $T = T(t_0, \alpha)$ such that

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \alpha$$

implies

$$E[\|x(t_0, \phi_0)(t)\|] < N, \quad t \geq t_0 + T;$$

(iv) *quasi-uniformly-ultimately bounded in the mean*, if the number T in (iii) is independent of t_0;

(v) *equi-ultimately bounded in the mean*, if (i) and (iii) hold at the same time;

(vi) *uniformly ultimately bounded in the mean*, if (ii) and (iv) hold simultaneously.

Consider now the comparison functional differential equation

$$u' = g(t, u) + G(t, u_t), \quad u_{t_0} = \sigma_0 \tag{2.2}$$

where $g \in C[R^+ \times R^+, R]$, $G \in C[R^+ \times C^+, R]$ and $G(t, \sigma)$ is nondecreasing in σ for each $t \in R^+$.

Relative to the comparison functional differential equation (2.2), we need the corresponding definitions in our discussion which may be defined analogously. For example, the definition of equi-stability runs as follows: The trivial solution $u \equiv 0$ of (2.2) is said to be equi-stable, if for each $\epsilon > 0$, $t_0 \in R^+$, there exists a positive function $\delta = \delta(t_0, \epsilon)$ that is continuous in t_0 for each $\epsilon > 0$ such that $|\sigma_0|_0 \leq \delta$ implies $u(t_0, \sigma_0)(t) < \epsilon, t \geq t_0$

Definition 2.3: A function $b(r)$ is said to belong to the class K, if $b \in C[R^+, R^+]$, $b(0) = 0$ and $b(r)$ is strictly increasing in r.

Definition 2.4: A function $a(t, r)$ is said to belong to the class CK, if $a \in C[[-\tau, \infty) \times R^+, R^+]$, $a(t, 0) \equiv 0$ and $a(t, r)$ is concave and increasing in r for each $t \in R^+$.

3. COMPARISON RESULTS

In this section, we wish to prove some comparison theorems for stochastic functional differential system. This is achieved by employing the notion of Lyapunov function and the theory of functional differential and integral inequalities analogous to the deterministic case.[3] These results play an important role to study the qualitative behavior of (2.1). Note that these results are extension of corresponding result.[3]

Let the function $V \in C[[-\tau, \infty) \times R^n, R^+]$, V_t, V_x, V_{xx} exist and are continuous for $(t, x) \in R^+ \times R^n$, the calculus introduced by Itô[5] shows that

$$dV(t, x, \phi) = LV(t, x, \phi)dt + \frac{\partial V}{\partial x} \cdot \sigma(t, \phi) dz(t), \tag{3.1}$$

where

$$LV(t, x, \phi) = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} \cdot f(t, \phi) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 V}{\partial x_i \partial x_j} b_{ij}(t, \phi), \tag{3.2}$$

$\phi \in C^n$, and $(b_{ij}) = \sigma \cdot \sigma^T$.

Here and after, we shall assume that Eq. (2.2) and the function V satisfy the following hypotheses:

(H_1) $g \in C[R^+ \times R^+, R]$, $g(t, u)$ is concave and nondecreasing in u for each $t \in R^+$.

(H_2) $G \in C[R^+ \times C^+, R]$, $G(t, \sigma)$ is nondecreasing in σ for each $t \in R^+$ and for any $\sigma \in S[C^+]$, $G(t, \sigma)$ satisfies the relation

$$E[G(t, \sigma)] \leq G(t, E\sigma), \tag{3.3}$$

where $E\sigma = E[\sigma(s)]$, $s \in [-\tau, 0]$.

(H_3) Let $r(t_0, \sigma_0)(t)$ be the maximal solution of the functional differential equation (2.2) existing for $t \geq t_0$, $t_0 \in R^+$.

(H_4) Assume that $g(t, 0) \equiv 0 \equiv G(t, 0)$.

(H_5) $V \in C[[-\tau, \infty) \times R^n, R^+]$, $\partial V/\partial t$, $\partial V/\partial x$, $\partial^2 V/\partial x \partial x$ exist and are continuous for $(t, x) \in R^+ \times R^n$. Furthermore, for $(t, \phi(0), \phi) \in R^+ \times R^n \times C^n$,

$$\frac{\partial V}{\partial x}\cdot\sigma(t,\phi)\le K(t)[1+\int_{-\tau}^{0}\|\phi(s)\|_{ds}^{u}], \tag{3.4}$$

and

$$LV(t,\phi(0),\phi)\le g(t,V(t,\phi(0)))+G(t,V_t), \tag{3.5}$$

where $K\in C[R^+,R^m]$, $1\le u\le 4$, and L is the operator as defined in (3.2).

(H_6) Assume that the hypothesis (H_5) holds except that the inequality (3.5) is strengthened to

$$A(t)LV(t,\phi(0),\phi)+V(t,\phi(0))A'(t)\le g(t,V(t,\phi(0))A(t))+G(t,A_tV_t), \tag{3.6}$$

where $A(t)$ is continuously differentiable positive function for $[-\tau,\infty)$ and $A_tV_t=A(t+s)V(t+s,\phi(s))$ for $s\in[-\tau,0]$.

(H_7) For $(t,x)\in R^+\times R^n$,

$$b(\|x\|)\le V(t,x)\le a(t,\|x\|),$$

where $b\in\mathcal{K}$, b is convex, and $a\in C\mathcal{K}$.

(H_8) For $(t,x)\in R^+\times R^n$,

$$b(\|x\|)\le V(t,x)\le a(t,\|x\|),$$

where $a\in C\mathcal{K}$, $b\in\mathcal{K}$, b is convex, and $b(u)\to\infty$ as $u\to\infty$.

We shall state and prove the following main comparison theorem.

Theorem 3.1: Let the hypotheses (H_1), (H_2), (H_3), and (H_5) be satisfied. Assume that, for the stationary solution process $x(t)=x(t_0,\phi_0)(t)$ of (2.1), $E[V(t,x(t)$ exists and $E[V(t_0+s,\phi_0(s))]\le\sigma_0(s)$, $s\in[-\tau,0]$. Then, we have

$$E[V(t,x(t_0,\phi_0)(t))]\le r(t_0,\sigma_0)(t),\quad t\ge t_0. \tag{3.7}$$

Proof: Let $x(t)=x(t_0,\phi_0)(t)$ be any stationary solution process of (2.1) with initial data (t_0,ϕ_0). By the hypothesis (H_5) and the calculus introduced by Itô,[5] we have

$$\begin{aligned}dV(t,x(t),x_t)&=\frac{\partial V}{\partial t}(t,x(t))dt+\frac{\partial V}{\partial x}(t,x(t))\cdot f(t,x_t)dt\\&\quad+\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2V}{\partial x_i\partial x_j}(t,x(t))b_{ij}(t,x_t)dt\\&\quad+\frac{\partial V}{\partial x}(t,x(t))\cdot\sigma(t,x_t)dz(t)\\&=LV(t,x(t),x_t)dt+\frac{\partial V}{\partial x}(t,x(t))\cdot\sigma(t,x_t)dz(t).\end{aligned}$$

This together with the hypotheses (H_1), (H_2), (H_5) and the existence of $E[V(t,x(t))]$ yields the inequality

$$E[V(t,x(t))]-E[V(t_0,\phi_0(s))]\le\int_{t_0}^{t}[g(s,EV(s,x(s)))+G(s,EV_s)]ds. \tag{3.8}$$

Set

$$m(t)=E[V(t,x(t))],\quad m(t_0)=E[V(t_0,\phi_0(t_0))],$$

then in view of (3.8), we have

$$m(t)\le m(t_0)+\int_{t_0}^{t}[g(s,m(s))+G(s,m_s)]ds. \tag{3.8'}$$

Define

$$u(t)=\begin{cases}m(t_0)+\int_{t_0}^{t}[g(s,m(s)),+G(s,m_s)]ds, & \text{for } t\ge t_0,\\ m(t), & \text{for } t\in[t_0-\tau,t_0],\end{cases}$$

so that

$$m(t)\le u(t),\quad \text{for } t\in[t_0-\tau,\infty), \tag{3.9}$$

and

$$u'(t)=g(t,m(t))+G(t,m_t),\quad \text{for } t\ge t_0. \tag{3.10}$$

From (3.9), (3.10), and nondecreasing property of g and G in u and σ respectively, we get

$$u'(t)\le g(t,u(t))+G(t,u_t),\quad \text{for } t\ge t_0, \tag{3.11}$$

and

$$m_{t_0}=u_{t_0}\le\sigma_0. \tag{3.12}$$

From an application of Theorem 6.9.4 in Ref. 3, we deduce that

$$u(t)\le r(t_0,\sigma_0)(t),\quad t\ge t_0.$$

The assertion (3.7) is now immediate in view of (3.9) and the definition of $m(t)$. The proof is complete.

The following variant of Theorem 3.1 is often more useful in applications.

Theorem 3.2: Let the hypotheses of Theorem 3.1 hold except that (H_5) is replaced by (H_6). Then

$$A(t_0+s)E[V(t_0+s,\phi_0(s))]\le\sigma_0(s)$$

implies

$$E[V(t,x(t))]\le R(t_0,\psi_0)(t),\quad t\ge t_0, \tag{3.13}$$

where $R(t_0,\psi_0)(t)$ is the maximal solution of the differential equation

$$v'=[-A'v+g(t,vA(t))+G(t,A_tv_t)]/A(t),\quad v_{t_0}=\psi_0, \tag{3.14}$$

existing for $t\ge t_0$.

Proof: Setting

$$W(t,\phi(0))=V(t,\phi(0))A(t),\quad t\in[-\tau,\infty),$$

we see, because of (3.6), that

$$\begin{aligned}LW(t,\phi(0),\phi)&=A(t)LV(t,\phi(0),\phi)+A'(t)V(t,\phi(0))\\&\le g(t,W(t,\phi(0)))+G(t,W_t).\end{aligned}$$

This shows that $W(t,x)$ satisfies all the hypotheses of Theorem 3.1 and, as a consequence, we have

$$E[W(t,x(t))]\le r(t_0,\sigma_0)(t),\quad t\ge t_0, \tag{3.15}$$

provided that

$$W(t_0+s,\phi_0(s))\le\sigma_0(s).$$

Here $r(t_0,\sigma_0)(t)$ is the maximal solution of (2.2). It is easy to verify $r(t_0,\sigma_0)(t)=A(t)R(t_0,\psi_0)(t)$ with $\psi_0(s)A(t_0+s)=\sigma_0(s)$. This implies because of (3.15) and the definition of $W(t,x)$, the desired inequality (3.13).

Remark 3.1: Theorem 3.1 is analogous to Theorem 8.1.4 in Ref. 3 for determinstic case. However, it is required that u, u_t are "separated" on the right-hand side of 8.1.16 in Ref. 3, i.e.,

$$g(t,u,u_t)=g(t,u)+G(t,u_t).$$

Furthermore, our results are obtained by employing integral inequalities, so we require monotonicity in u. However, if (2.1) is a system of diffusion equations, the differential inequalities can be employed to discuss such results. For more details see Ref. 4.

Remark 3.2: The drawback of Theorem 3.1 is the assumption that $E[V(t,x(t))]$ exists for each $t\in R^+$. Under certain conditions, one could show that this assumption holds. For example, let $V(t,x)\leq a(t,\|x\|)$, where $a \in C[R^+\times R^+,R^+]$ and $a(t,u)$ is concave in u for fixed $t\in R^+$. Then we would have

$$0\leq E[V(t,x(t))]\leq a(t,E[\|x(t)\|]),$$

in view of the property (III) of the solution process $x(t)$.

Remark 3.3: Observe that the nondecreasing nature of $g(t,u)$ in u can be dropped, if the inequalities (3.3), (3.4), and (3.5) are equalities. However, we do require nondecreasing nature of $G(t,\sigma)$ in σ, in order to insure the existence of the maximal solution for (2.2). See Ref. 3 for more details.

Remark 3.4: The restriction $E[G(t,\sigma)]\leq G(t,E\sigma)$ for $\sigma\in S[C^+]$ on $G(t,u_t)$ is natural. However, the class of functionals $G(t,u_t)$ having such a property is nonempty. For example,

$$G(t,\sigma)=\lambda(t)\int_{-\tau}^{0}\sigma(s)dK(s),\quad \text{for } \sigma\in S[C^+],$$

where dK is a bounded measure on $[-\tau,0]$. One can easily see that

$$E[G(t),\sigma)]=\lambda(t)\int_{-\tau}^{0}E[\sigma(s)]dK(s),$$

whenever $E[\sigma(s)]$ exists for $s\in[-\tau,0]$.

4. STABILITY AND BOUNDEDNESS IN MEAN

In this section, by employing the comparison theorems developed in the preceding section, we shall present various results giving sufficient conditions for stability and boundedness in the mean.

Theorem 4.1: Assume that the hypotheses (H_1), (H_2), (H_3), (H_4), (H_5), and (H_7) hold. Furthermore, the system (2.1) possesses the trivial solution. Then:

(i) equi-stability of the trivial solution of (2.2) implies equi-stability in the mean of the trivial solution of (2.1);

(ii) quasi-equi-asymptotic stability of the trivial solution of (2.2) implies quasi-equi-asymptotic stability in the mean of the trivial solution of (2.1);

(iii) equi-asymptotic stability of the trivial solution of (2.2) implies equi-asymptotic stability in the mean of the trivial solution of (2.1).

Proof: Let $x(t_0,\phi_0)(t)$ be any stationary solution process of (2.1). By the hypothesis (H_7), we have

$$0\leq E[b(\|x(t_0,\phi_0)(t)\|)]\leq E[V(t,x(t_0,\phi_0)(t))]$$
$$\leq a(t,E[\|x(t_0,\phi_0)(t)\|]),$$

which implies that $E[V(t,x(t_0,\phi_0)(t))]$ exists, in view of the Remark 3.2. Hence, by Theorem 3.1, we have the following inequality,

$$E[V(t,x(t_0,\phi_0)(t))]\leq r(t_0,\sigma_0)(t),\quad t\geq t_0, \tag{4.1}$$

whenever

$$E[V(t_0+s,\phi_0(s))]\leq\sigma_0(s),\quad s\in[-\tau,0]. \tag{4.2}$$

Let us first prove statement (i). Let $\epsilon>0$ and $t_0\in R^+$ be given. Assume that the trivial solution $u\equiv 0$ of (2.2) is equi-stable. Then given $b(\epsilon)$, $t_0\in R^+$, there exists a positive function $\delta_1=\delta_1(t_0,\epsilon)$ that is continuous in t_0 for each ϵ such that $|\sigma_0|_0\leq\delta_1$ implies

$$u(t_0,\sigma_0)(t)<b(\epsilon),\quad t\geq t_0, \tag{4.3}$$

where $u(t_0,\sigma_0)(t)$ is any solution of (2.2). We choose

$$\sigma_0(s)=a(t_0+s,E[\|\phi_0(s)\|]).$$

Since $a\in CK$, for fixed $s\in[t_0-\tau,t_0]$, we can find $\delta(t_0+s,\epsilon)=\delta_s>0$ that is continuous in t_0 for each ϵ, such that

$$E[\|\phi_0(s)\|]<\delta_s$$
$$\text{implies } a(t_0+s,E[\|\phi_0(s)\|])\leq\delta_1. \tag{4.4}$$

Our aim is to choose δ which is independent of $s\in[-\tau,0]$. From the continuity of $E[\|\phi_0(s)\|]$ in s and (4.4), we can find η_s, for fixed $s\in[-\tau,0]$ such that

$$E[\|\phi_0(\theta)\|]<\delta_s,\quad \text{for } \theta\in(-\eta_s,\eta_s)\cap[-\tau,0].$$

This is true for each $s\in[-\tau,0]$. Consider the collection of open sets in $[-\tau,0]$ defined by

$$U=\{O_s: O_s=(-\eta_s,\eta_s)\cap[-\tau,0],\ \text{for } s\in[-\tau,0]\}.$$

It is easy to verify that it is an open covering of $[-\tau,0]$ and hence by Heine—Borel theorem, we can extract a finite subcover corresponding to η_{s_1}, $\eta_{s_2},\ldots,\eta_{s_n}$ for some fixed integer n. Take the corresponding numbers δ_{s_1}, $\delta_{s_2},\ldots,\delta_{s_n}$ and set

$$\delta=\min\{\delta_{s_1},\delta_{s_2},\ldots,\delta_{s_n}\}.$$

Then, we have

$$\sup_{-\tau\leq s\leq 0}E[\|\phi_0(s)\|]\leq\delta$$
$$\text{implies } a(t_0+s,E[\|\phi_0(s)\|])\leq\delta_1. \tag{4.5}$$

Now, we claim that if

$$\sup_{-\tau\leq s\leq 0}E[\|\phi_0(s)\|]\leq\delta,$$

then

$$E[\|x(t_0,\phi_0(t))\|]<\epsilon,\quad t\geq t_0.$$

Suppose that this is false. Then there would exist a solution $x(t_0,\phi_0)(t)$ with

$$\sup_{-\tau\leq s\leq 0}E[\|\phi_0(s)\|]\leq\delta \text{ and } t_1>t_0$$

such that

$$E[\|x(t_0,\phi_0)(t_1)\|]=\epsilon. \tag{4.6}$$

This, in view of the hypothesis (H_7), gives

$$b(E[\|x(t_0,\phi_0)(t_1)\|])\leq E[V(t_1,x(t_0,\phi_0)(t_1))]. \tag{4.7}$$

The relations (4.1), (4.3), (4.6), and (4.7) lead us to the contradiction

$$b(\epsilon)\leq E[V(t_1,x(t_0,\phi_0)(t_1))]\leq r(t_0,\sigma_0)(t_1)<b(\epsilon).$$

This proves the conclusion (i).

Now, we shall prove the conclusion (ii). Let $\epsilon > 0$, $t_0 \in R^+$, be given. By quasi-asymptotic stability of (2.2), given $b(\epsilon)$ and $t_0 \in R^+$, there exist numbers $\delta^0(t_0) = \delta^0 > 0$ and $T(t_0, \epsilon) = T > 0$ such that

$$u(t_0, \sigma_0)(t) < b(\epsilon), \quad t \geq t_0 + T, \tag{4.8}$$

whenever $|\sigma_0|_0 \leq \delta^0$. Choosing $\sigma_0(s) = a(t_0 + s, E[\|\phi_0(s)\|])$, we can find, as before a $\delta_0 = \delta_0(t_0) > 0$ such that

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] < \delta_0$$

and $a(t_0 + s, E[\|\phi_0(s)\|]) \leq \delta^0$

hold at the same time. We claim that

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \delta_0$$

implies $E[\|x(t)\|] < \epsilon$, for $t \geq t_0 + T$. As a result, we have, because of (4.1), (4.8), and (H_7),

$$b(E[\|x(t)\|]) \leq E[V(t, x(t))]$$

$$\leq r(t_0, \sigma_0)(t) < b(\epsilon), \quad t \geq t_0 + T,$$

which implies, arguing as before that

$$E[\|x(t)\|] < \epsilon, \quad t \geq t_0 + T,$$

whenever

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \delta_0.$$

This proves (ii).

The proof of (iii) follows from the proof of (i) and (ii). Thus the proof of the theorem is complete.

Theorem 4.2: Let the hypotheses (H_1), (H_2), (H_3), (H_5), and (H_8) be satisfied. Then:

(i) equi-boundedness of solutions of (2.2) implies the equi-boundedness in the mean of solutions of (2.1);

(ii) quasi-equi-ultimately boundedness of solutions of (2.2) implies the quasi-equi-ultimately boundedness in the mean of solutions of (2.1);

(iii) equi-ultimate-boundedness of solutions of (2.2) implies equi-ultimate boundedness in the mean of solutions of (2.1).

Proof: Let $x(t_0, \phi_0)(t)$ be any stationary solution process of (2.1). By following the proof of the Theorem 4.1, we have the inequality (4.1).

Let $\alpha \geq 0$ and $t_0 \in R^+$ be given and let

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \alpha.$$

Define

$$\alpha_1 = \sup_{-\tau \leq s \leq 0} a(t_0 + s, \alpha).$$

Assume that the solutions of (2.2) are equi-bounded. Then, given $\alpha_1 \geq 0$ and $t_0 \in R^+$, there exists a positive function

$$\beta_1(t_0, \alpha) = \beta_1 > 0$$

that is continuous in t_0 for each α such that $|\sigma_0|_0 \leq \alpha_1$ implies

$$u(t_0, \sigma_0)(t) < \beta_1, \quad t \geq t_0, \tag{4.9}$$

where $u(t_0, \sigma_0)(t)$ is any solution of (2.2). Since $b(u) \to \infty$ as $u \to \infty$, we can find a $\beta(t_0, \alpha) = \beta > 0$ satisfying the relation

$$\beta_1 \leq b(\beta). \tag{4.10}$$

Now let $\sigma_0(s) = a(t_0 + s, E[\|\phi_0(s)\|])$ so that $|\sigma_0|_0 \leq \alpha_1$. With the β obtained in (4.10), we claim that the solutions of (2.1) are equi-bounded in the mean. If this is not true, then there exists a solution $x(t_0, \phi_0)(t)$ with

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \alpha$$

and $t_1 > t_0$ such that

$$E[\|x(t_0, \phi_0)(t_1)\|] = \beta. \tag{4.11}$$

This together with (4.1), (4.9), (4.10), and (H_8), we have

$$b(\beta) \leq E[V(t_1, x(t_1))] \leq r(t_0, \sigma_0)(t_1) < b(\beta),$$

which completes the proof of (i).

To prove (ii), let $\alpha \geq 0$ and $t_0 \in R^+$ be given and let

$$\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \alpha.$$

Define

$$\alpha_1 = \sup_{-\tau \leq s \leq 0} a(t_0 + s, \alpha).$$

Assume that the solutions of (2.2) are quasi-equi-ultimately bounded. Then, given $\alpha_1 \geq 0$ and $t_0 \in R^+$, there exist positive numbers N_1 and $T = T(t_0, \alpha)$ such that

$$u(t_0, \sigma_0)(t) < N_1, \quad t \geq t_0 + T, \tag{4.12}$$

whenever $|\sigma_0|_0 \leq \alpha_1$. Since $b(u) \to \infty$ as $u \to \infty$, it is possible to fine $N > 0$ such that

$$N_1 \leq b(N). \tag{4.13}$$

As before, choosing $\sigma_0(s) = a(t_0 + s, E[\|\phi_0(s)\|])$, we can conclude, because of the relations (4.1), (4.12), and (4.13),

$$E[V(t, x(t_0, \phi_0)(t))] < b(N), \quad t \geq t_0 + T.$$

From this and (H_8), it is easy to deduce that $E[\|x(t)\|] < N$ for $t \geq t_0 + T$, whenever $\sup_{-\tau \leq s \leq 0} E[\|\phi_0(s)\|] \leq \alpha$. This proves (ii).

The proof of (iii) follows from the proof of (i) and (ii). Hence the theorem is proved.

In Theorem 4.1, the assumption that the trivial solution of (2.2) is asymptotically stable, in general, may not be valid. In such cases, the following result which is based, on the comparison Theorem 3.2, is useful to discuss the asymptotic stability of (2.1). We state the result in the following.

Theorem 4.3: Assume that the hypotheses of the Theorem 4.1 hold except that (H_5) is replaced by (H_6). Then:

(i) equi-stability of the trivial solution of (3.14) implies equi-stability in the mean of the trivial solutions of 2.1;

(ii) equi-quasi-asymptotic stability of trivial solution of (3.14) implies equi-quasi-asymptotic stability in the mean of the trivial solution of (2.1);

(iii) equi-asymptotic stability of the trivial solution of (3.14) implies equi-asymptotic stability in the mean of the trivial solution of (2.1).

Proof: Let $x(t_0,\phi_0)(t)$ be any solution of (2.1). As before, by the hypotheses (H_7), $E[V(t,x(t_0,\phi_0)(t))]$ exists for $t\geq t_0$ and hence by Theorem 3.2, we have

$$E[V(t,x(t_0,\phi_0(t))]\leq R(t_0\psi_0)(t),\quad t\geq t_0,$$

whenever $E[V(t_0+s,\phi_0(s))]\leq\sigma_0(s)$, where $R(t_0,\psi_0)(t)$ is the maximal solution of (3.14). We now follow an argument similar to the proof of Theorem 4.1 to complete the proof of the theorem.

Also, one could formulate the result corresponding to uniform notions by assuming $V(t,x)$ is decrescent and the corresponding notion of the comparison equation (2.2) or (3.14) is also uniform.

5. EXAMPLES

In this section, we shall give some examples to demonstrate the usefulness of our results.

Example 1: Relative to the system (2.1), assume that

(a) $\phi(0)\circ f(t,\phi)\leq 0$,

$$\phi(0)\circ\sigma(t,\phi)\leq L(t)\left(1+\int_{-\tau}^{0}\|\phi(s)\|^{\mu}ds\right),$$

and

$$\sum_{i,j=1}I_{ij}b_{ij}(t,\phi)\leq\lambda(t)\int_{-\tau}^{0}\|\phi(s)\|^2ds$$

for $(t,\phi(0),\phi)\in R^+\times R^n\times\mathcal{C}^n$,

where I is the identity matrix,

$$b(t,\phi)=\sigma(t,\phi)\circ\sigma^T(t,\phi),$$

$L\in C[R^+,R^m]$, $1\leq\mu\leq 4$, and $\lambda\in L^1[0,\infty)$.

(b) $\phi(0)\circ f(t,\phi)\leq-\alpha\|\phi(0)\|^2$, $\alpha>0$,

$$\phi(0)\sigma(t,\phi)\leq L(t)\left(1+\int_{-\tau}^{0}\|\phi(s)\|^{\mu}ds\right)$$

and

$$\sum_{i,j=1}I_{ij}b_{ij}(t,\phi)\leq\frac{P'(t)}{P(t)}\int_{-\tau}^{0}\exp(\alpha s)\|\phi(s)\|^2ds,$$

for $(t,\phi(0),\phi)\in R^+\times R^n\times\mathcal{C}^n$, where $P(t)$ is a polynomial in t. Then taking $V(t,x)=\|x\|^2$, we see that

$$LV(t,\phi(0),\phi)\leq\lambda(t)\int_{-\tau}^{0}V(s+t,\phi(s))ds,$$

when (a) holds. The comparison equation is

$$u'=\lambda(t)\int_{-\tau}^{0}u(t+s)ds$$

and consequently $u\equiv 0$ is stable. Hence, by Theorem 4.1, it follows that the trivial solution of (2.1) is stable in the mean. If on the other hand, (b) holds, we get

$$LV(t,\phi(0),\phi)\leq-\alpha V(t,\phi(0))+\frac{P'(t)}{P(t)}\int_{-\tau}^{0}\exp(\alpha s)V(t+s,\phi(s))ds$$

so that

$$u'=-\alpha u+\frac{P'(t)}{P(t)}\int_{-\tau}^{0}\exp(\alpha s)u(t+s)ds$$

is the corresponding comparison equation. It is easy to see that $u\equiv 0$ is quasi-asymptotically stable. As a result, Theorem (4.3) gives quasi-asymptotic stability in the mean of the trivial solution of (2.1).

Example 2: Consider the stochastic functional differential equations

$$dx=-f(t)x(t)dt+F(x(t-\tau))dz(t),\tag{5.1}$$

where $\tau\geq 0$, $f\in C[R^+,R^+]$ and $F\in C[R,R]$, F satisfies the Lipschitz condition with Lipschitz constant K and $F(0)=0$. Furthermore, assume that

$$\liminf_{t\to\infty}\left(\frac{1}{t-t_0}\int_{t_0}^{t}f(u)du\right)>\frac{K^2}{2}.\tag{5.2}$$

By taking $V(t,x)=x^2$, $A(t)=\exp[2\int_0^t f(u)du]$, it is easy to see that

$$A(t)LV(t,\phi(0),\phi)+A'(t)V(t,\phi(0))\leq K^2V(t-\tau,x(t-\tau))A(t).\tag{5.3}$$

The comparison equation is $u'=-2f(t)u+K^2u(t-\tau)$. Consequently, $u\equiv 0$ is asymptotically stable, in view of (5.2). Hence, by Theorem 4.3, it follows that the trivial solution of (5.1) is asymptotically stable in the mean.

[1]K. Itô and M. Nisio, J. Math. Kyoto Univ. 4-1, 1 (1964).
[2]V. B. Kolmanovskiĭ, Probl. Predači Inform. 5, 59 (1969).
[3]V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities, Theory and Applications, Vols. I and II* (Academic, New York, 1969).
[4]G. S. Ladde, V. Lakshmikantham, and P. T. Liu, "Differential inequalities and Itô type stochastic differential equations," *Proc. of the Int. Conference on Nonlinear Differential and Functional Equations, Bruxelles et Louvain, Belgium* (Harman, Paris, 1973), pp. 611—40.
[5]I. I. Gikhmen and A. V. Skorohod, *Introduction to the Theory of Random Processes* (Saunders, Philadelphia, 1969).

The evaluation of "Kondo" and other integrals of arbitrary range

Norman Louat

Department of Physics, The University of Connecticut, Storrs, Connecticut 06268
(Received 9 October 1973)

It is pointed out that integrals over arbitrary ranges and indefinite integrals may often be obtained very simply by the methods of contour integration.

By resorting to Fourier transforms Glasser[1] has evaluated integrals of the form

$$I_{nm}=\int_{-D}^{D}\frac{\omega^n}{(\omega^2+\Delta^2)^m}\,\frac{d\omega}{(e^{B\omega}+1)} \tag{1}$$

where n and m are integers. The difficulty overcome by Glasser lies in the fact that D is arbitrary. We shall show here that such integrals may be evaluated through contour integration as a matter of routine. In essence this procedure is not new but seems to have been repeatedly forgotten and rediscovered. The first reference of which the author is aware is in Whittaker and Watson[2] in the form of a problem for the student. Later discussions have been couched in complex terms. Accordingly, we shall here present the matter stripped to its essentials and then point out some rather obvious generalizations which do not appear to have been noted.

Given the integral

$$I=\int_a^b f(x)dx \tag{2}$$

we consider the contour integral

$$I_c=\int f(z)\ln[(z-b)/(z-a)]dz \tag{3}$$

with a contour which consists of the two lines joining the branch points at $z=b$, $z=a$ and the circle at infinity.

Now provided $f(z)$ is a single-valued function in the interval of integration, the integrand of Eq. (3) takes the values

$$[\ln(x-a)/(x-b)+\pi i]f(x)$$

and

$$[\ln(x-a)/(x-b)-\pi i]f(x)$$

on the upper and lower lines a-b, respectively. Accordingly, provided $f(z)$ is single valued everywhere and without poles in the interval of integration, we have

$$\begin{aligned}I_c&=2\pi i\int_a^b f(x)dx+\int_\Gamma f(z)\ln[(z-b)/z-a]\,dz\\&=2\pi i\times\text{sum of residues of } f(z)\ln[z-b)/(z-a)]\end{aligned} \tag{4}$$

where the second integral represents the contribution from the great circle Γ. It may be noted that: (1) This procedure may be generalized immediately to cover integration over any number of segments. Furthermore, the weight accorded to each interval need not be the same. This is clear from a consideration of the function

$$\ln\left[\left(\frac{z-a_1}{z-b_1}\right)^{\alpha_1}\left(\frac{z-a_2}{z-b_2}\right)^{\alpha_2}\cdots\left(\frac{z-a_n}{z-b_n}\right)^{\alpha_n}\right]. \tag{5}$$

(2) The method also lends itself to the evaluation of indefinite integrals. Thus, e.g., the use of the function $\ln[(z-b)/(z-a)]$ with the function to be integrated $f(x)$ leads to a result

$$I=F(b)-F(a) \tag{6}$$

where

$$\frac{dF(x)}{dx}=f(x)\,. \tag{7}$$

To illustrate the method we shall now evaluate the generic integral considered by Glasser:

$$I=\int_{-D}^{D}\frac{\omega}{(\omega^2+\Delta^2)}\,\frac{d\omega}{(e^{B\omega}+1)}\,. \tag{8}$$

We consider then

$$I_c=\oint\frac{z}{(z^2+\Delta^2)}\,\frac{1}{(e^{Bz}+1)}\,\ln\frac{z-D}{z+D}\,dz\,.$$

For this function there are simple poles at $z=\pm i\Delta$ and at $z=\pm i(k\pi/B)$ provided $B\Delta\neq k\pi$ where k is any positive odd integer.

It is clear from inspection that the integral around Γ vanishes. The evaluation of the residues is straightforward and we obtain

$$I=\tan^{-1}\frac{D}{\Delta}\tan\frac{B\Delta}{2}+4\pi\sum_{k\ \text{odd}}\frac{k}{B^2\Delta^2-k^2\pi^2}\tan^{-1}\frac{\Delta B}{k\pi}$$

which is the result given by Glasser!

[1] M. L. Glasser, J. Math. Phys. **10**, 1105 (1969).
[2] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., Cambridge, 1902), p. 122.

The unitarity equation for scattering in the absence of spherical symmetry

Michael Tortorella

Department of Mathematics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211
(Received 26 December 1973)

We consider the problem of determining the scattering amplitude from the differential cross section at a fixed energy by using the unitarity equation when the scattering potential does not have spherical symmetry. We indicate some of the problems peculiar to this case. We prove two existence and uniqueness theorems. We give an example of nonuniqueness.

1. INTRODUCTION

The purpose of this article is to give an account of some of the existence and uniqueness questions which arise in connection with the problem of determining the scattering amplitude from the differential cross section at a fixed energy by using the unitarity condition for the scattering of scalar waves in the absence of spherical symmetry. We take the differential cross section as given with infinite accuracy. A forthcoming article will discuss the interesting question of how experimental uncertainties affect the construction of solutions of the unitarity equation, both in this case and in the case in which there is spherical symmetry.

The principal results in this article can be summarized as follows. We first show that the additional necessary condition imposed by the absence of spherical symmetry leads us naturally to consideration of a set of functions which is nowhere dense in the space $C_{\mathbb{C}}(S\times S)$, the Banach space of continuous complex-valued functions defined on the compact set $S\times S$, where S is the unit sphere in $\mathbb{R}^3$, with the usual maximum norm. Next we prove two uniqueness theorems for the unitarity equation in this case. One of these is quite elementary, and has been known to the author for some time. The other is a generalization of a theorem given by Atkinson, Johnson, and Warnock.[1] We also point out a minor error in their proof, and it is easily corrected, as is seen below. Finally, we give an example to show that at least one of the hypotheses of this last-mentioned theorem cannot be dispensed with.

2. A FEATURE PECULIAR TO THIS CASE

We shall use the unitarity equation in the form

$$4\pi \mathrm{Im} f(\hat{n}_1,\hat{n}_2)=\int_S f(\hat{n}_1,\hat{n})\overline{f(\hat{n}_2,\hat{n})}\,d\Omega(\hat{n}), \tag{2.1}$$

where f is proportional to the scattering amplitude, and $\hat{n}_1$, $\hat{n}_2$, $\hat{n}$ are unit vectors in $\mathbb{R}^3$ [see Ref. 2, expression (10), or Ref. 3, expression (1)]. We shall suppose that there is no spherical symmetry, but that the scattering potential has inversion symmetry, so that we have $f(\hat{n}_2,\hat{n}_1)=f(\hat{n}_1,\hat{n}_2)$ and (2.1) is valid (Ref. 2, part I). As usual, we obtain the two equations

$$4\pi \mathrm{Im} f(\hat{n}_1,\hat{n}_2)=\int_S[\mathrm{Re} f(\hat{n}_1,\hat{n})\,\mathrm{Re} f(\hat{n}_2,\hat{n}) + \mathrm{Im} f(\hat{n}_1,\hat{n})\,\mathrm{Im} f(\hat{n}_2,\hat{n})]\,d\Omega, \tag{2.2a}$$

$$0=\int_S[\mathrm{Re} f(\hat{n}_1,\hat{n})\,\mathrm{Im} f(\hat{n}_2,\hat{n})-\mathrm{Re} f(\hat{n}_2,\hat{n})\,\mathrm{Im} f(\hat{n}_1,\hat{n})]\,d\Omega. \tag{2.2b}$$

It is well known that if the potential is spherically symmetric, (2.2b) is identically satisfied. However, if the potential possesses only inversion symmetry, (2.2b) is not, in general, identically satisfied [here the two sentences following (2′) of Ref. 4 are in error]. In fact, we have the following interesting observation:

Proposition 2.1: Let $\mathcal{J}=\{x\in C_{\mathbb{C}}(S\times S)$: $\int_S x(\hat{n}_1,\hat{n})\overline{x(\hat{n}_2,\hat{n})}\,d\Omega$ is real for every $\hat{n}_1,\hat{n}_2\in S\}$. Then $\mathcal{J}$ is nowhere dense in $C_{\mathbb{C}}(S\times S)$.

Proof: Let $\mathcal{A}=\{x\in C_{\mathbb{C}}(S\times S)$: $\mathrm{Im}\int_S x(\hat{n}_1,\hat{n})\overline{x(\hat{n}_2,\hat{n})}\,d\Omega>0$, for every $\hat{n}_1,\hat{n}_2\in S\}$. Then the boundary of $\mathcal{A}$, $\partial\mathcal{A}=\bar{\mathcal{A}}\cap\overline{\mathcal{A}^C}$, is nowhere dense in $C_{\mathbb{C}}(S\times S)$, and $\mathcal{J}$ is a closed subset of $\partial\mathcal{A}$, so that $\mathcal{J}$ is itself nowhere dense in $C_{\mathbb{C}}(S\times S)$.

The following not very surprising corollary follows immediately.

Corollary 2.2. The set of solutions of the unitarity equation (2.1) is nowhere dense in $C_{\mathbb{C}}(S\times S)$.

Proposition 2.1 shows that if the potential has only inversion symmetry, the set of solutions of (2.2b) does not occupy very much space in $C_{\mathbb{C}}(S\times S)$. In fact, a nowhere dense set is sometimes called a *sieve*.[5] This presents a problem in that the results we have so far for this case all deal with Eq. (2.2a) only, and Eq. (2.2b) then represents an additional necessary condition on the solutions spoken of in these results. The theorems below then, are uniqueness and construction theorems for Eq. (2.2a) as they stand. If the solutions constructed by the iteration procedures also satisfy (2.2b), we get the existence of a solution of (2.1) as well. However, the set of solutions of (2.2b) is rather sparse.

3. TWO UNIQUENESS THEOREMS

For the notation used in the remainder of this article, see Sec. 2 of Ref. 4. To solve (2.2a), we seek a fixed point of the transformation $\mathcal{H}$ defined on $C_{\mathbb{R}}(S\times S)$ by

$$\mathcal{H}(\varphi)(\hat{n}_1,\hat{n}_2)=\arcsin\int_S H(\hat{n}_1,\hat{n}_2,\hat{n})\cos[\varphi(\hat{n}_1,\hat{n})-\varphi(\hat{n}_2,\hat{n})]\,d\Omega.$$

We introduce the notation $M(G)=\sup\{\int_S H(\hat{n}_1,\hat{n}_2,\hat{n})\,d\Omega$: $\hat{n}_1,\hat{n}_2\in S\}$ and $M_1(G)=(2\pi)^{-1}M(G)^3[1-M(G)^2+M(G)^4]^{-1}\times\sup\{\int_S G(\hat{n}_1,\hat{n})\,d\Omega$: $\hat{n}_1\in S\}$. We first make the trivial remark that, for (2.2a) to have a solution, it is necessary that $\int_S G(\hat{n}_1,\hat{n})\,d\Omega\le 4\pi G(\hat{n}_1,\hat{n}_1)$ for every $\hat{n}_1\in S$. In the spherically symmetric case, the corresponding observation was made by Newton [Ref. 3, expression (8)].

Lemma 3.1: Let $G: S\times S\to \mathbb{R}^+$ be continuous, $M(G)<1$, and let φ be a square-integrable solution of the equation $\varphi = \mathcal{H}(\varphi)$. Then $0\leq\varphi(\hat{n}_1,\hat{n}_2)\leq \arcsin M(G)$ for almost every $[d_2\Omega]\hat{n}_1$ and $\hat{n}_2$ in S.

Proof: The proof of this result given by Martin[6] for continuous solutions in the spherically symmetric case is still valid in this more general case.

The importance of this lemma is that if we restrict $M(G)$ to be less than one, we need only look for fixed points of $\mathcal{H}$ in the set $D=\{\varphi\in X: 0\leq\varphi(\hat{n}_1,\hat{n}_2)\leq\arcsin M(G),\hat{n}_1,\hat{n}_2\in S\}$, where X is the appropriate function space.

Theorem 3.2: Let $G: S\times S\to \mathbb{R}^+$ be continuous and nonvanishing, and suppose $M(G)<[8^{-1}(\sqrt{17}-1)]^{1/2}\approx 0.6255$. Then there is a unique continuous solution of the equation $\varphi=\mathcal{H}(\varphi)$. The solution is the limit of a sequence of successive approximations which converges uniformly on $S\times S$.

Proof: Let $X=C_{\mathbb{R}}(S\times S)$ and let D be as above. Then because $M(G)<1$, $\mathcal{H}: D\to D$. Let $\varphi_1,\varphi_2\in D$. From the identity $\cos A-\cos B=2\sin\frac{1}{2}(A+B)\sin\frac{1}{2}(A-B)$ follows the Lipschitz condition

$$||\mathcal{H}(\varphi_1)-\mathcal{H}(\varphi_2)||\leq 2M(G)^2[1-M(G)^2]^{-1/2}||\varphi_1-\varphi_2||,$$

and then the result follows from the Banach contraction mapping principle.

The next theorem gives other sufficient conditions for the existence of a unique solution of Eq. (2.2a).

Theorem 3.3: Let $G: S\times S\to \mathbb{R}^+$ be continuous and nonvanishing, and suppose $M(G)<1$ and $M_1(G)<1$. Then there is a unique square-integrable solution of the equation $\varphi=\mathcal{H}(\varphi)$. The solution is the limit of a sequence of successive approximations which converges in the norm of $L^2(S\times S, d_2\Omega)$.

Proof: Let X be the completion of $C_{\mathbb{R}}(S\times S)$ in the norm

$$||x||^2=\int_{S\times S}G(\hat{n}_1,\hat{n}_2)|x(\hat{n}_1,\hat{n}_2)|^2 d_2\Omega. \tag{3.1}$$

X is the Hilbert space of real-valued measurable functions on $S\times S$ which are square-integrable in the measure $Gd_2\Omega$, which is the measure $d\Omega\times d\Omega$ on $S\times S$ with the weight G. Since $G>0$, the norm (3.1) is equivalent to the usual norm on $L^2(S\times S, d_2\Omega)$. Let

$$D=\{\varphi\in X: 0\leq\varphi(\hat{n}_1,\hat{n}_2)\leq\arcsin M(G) \text{ a.e. } [d_2\Omega]\};$$

D is a closed convex subset of X, and hence is a complete metric space with the distance induced by the norm (3.1). To solve (2.2a), we seek a fixed point of $\mathcal{H}$ in D. Since $M(G)<1$, $\mathcal{H}: X\to D$; in particular, $\mathcal{H}: D\to D$. Also, at each $\varphi\in D$, $\mathcal{H}$ has a Gateaux variation $\delta\mathcal{H}(\varphi;x)$ which satisfies the following equation: for every $x\in X$ and $\varphi\in D$,

$$\begin{aligned}4\pi G(\hat{n}_1,\hat{n}_2)\cos\mathcal{H}(\varphi)(\hat{n}_1,\hat{n}_2)\delta\mathcal{H}(\varphi;x)(\hat{n}_1,\hat{n}_2)&\\ =\int_S G(\hat{n}_1,\hat{n})G(\hat{n}_2,\hat{n})\sin[\varphi(\hat{n}_1,\hat{n})-\varphi(\hat{n}_2,\hat{n})]&\\ \times[x(\hat{n}_2,\hat{n})-x(\hat{n}_1,\hat{n})]\,d\Omega.&\end{aligned} \tag{3.2}$$

This is the defining equation for the Gateaux variation of $\mathcal{H}$ at φ in the direction x.[7] Using (2.2a), (3.2), and the method of Martin (Ref. 6, pp. 137–39), we obtain the inequality

$$\begin{aligned}&\int_{S\times S}G(\hat{n}_1,\hat{n}_2)\,|\delta\mathcal{H}(\varphi;x)(\hat{n}_1,\hat{n}_2)|^2 d_2\Omega\\ &\quad\leq M_1(G)\int_{S\times S}G(\hat{n}_1,\hat{n}_2)\,|x(\hat{n}_1,\hat{n}_2)|^2 d_2\Omega.\end{aligned}$$

In the language of the norm (3.1), this is $\|\delta\mathcal{H}(\varphi;x)\|\leq M_1(G)^{1/2}\|x\|$ for every $x\in X$ and $\varphi\in D$, or $\|\delta\mathcal{H}(\varphi;\cdot)\|\leq M_1(G)^{1/2}$ for every $\varphi\in D$. Let $\varphi_1,\varphi_2\in D$ and apply a mean value theorem (Ref. 7, theorem 5.4) to obtain the estimate

$$\begin{aligned}||\mathcal{H}(\varphi_1)-\mathcal{H}(\varphi_2)||&\leq||\varphi_1-\varphi_2||\\ &\times\sup\{||\delta\mathcal{H}((1-t)\varphi_1+t\varphi_2;\cdot)||: 0\leq t\leq 1\}\\ &\leq M_1(G)^{1/2}||\varphi_1-\varphi_2||.\end{aligned}$$

By the Banach contraction mapping principle, there is then exactly one solution of (2.2a) in D. But by Lemma 3.1, because $M(G)<1$, all solutions of (2.2a) are in D, so that this solution is the only one in all of X. The square-integrability of the solution and the convergence of the approximations in $L^2(S\times S, d_2\Omega)$ follow from the equivalence of the norm (3.1) and the usual norm on $L^2(S\times S, d_2\Omega)$.

Remarks: (1) In the spherically symmetric case, $\int_S G(\hat{n}_1,\hat{n})\,d\Omega$ is independent of $\hat{n}_1$; and so no "sup" is required in the definition of $M_1(G)$, and in this case the condition $M_1(G)<1$ is the original condition of Martin (Ref. 6, expression 31).

(2) Theorem 3.2 is not included in Theorem 3.3 because it is presumably possible that $M(G)<0.6255$ while $M_1(G)>1$. Of course, in the spherically symmetric case this cannot occur.

(3) A significant difference in this case in which no spherical symmetry is present is that we cannot guarantee that the solution φ which corresponds to the continuous G is itself continuous. In the spherically symmetric case, an easy approximation argument shows that if G is continuous and $M(G)<1$, then $\mathcal{H}$ takes L^2 functions onto continuous functions (Anderson, Johnson, and Warnock also show this in a different way in Ref. 1, Theorem 2). When there is no spherical symmetry, if G is continuous and satisfies $\int_S G(\hat{n}_1,\hat{n})\,d\Omega\leq 4\pi\mu G(\hat{n}_1,\hat{n}_1)$ for some $\mu<1$ and every $\hat{n}_1,\hat{n}_2\in S$, then $\mathcal{H}(\varphi)$ is at least continuous in the forward direction (i.e., on the diagonal $\hat{n}_2=\hat{n}_1$), because

$$\begin{aligned}4\pi|\mathcal{H}(\varphi)(\hat{n}_1,\hat{n}_1)-\mathcal{H}(\varphi)(\hat{n}_1',\hat{n}_1')|&\\ \leq(1-\mu^2)^{-1/2}\int_S\left|\frac{G(\hat{n}_1,\hat{n})^2}{G(\hat{n}_1,\hat{n}_1)}-\frac{G(\hat{n}_1',\hat{n})^2}{G(\hat{n}_1',\hat{n}_1')}\right|d\Omega.&\end{aligned}$$

(4) The transformation $\mathcal{H}$, even in the spherically symmetric case, is not Frechet differentiable in the space X (or in $L^2[-1,1]$ in the spherically symmetric case). Thus, the mean value theorem as it appears in Dieudonne,[8] which is used in Ref. 1, Theorem 2, is not applicable. However, this is not at all serious, since, as we have seen, $\mathcal{H}$ has a Gateaux variation (in fact it can be shown that $\mathcal{H}$ has a Gateaux derivature) and there

are stronger mean value theorems available than 8.5.4 of Dieudonne: for example, the one we used in Theorem 3.3, or proposition 2.3 of Ref. 9. Also, as indicated in remark (1), our Theorem 3.3 includes the spherically symmetric case as a special case. An excellent discussion of many of the aspects of differentials in nonlinear analysis is given in Ref. 7.

4. AN EXAMPLE OF NONUNIQUENESS

The following example illustrates that the hypothesis that G be nonvanishing cannot be relaxed in uniqueness theorems such as Theorems 3.2 and 3.3. The idea for this example is due to Professor Michael Golomb, and the author wishes to express his gratitude to Professor Golomb for his help.

Let ψ be a real-valued continuous function on S such that ψ assumes both positive and negative values and $\int_S |\psi(\hat{n})|^2 d\Omega = 4\pi$. Let $\alpha \in (0,1)$. Put $G(\hat{n}_1,\hat{n}_2) = \alpha^{1/2}|\psi(\hat{n}_1)||\psi(\hat{n}_2)|$ and $F(\hat{n}_1,\hat{n}_2) = \beta\psi(\hat{n}_1)\psi(\hat{n}_2)$, where $\beta = (\alpha - \alpha^2)^{1/2} + i\alpha$. Then $|F| = G$ and F satisfies (2.1) since $|\beta|^2 = \alpha$ and ψ is real-valued. Now let ψ_1 be constructed from ψ by replacing $\psi(\hat{n}) - \psi(\hat{n})$ in some places in such a way that ψ_1 is continuous. Then $F_1(\hat{n}_1,\hat{n}_2) = \beta\psi_1(\hat{n}_1)\psi_1(\hat{n}_2)$ also satisfies (2.1), and $|F_1| = G$, but $F_1 \neq F$ and $F_1 \neq -\bar{F}$.

Also, we have $H(\hat{n}_1,\hat{n}_2,\hat{n}) = (4\pi)^{-1}\alpha^{1/2}\psi(\hat{n})^2$, and so $M(G) = \alpha^{1/2}$. Putting $p^* = \sup\{|\psi(\hat{n})| : \hat{n} \in S\}$, we have $M_1(G) \leq 2p^*\alpha^{1/2}(1 - \alpha + \alpha^2)^{-1}$, and both $M(G)$ and $M_1(G)$ can be made less than one by choosing α small enough. Then all the hypotheses of Theorem 3.3 (or Theorem 3.2) are satisfied, except that G be nonvanishing, and we have two (nontrivially different, see Ref. 3, Sec. 2) solutions of (2.1).

From the physical point of view, this example certainly seems rather contrived. However, it does serve to point out a difficulty which arises in treating the unitarity equation in a purely mathematical way. In fact, the example is all the more useful because it deals directly with Eq. (2.1), rather than with (2.2a).

5. CONCLUSION

While giving some results for the solution of the unitarity equation for scattering in the absence of spherical symmetry when the differential cross section is known, we have tried to point out some of the problems which arise which are peculiar to this case. The most prominent of these is that, in this case, there is an additional necessary condition on the scattering amplitude, represented by equation (2.2b). There is no reason to believe *a priori* that a function which satisfies (2.2a) will also satisfy (2.2b) without some further conditions; in fact, we have seen (Proposition 2.1) that the set of solutions of (2.2b) is rather widely distributed in $C_c(S\times S)$. Thus the conditions given in Theorems 3.2 and 3.3 are sufficient conditions only for the uniqueness of solutions of (2.1) [while they do give existence for (2.2a)]; the treatment of the question of existence of solutions of (2.1) must also take into account Eq. (2.2b). We do know that using these theorems, we cannot expect any better conditions for existence of solutions of (2.1); however, other methods may yield more promising results.

We have also given an example to show that even under fairly restricted assumptions on the differential cross section, if this cross section vanishes, Eq. (2.1) may have two essentially different solutions. This example as given is peculiar to this case in which there is no spherical symmetry, for this construction cannot be made if the potential is spherically symmetric.

ACKNOWLEDGMENTS

Parts of the above material are from the author's Ph.D. Dissertation, completed at Purdue University. The author would like to thank Professor Vaughan Weston for suggesting this problem and for his direction in the preparation of the dissertation.

[1]D. Atkinson, P.W. Johnson, and R.L. Warnock, Commun. Math. Phys. 28, 133 (1972), Theorem 2.
[2]R. Glauber and V. Schomaker, Phys. Rev. 89, 667 (1953).
[3]R.G. Newton, J. Math. Phys. 9, 2050 (1968).
[4]M. Tortorella, J. Math. Phys. 13, 1764 (1972).
[5]J. Dugundji, *Topology* (Allyn and Bacon, Boston, 1966), p. 250.
[6]A. Martin, Nuovo Cimento A 59, 131 (1969); see p. 133.
[7]M.Z. Nashed, "Differentiability and Related Properties of Nonlinear Operations: Some Aspects of the Role of Differentials in Nonlinear Functional Analysis," in *Nonlinear Functional Analysis and Applications*, edited by L.B. Rall (Academic, New York, 1971), Sec. 1.1.
[8]J. Dieudonne, *Foundations of Modern Analysis* (Academic, New York, 1960), Vol. I, 8.5.4.
[9]R.A. Tapia, "The Differentiation and Integration of Nonlinear Operators," in Ref. 7.

The power spectrum of the Mellin transformation with applications to scaling of physical quantities

H. E. Moses and A. F. Quesada

Air Force Cambridge Laboratories, L. G. Hanscom Field, Bedford, Massachusetts 01730
(Received 14 November 1973)

The Mellin transform is used to diagonalize the dilation operator in a manner analogous to the use of the Fourier transform to diagonalize the translation operator. A power spectrum is also introduced for the Mellin transform which is analogous to that used for the Fourier transform. Unlike the case for the power spectrum of the Fourier transform where sharp peaks correspond to periodicities in translation, the peaks in the power spectrum of the Mellin transform correspond to periodicities in magnification. A theorem of Wiener–Khinchine type is introduced for the Mellin transform power spectrum. It is expected that the new power spectrum will play an important role extracting meaningful information from noisy data and will thus be a useful complement to the use of the ordinary Fourier power spectrum.

1. INTRODUCTION

In Refs. 1 and 2 we introduced the scale-Euclidean group in one and three dimensions and showed how physical quantities could be expanded in terms of the irreducible representations of the group. One could then use the expansion to construct correlations between physical quantities which are independent of the translation of the origin of coordinates, of the units used in measuring the physical quantities, and of the orientation of the axes (in three dimensions). The expansion which was given is a generalization of a Fourier expansion and the Fourier coefficients provide a basis for diagonalizing the translation operator. One can introduce a power spectrum in the Fourier transform space, which is a slight modification of the usual power spectrum that has been found so useful in data analysis, particularly when noise is present.

One of the elements of the scale-Euclidean group is the dilation. One can expand the physical quantities in terms of the irreducible representations of the dilation operator and define a power spectrum in terms of the amplitudes. This expansion corresponds to the use of the (imaginary) Mellin transform. The power spectrum associated with the Fourier transform can be used to detect periodicities in the physical function, since the wavenumbers at which sharp peaks of the spectrum occur give the wavelength of such periodicities. By contrast, the positions of the peaks in the spectrum associated with the Mellin transform give the magnification (or compression) which will reproduce features in the physical function. We believe that the power spectrum associated with the Mellin transform is as important as the power spectrum associated with the Fourier transform. The use of both power spectra together appears to be a most useful tool in extracting meaningful features from data.

It is the object of the present paper to show the utility (theoretically, at least) of the Mellin transform in the one-dimensional case. We shall therefore review briefly the expansions of physical quantities in terms of the scale-Euclidean group in the one-dimensional case and show how the Mellin transform makes its appearance. We shall then review properties of the power spectrum associated with the Fourier transform and give the analogs of those associated with the Mellin transform. Finally, we shall derive an analog of the Wiener–Khinchine theorem which will greatly simplify the calculation of the new power spectrum.

2. THE SCALE-TRANSLATION GROUP. EXPANSIONS OF PHYSICAL QUANTITIES IN TERMS OF THE IRREDUCIBLE REPRESENTATIONS

A. The group and its irreducible representations

We shall concern ourselves with the scale-translation group in one dimension which is a subgroup of the scale-Euclidean group in one dimension. Let us consider a one-dimensional coordinate system, in which the coordinate is labelled x. The variable x may be a space variable or a time variable. For many applications, the interpretation as a time variable is more interesting.

We label the transformation of the coordinate

$$x' = x - a \tag{1}$$

$T(a)$. This transformation is called the translation.

The transformation

$$x' = e^{\lambda} x \tag{2}$$

is called the scale transformation or dilation and corresponds to a change of scale. The transformation is denoted by $S(\lambda)$. The set of transformations form a group with the multiplication laws

$$T(0) = S(0) = I, \quad T(a)T(b) = T(a+b),$$
$$S(\lambda)S(\mu) = S(\lambda+\mu), \quad S(\lambda)T(a) = T(e^{\lambda}a)\,S(\lambda), \tag{3}$$

where I is the identity transformation.

The scale-translation group is the abstract group whose elements satisfy the multiplication rules (3). In Ref. 1 it is shown that irreducible unitary representations of the group in a separable Hilbert space are two in number. The Hilbert space corresponding to one of the representations is the space of complex functions $\{f(p)\}$ defined for $p > 0$. The inner product and unitary operators are defined by

$$(f^{(1)}, f) = \int_0^{\infty} f^{(1)*}(p) f(p) \frac{dp}{|p|}, \tag{4}$$

$$T(a)f(p) = \exp(iap) f(p), \quad S(\lambda)f(p) = f[\exp(\lambda)p], \tag{5}$$

where the asterisk means complex conjugate and the unitary operators are given the same name as the corresponding elements of the transformation and abstract

groups. We shall always label the operators this way in any representation.

The second irreducible representation consists of the space of functions $\{f(p)\}$ defined for $p<0$ with the inner product

$$(f^{(1)},f)=\int_{-\infty}^{0} f^{(1)*}(p)f(p)\frac{dp}{|p|}. \tag{6}$$

The operators representing the group elements are given by Eq. (5).

[It should be mentioned that our results for the irreducible representations of the scale-translation group in Ref. 1 were anticipated in Ref. 3 in a different form and with a different derivation. However, the expansion of physical quantities in terms of the irreducible representations, the definition of correlations, and the generalization to three dimensions which are given in Refs. 1 and 2 are not in Ref. 3.]

B. Physical quantities and their expansion in terms of the irreducible representations

We now consider field quantities or physical quantities $v(x)$. The quantity v might, for example, be a linear density or velocity component in one dimension. It will have a dimensionality in units of length L. For example, if $v(x)$ is a linear density, $v(x)\sim L^{-1}$. If $v(x)$ is a velocity, $v(x)\sim L$. If we now regard x as time instead of a distance, and take $v(x)$ to be an acceleration, $v(x)\sim L^{-2}$. Generally we shall write as an expression of dimensionality,

$$v(x)\sim L^{N}, \tag{7}$$

where N is any real number.

Under the transformation $T(a)$, $v(x)$ transforms to $v'(x')$ which is the same variable measured in the transformed frame. Clearly

$$v'(x)=v(x+a). \tag{8}$$

We shall write Eq. (8) as an operator relation

$$T(a)v(x)=v(x+a). \tag{9}$$

Similarly under the scale transformation $S(\lambda)$

$$v'(x)=\exp(N\lambda)v[\exp(-\lambda)x] \tag{10}$$

which is in operator notation

$$S(\lambda)v(x)=\exp(N\lambda)v[\exp(-\lambda)x]. \tag{11}$$

On constructing the space of functions $\{v(x)\}$ with the same dimensionality N, it is seen that the operators $T(a)$ and $S(\lambda)$ are linear operators. Furthermore, they satisfy the multiplication laws Eq. (3) for the scale-translation group. Hence the space $\{v(x)\}$ provides a representation of the scale-translation group for every N.

Let us now assume that the functions $v(x)$ have a Fourier transform and that their average is zero, i.e.,

$$\int_{-\infty}^{+\infty} v(x)\,dx=0. \tag{12}$$

That is, we regard $v(x)$ as being measured from a mean value.

We may then write

$$v(x)=(2\pi)^{-1/2}\int_{-\infty}^{+\infty}|p|^{-(N+1)}f(p)\exp(ipx)\,dp. \tag{13}$$

From Eq. (12) it follows that

$$\lim_{|p|\to 0}|p|^{-(N+1)}f(p)=0. \tag{14}$$

It is readily seen that

$$T(a)v(x)=(2\pi)^{-1/2}\int_{-\infty}^{+\infty}|p|^{-(N+1)}[T(a)f(p)]\exp(ipx)\,dp, \tag{15}$$

$$S(\lambda)v(x)=(2\pi)^{-1/2}\int_{-\infty}^{+\infty}|p|^{-(N+1)}[S(\lambda)f(p)]\exp(ipx)\,dp, \tag{16}$$

where $T(a)f(p)$ and $S(\lambda)f(p)$ are defined in Eq. (5). Thus, by rewriting the Fourier amplitudes slightly, we have been able to reduce the representation of the group operators acting on the physical quantities. In particular, it is noted that the translation operator is diagonalized. It is also to be noted that the functions $f(p)$ are dimensionless.

We shall now introduce an inner product between two functions $v^{(1)}(x)$ and $v(x)$ of the set $\{v(x)\}$, namely

$$(v^{(1)},v)=\int_{-\infty}^{+\infty}f^{(1)*}(p)f(p)\frac{dp}{|p|}. \tag{17}$$

With this inner product, the representation of the operators acting on $\{v(x)\}$ is a unitary representation and Eqs. (13), (15), and (16) give the reduction of this representation to the irreducible unitary representations of the group.

The inner product $(v^{(1)},v)$ is invariant under the transformations of the group, that is, it has the same value under the translation of the coordinate Eq. (1) and under the change of scale of units Eq. (2). In Ref. 1 we defined the correlation between $v^{(1)}(x)$ and $v(x)$ as being the inner product (17). Hence this correlation is invariant under the transformations of the group. We further defined the translation autocorrelation (TAC) as being given by $(v, T(a)v)$ which is closely related to the usual autocorrelation except that it is now scale-invariant. Furthermore, we defined a new type of autocorrelation called the scale autocorrelation (SAC) by $(v, S(\lambda)v)$. The TAC compares a function with itself when the function is shifted. The SAC compares a function with itself when it is stretched.

As usual the norm of $v(x)$, denoted by $\|v\|$, is defined by

$$\|v\|=[(v,v)]^{1/2} \tag{18}$$

and is an invariant magnitude of the physical quantity.

It is readily seen that a necessary and sufficient condition that $v(x)$ be real is that

$$f(-p)=f^{*}(p) \tag{19}$$

in the expansion Eq. (13).

The power spectrum of $v(x)$ in this representation, in which the operator $T(a)$ is diagonal, is defined by $|f(p)|^2$ which is nondimensional. The amount of power in the interval $a<p<b$ is

$$\int_{a}^{b}|f(p)|^{2}\frac{dp}{|p|}.$$

3. DIAGONALIZATION OF THE SCALE TRANSFORMATION OPERATOR THROUGH THE USE OF THE IMAGINARY MELLIN TRANSFORM

A. Transformation from the space in which $T(a)$ is diagonal to the space in which $S(\lambda)$ is diagonal

For $-\infty < s < +\infty$ and $m = \pm 1$, let us define the complex function $G(s, m)$ by

$$G(s, +1) = (2\pi)^{-1/2} \int_0^\infty |p|^{-is-1} f(p)\, dp, \tag{20}$$

$$G(s, -1) = (2\pi)^{-1/2} \int_{-\infty}^0 |p|^{-is-1} f(p)\, dp. \tag{21}$$

It is readily seen that the inverse of Eqs. (20) and (21) are respectively

$$f(p) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} |p|^{is} G(s, +1)\, ds \quad \text{for } p > 0, \tag{22}$$

$$f(p) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} |p|^{is} G(s, -1)\, ds \quad \text{for } p < 0. \tag{23}$$

These transformations provide another basis for the scale-translation group. From Eqs. (22) and (23) it follows that

$$S(\lambda) f(p) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} |p|^{is} [S(\lambda) G(s, +1)]\, ds, \quad \text{for } p > 0, \tag{24}$$

$$S(\lambda) f(p) = (2\pi)^{-1/2} \int_{-\infty}^{+\infty} |p|^{is} [S(\lambda) G(s, -1)]\, ds, \quad \text{for } p < 0, \tag{25}$$

where

$$S(\lambda) G(s, m) = \exp(i\lambda s) G(s, m). \tag{26}$$

One also has preservation of the inner product in the form

$$\int_{-\infty}^{+\infty} G^{(1)*}(s, +1) G(s, +1)\, ds = \int_0^{+\infty} f^{(1)*}(p) f(p) \frac{dp}{|p|},$$

$$\int_{-\infty}^{+\infty} G^{(1)*}(s, -1) G(s, -1)\, ds = \int_{-\infty}^0 f^{(1)*}(p) f(p) \frac{dp}{|p|}. \tag{27}$$

The variable $m = \pm 1$ labels the two irreducible representations of the group. While the above statements can be verified directly, the motivation for the form of the transformation will be given in Appendix A, using techniques of Ref. 1.

The transformations Eqs. (20)—(23) are seen to be Mellin transforms with an imaginary argument is (see, e.g., Ref. 4). This transformation is related to the usual Mellin transform as the Laplace transform is related to the Fourier transform.

B. Expansion of physical quantities in the new basis. Power spectra

By substituting Eqs. (22) and (23) into Eq. (13) we obtain the following expansion for $v(x)$:

$$v(x) = (2\pi)^{-1} \sum_m \int_{-\infty}^{+\infty} (-imx)^{N-is} \Gamma(-N + is) G(s, m)\, ds. \tag{28}$$

The inverse transformation is

$$G(s, m) = (2\pi)^{-1} \Gamma(N - is + 1) \int_{-\infty}^{+\infty} (imx)^{is-N-1} v(x)\, dx. \tag{29}$$

Equations (28) and (29) are verified in Appendix B. From Eqs. (17) and (27) the inner product of $v^{(1)}$ and v is

$$(v^{(1)}, v) = \sum_m \int_{-\infty}^{+\infty} G^{(1)*}(s, m) G(s, m)\, ds. \tag{30}$$

We note that Eqs. (28) and (29) are a generalized imaginary Mellin transform pair.

We define the power spectrum with respect to the Mellin transform as

$$\sum_m |G(s, m)|^2.$$

A necessary and sufficient condition that $v(x)$ be real is

$$G^*(s, 1) = G(-s, -1). \tag{31}$$

4. SIGNIFICANCE OF PEAKS IN THE POWER SPECTRA

A. Fourier power spectrum

Let us assume that the power spectrum $|f(p)|^2$ has a peak $p \sim k$. There is then a contribution to $v(x)$ from $f(p)$ in the vicinity of the peak given by

$$v_k(x) \sim (2\pi)^{-1/2} |k|^{-(N+1)} f(k) \exp(ikx). \tag{32}$$

If $v(x)$ is real, $f(p)$ also has a contribution near $p \sim -k$. The sum of the two contributions give

$$v_k(x) \sim (2/\pi)^{1/2} |k|^{-(N+1)} |f(k)| \cos(kx + \phi), \tag{33}$$

where ϕ is given by $f(k) = |f(k)| \exp(i\phi)$. For either Eq. (32) or Eq. (33).

$$v_k(x + 2\pi n/k) \sim v_k(x), \tag{34}$$

where n is any integer, positive or negative. Equation (34) states that $v(x)$ contains a feature which is invariant under translations which are multiples of the wavelength $\lambda = (2\pi/k)$. Of course, this discussion is the familiar one for identifying peaks of the power spectrum with translationally invariant components. We are reviewing the theory because we shall give analogs to it when we consider the power spectrum associated with the Mellin transform.

If $f(p)$ were infinitely sharp at $p = k$, i.e., if $f(p)$ were a delta function centered at $p = k$, the approximations of Eqs. (32), (33), and (34) would be replaced by equalities. However, the peaks have a "line width." The approximations become increasingly poor with increasing line width. In particular, one expects the periodicity condition Eq. (34) to become increasingly poor with increasing values of $|n|$. We shall characterize the degree of approximation by giving a bound such that, if $|n|$ does not exceed this bound, the periodicity condition of Eq. (34) is satisfied within a prescribed error.

To obtain this criterion, we shall consider only the contribution to $v(x)$ from the peak. We shall thus take

$$f(p) \equiv 0 \quad \text{for } p < k - \Delta \text{ and for } p > k + \Delta, \tag{35}$$

where 2Δ is the line width $(0 < 2\Delta < |k|)$.

Let us define the relative error in periodicity by

$$E = \|[T(2\pi n/k) - I] v\| / \|v\|. \tag{36}$$

The error will depend on n and on Δ.

From Eqs. (17), (15), and (5),

$$\|v\|^2 = \int_{k-\Delta}^{k+\Delta} |f(p)|^2 \frac{dp}{|p|}, \tag{37}$$

$$\left\| \left[T\left(\frac{2\pi n}{k}\right) - I \right] v \right\|^2 = 4 \int_{k-\Delta}^{k+\Delta} \left(\sin \frac{\pi n p}{k} \right)^2 |f(p)|^2 \frac{dp}{|p|}. \quad (38)$$

But by the second mean value theorem of integral calculus and from Eq. (37)

$$\| [T(2\pi n/k) - I] v \|^2 = 4[\sin(\pi n/k)(k+\xi)]^2 \| v \|^2$$
$$= 4[\sin(\pi n \xi/k)]^2 \| v \|^2, \quad (39)$$

where

$$-\Delta < \xi < +\Delta. \quad (39')$$

Thus

$$E < 2\pi(\Delta/|k|)\,|n|. \quad (40)$$

Thus if

$$|n| < (|k|/2\pi\Delta) E_0, \quad (41)$$

where E_0 is a prescribed (positive) limit of error, then

$$E < E_0. \quad (42)$$

Equations (41) and (42) provide the criterion.

B. Mellin power spectrum

Let us assume that the Mellin power transform $\sum_m |G(s,m)|^2$ has a peak at $s = r$. Then at least one of the functions $G(s,+1)$ or $G(s,-1)$ also has a peak in absolute value near $s = r$. Without much loss in generality, let us assume that $G(s,+1)$ has this peak and that $G(s,-1)$ is small. The contribution to $v(x)$ from $G(s,+1)$ near the peak is

$$v_r(x) \sim (2\pi)^{-1}(-ix)^{N-ir}\Gamma(-N+ir)\,G(r,+1). \quad (43)$$

If $v(x)$ is real, $G(s,-1)$ has a peak near $-r$. The contribution of both peaks is given by

$$v_r(x) \sim (\pi)^{-1} A\,|x|^N \exp[-\rho\pi r/2)]$$
$$\times \cos[r \log|x| + N\rho\pi/2 + \phi], \quad (44)$$

where

$$\rho = \operatorname{sgn} x, \quad A = |G(r,+1)\Gamma(-N+ir)|, \quad (44'a)$$

and ϕ is given by

$$G(r,+1)\Gamma(-N+ir) = Ae^{-i\phi}. \quad (44'b)$$

In either case $v_r(x)$ exhibits the property

$$\exp(N2\pi n/r) v_r[\exp(-2\pi n/r)x] = S(2\pi n/r) v_r(x) \sim v_r(x), \quad (45)$$

where n is a positive or negative integer.

The significance of Eq. (45) is that a peak in the power spectrum of the Mellin transform gives a contribution to $v(x)$ such that when this feature is stretched (for positive n) or compressed (for negative n) by multiplying the units by the scale factor $\exp(2\pi n/r)$ it is identical to the original feature. Furthermore, the stretching and compressing needed to obtain the original feature is *periodic*. We believe that periodicities in scale may be as important as periodicities in translation, especially where noise is present. Moreover, one can set up a theory of noise using the Mellin transform in a very similar way to the usual theory in terms of the Fourier transform. It seems likely that the use of both transforms can lead to better ways of extracting meaningful information from a noisy background.

As in the Fourier power spectrum case, the width of the peak in the power spectrum determines the accuracy of Eqs. (43), (44), and (45). We can set up an error estimate similar to that of Eqs. (41) and (42). To study such a peak, let us defines

$$G(s,m) \equiv 0, \quad \text{for } r-\Delta < s < r+\Delta. \quad (46)$$

We also define a relative error E as in Eq. (36) but replace $T(2\pi n/k)$ by $S(2\pi n/r)$. Then Eqs. (41) and (42) hold with $|k|$ replaced by $|r|$.

5. A THEOREM OF WIENER-KHINCHINE TYPE

The direct way to obtain the new power spectrum $\sum_m |G(s,m)|^2$ is to use the Mellin transformation Eq. (29). While in principle there is no difficulty in using the Mellin transform, the numerical evaluation of such a transform has not yet been sufficiently studied to enable one to do this easily. We shall show that the power spectrum can be obtained using a Fourier transform. Of course, the Fourier transform has been studied exhaustively from a numerical point of view and therefore offers, for the present, great advantages.

The scale autocorrelation $F(\lambda)$ is defined by

$$F(\lambda) = (v, S(\lambda)v). \quad (47)$$

But from Eqs. (5) and (17)

$$F(\lambda) = \int_{-\infty}^{+\infty} f^*(p) f[\exp(\lambda)p] \frac{dp}{|p|}, \quad (48)$$

where $f(p)$ is obtained from $v(x)$ using the somewhat generalized Fourier transform which is the inverse of Eq. (13):

$$f(p) = (2\pi)^{-1/2} |p|^{(N+1)} \int_{-\infty}^{+\infty} v(x) \exp(-ipx)\,dx. \quad (49)$$

In Refs. 1 and 5 it is noted that if $|f(p)|$ has sharp maxima at $p = k_i$ $(i = 1, 2, \cdots)$, then $|F(\lambda)|$, in general, has maxima at $\lambda = \lambda_{ij}$ where $\lambda_{ij} = \log(k_i/k_j)$. This property of $F(\lambda)$ helps one identify periodicities in $v(x)$ or, equivalently, peaks in the Fourier power spectrum.

On using Eq. (26) and Eq. (30), one obtains another expression for $F(\lambda)$, namely,

$$F(\lambda) = \int_{-\infty}^{+\infty} \left(\sum_m |G(s,m)|^2 \right) \exp(i\lambda s)\,ds. \quad (50)$$

Thus the Mellin power spectrum is the Fourier transform of $F(\lambda)$:

$$\sum_m |G(s,m)|^2 = (2\pi)^{-1} \int_{-\infty}^{+\infty} F(\lambda) \exp(-i\lambda s)\,d\lambda. \quad (51)$$

The substitution of the right-hand side of Eq. (48) gives the Mellin power spectrum entirely in terms of Fourier transforms.

Equations (48) and (51) constitute a theorem of the Wiener–Khinchine type.

ACKNOWLEDGMENT

We wish to thank Dr. Albion D. Taylor of the Air Resources Laboratories for his suggestion of the use of the Mellin transform to diagonalize the dilation operator and for bringing to our notice Ref. 3.

APPENDIX A: DERIVATION OF THE TRANSFORMATION BETWEEN THE REPRESENTATION WHICH DIAGONALIZES THE TRANSMISSION AND THAT WHICH DIAGONALIZES THE DILATION

For brevity we shall assume that the reader is acquainted with Ref. 1, particularly Sec. 4. References to equations therein will be denoted by a prime.

We consider the space of kets $|p\rangle$ for $p\rangle 0$ and for which the spectrum is continuous. Hence the kets satisfy the completeness relation Eq. (69)'.

We recollect that

$$|p\rangle = \exp(-i\mu D)|1\rangle, \quad \mu = \log p \tag{A1}$$

from Eqs. (50)', (37)', and (47)'.

Let the eigenvalues of the Hermitian operator D be denoted by the real numbers s and let the corresponding eigenkets be denoted by $|s\rangle$. Hence

$$D|s\rangle = s|s\rangle. \tag{A2}$$

Of course, the kets $|s\rangle$ also diagonalize $S(\lambda)$.

$$S(\lambda)|s\rangle = \exp(i\lambda s)|s\rangle. \tag{A2'}$$

The transformation between the basis which diagonalizes the translation operator and that which diagonalizes the dilation operator is $\langle s|p\rangle$. But

$$\langle s|p\rangle = \langle s|\exp(-i\mu D)|1\rangle = \exp(-i\mu s)K(s)$$
$$= \exp(-is\log p)K(s) = p^{-is}K(s), \tag{A3}$$

where $K(s) = \langle s|1\rangle$. Now if $|s\rangle$ diagonalizes $S(\lambda)$, so does $[(2\pi)^{1/2}K^*(s)]^{-1}|s\rangle$. On using the latter set of kets we have

$$\langle s|p\rangle = (2\pi)^{-1/2}p^{-is}. \tag{A4}$$

Now from the completeness relation Eq. (69)'

$$\langle s|s'\rangle = \int_0^\infty \langle s|p\rangle \frac{dp}{|p|}\langle p|s'\rangle. \tag{A4a}$$

On using Eq. (A4) and changing the variable of integration p in a suitable fashion, one finds

$$\langle s|s'\rangle = \delta(s - s'), \tag{A5}$$

which is the orthonormality relation between the kets.

Similarly one finds

$$\int_{-\infty}^{+\infty} \langle p|s\rangle ds \langle s|p'\rangle = \delta(p - p'). \tag{A6}$$

Equation (A6) is a completeness relation for the kets $|s\rangle$ and is equivalent to the resolution of the identity

$$I = \int_{-\infty}^{+\infty} |s\rangle ds \langle s|. \tag{A7}$$

Equations (A6) and (A7) show that the spectrum of D is continuous and extends over the entire real axis.

Let us define $G(s)$ as being $|\Phi\rangle$ in the $|s\rangle$ representation:

$$G(s) = \langle s|\Phi\rangle. \tag{A8}$$

Then from the completeness relation (69)'

$$G(s) = \int_0^\infty \langle s|p\rangle \frac{dp}{|p|}\langle p|\Phi\rangle. \tag{A9}$$

On using Eq. (A4) and (56)' one obtains (in a slightly different notation) Eq. (20). Equations (21), (22), and (23) are derived similarly.

APPENDIX B: THE TRANSFORMATION FROM THE PHYSICAL SPACE TO THE SPACE IN WHICH THE DILATION OPERATOR IS DIAGONAL

On substituting Eqs. (22) and (23) into Eq. (13) we obtain

$$v(x) = (2\pi)^{-1}\sum_m \int_0^{+\infty} p^{-(N+1)}\exp(imxp)\,dp \int_{-\infty}^{+\infty} p^{is}G(s,m)\,ds. \tag{B1}$$

Interchanging order of integration gives

$$v(x) = (2\pi)^{-1}\sum_m G(s,m)\,ds \int_0^{+\infty} p^{(is-N-1)}\exp(impx)\,dp. \tag{B2}$$

But from Eqs. (4 – 1 – 7) and (4 – 1 – 8) of Ref. 4

$$\int_0^{+\infty} p^{(is-N-1)}\exp(impx)\,dp = \Gamma(is - N)(-imx)^{N-is}. \tag{B3}$$

On substituting Eq. (B3) into Eq. (B2) we obtain Eq. (28). However, Eq. (B3) holds, as an expression for the integral, only for $-1\langle N\langle 0$. For other real values of N, Eq. (B3) is to be regarded as an equality in the sense of distributions. The proof that Eq. (B3) is an equality in the sense of distributions follows standard procedures. One multiplies both sides of Eq. (B3) by a suitable test function $f(x)$ and integrates with respect to x. On differentiating by parts in a suitable fashion one proves the result. For the sake of brevity we refrain from details.

Equation (29) is proved as follows: Eq. (49) is substituted into Eq. (20) and (21) and the order of integration is interchanged. The use of Eq. (B3) leads to the final result.

[1] H.E. Moses and A.F. Quesada, Arch. Rat. Mech. Anal. 44, 217 (1972).
[2] H.E. Moses and A.F. Quesada, Arch. Rat. Mech. Anal. 50, 194 (1973).
[3] I. Gel'fand and M. Neumark, C.R. (Dokl.) Acad. Sci. USSR 55, 567 (1947).
[4] I.N. Sneddon, *Use of Integral Transforms* (McGraw-Hill, New York, 1972).
[5] V. Corbin, H.E. Moses, and A.F. Quesada, J. Geoph. Res. 78, 6199 (1973).

Generating functions of the hypergeometric functions

E. Kyriakopoulos

Nuclear Research Center "Democritos," Aghia Paraskevi Attikis, Athens, Greece

*Theory Division, CERN, Geneva, Switzerland**

(Received 24 September 1973)

The Lie algebra, which was introduced in a previous paper to treat the hypergeometric functions by Lie theory techniques, is used to derive generating functions of the hypergeometric functions. Several generating functions are obtained from the theory of multiplier representations. Weisner's method is also applied, giving another generating function.

1. INTRODUCTION

In a previous paper[1] six operators forming the Lie algebra D_2 were introduced, which transform the hypergeometric functions among themselves. Therefore these operators can be used to treat the hypergeometric functions by Lie theory techniques. This is done in the present paper.

In Sec. 2 the Lie algebra D_2 is used to derive generating functions of the hypergeometric functions. First, three generators forming the subalgebra $sl(2)$ are employed. From the theory of multiplier representations of local Lie groups and the matrix elements of the representations[2] $D(u, m_0)$ and $\uparrow u$ of the algebra $sl(2)$, two generating functions of the hypergeometric functions are derived. Several interesting relations are obtained as special cases of them, in which a hypergeometric function is expanded in a series of hypergeometric functions of another variable.

Subsequently the boost operator in the direction x_3 is used to derive two more generating functions in a similar fashion. The matrix elements of this operator corresponding to finite rotations were employed, which were calculated recently[3-6] for both series of unitary representations of the Lorentz group[7,8], the principal series and the supplementary series. In Sec. 3 Weisner's method[9,10] is applied, and another generating function of the hypergeometric functions is derived.

In Ref. 1 we considered the operators

$$L_{12}=t\frac{\partial}{\partial t},$$

$$L_{14}=\frac{i}{2}\left[t\left(w\frac{\partial}{\partial w}-t\frac{\partial}{\partial t}+\sigma\right)+\frac{1}{t}\left((w-1)\frac{\partial}{\partial w}+t\frac{\partial}{\partial t}+\sigma\right)\right],$$

$$L_{24}=\frac{1}{2}\left[-t\left(w\frac{\partial}{\partial w}-t\frac{\partial}{\partial t}+\sigma\right)+\frac{1}{t}\left((w-1)\frac{\partial}{\partial w}+t\frac{\partial}{\partial t}+\sigma\right)\right], \tag{1.1}$$

$$L_{34}=i\left(2w(w-1)\frac{\partial}{\partial w}+t\frac{\partial}{\partial t}+2(\sigma-k)w+k-\sigma\right),$$

$$L_{32}=\frac{[w(t^2-1)+1](2w-1)}{2t}\frac{\partial}{\partial w}+\frac{t^2-1}{2}\frac{\partial}{\partial t}-k\frac{w(t^2-1)+1}{t}$$

$$+\sigma\frac{(t^2-1)(2w-1)}{2t},$$

$$L_{13}=i\left(\frac{[w(t^2+1)-1](2w-1)}{2t}\frac{\partial}{\partial w}+\frac{t^2+1}{2}\frac{\partial}{\partial t}-k\frac{w(t^2+1)-1}{t}\right.$$

$$\left.+\sigma\frac{(t^2+1)(2w-1)}{2t}\right)$$

which satisfy the commutation relations

$$[L_{\mu\nu},\ L_{\rho\lambda}]=i(g_{\nu\rho}L_{\mu\lambda}+g_{\mu\lambda}L_{\nu\rho}-g_{\mu\rho}L_{\nu\lambda}-g_{\nu\lambda}L_{\mu\rho}),$$
$$g_{11}=g_{22}=g_{33}=-g_{44}=1. \tag{1.2}$$

Therefore the above operators form the Lie algebra D_2 in the notation of Cartan. If we write

$$L_{12}=M_3,\ L_{23}=M_1,\ L_{31}=M_2,\ L_{41}=N_1,\ L_{42}=N_2,\ L_{43}=N_3,$$

the Casimir operators are

$$\mathbf{M}^2-\mathbf{N}^2=\sigma^2+(k+1)^2-1,\quad \mathbf{M}\cdot\mathbf{N}=-i\sigma(k+1),$$

i.e., the letters σ and k determine the eigenvalues of the Casimir operators, and therefore characterize the irreducible representations of the group generated by the operators $L_{\mu\nu}$.

We defined also in Ref. 1 the operators J_3, J^+, and $H^\pm$ by

$$L_{12}=J_3,\quad L_{14}=-\frac{i}{2}(J^++J^-),$$
$$L_{24}=\tfrac{1}{2}(J^+-J^-),\quad H^\pm=L_{23}\pm iL_{23}. \tag{1.3}$$

Equations (1.1) and (1.3) give

$$J_3=t\frac{\partial}{\partial t},$$

$$J^+=-t\left(w\frac{\partial}{\partial w}-t\frac{\partial}{\partial t}+\sigma\right),$$

$$J^-=-\frac{1}{t}\left((w-1)\frac{\partial}{\partial w}+t\frac{\partial}{\partial t}+\sigma\right) \tag{1.4}$$

which satisfy the commutation relations

$$[J_3,\ J^\pm]=\pm J^\pm,\quad [J^+,\ J^-]=2J_3. \tag{1.5}$$

Therefore the operators J_3, $J^\pm$ form the Lie algebra $sl(2)$.

It can be shown[1] that the operators $L_{\mu\nu}$ of Eqs. (1.1) transform among themselves the functions $|u,m\rangle = F(\sigma-u,\ \sigma+u+1;\ \sigma-m+1;w)\ t^m$, where $F(\sigma-u,\ \sigma+u+1;\ \sigma-m+1;w)$ is the well-known Gauss hypergeometric function. For convenience we introduce the functions $|u,m\rangle'$,

$$\left|u,m\right\rangle'=\frac{\Gamma(m-\sigma)}{\Gamma(m-u)}F(\sigma-u,\sigma+u+1;\sigma-m+1;w)t^m, \tag{1.6}$$

for which we get

$$L_{12}|u,m\rangle' = m|u,m\rangle',$$

$$J^+|u,m\rangle' = (m-u)|u,m+1\rangle',$$

$$J^-|u,m\rangle' = -(m+u)|u,m-1\rangle',$$

$$H^+\left|u,m\right\rangle' = \frac{(u-m)\sigma(k+1)}{u(u+1)}\left|u,m+1\right\rangle' - \frac{(u+\sigma+1)(u-k)}{(u+1)(2u+1)} \times \left|u+1,m+1\right\rangle'$$

$$- \frac{(u-\sigma)(u+k+1)(u-m)(u-m-1)}{u(2u+1)} \times \left|u-1,m+1\right\rangle',$$

$$H^-\left|u,m\right\rangle' = \frac{(u+m)\sigma(k+1)}{u(u+1)}\left|u,m-1\right\rangle' + \frac{(u+\sigma+1)(u-k)}{(u+1)(2u+1)}\left|u+1,m-1\right\rangle' \tag{1.7}$$

$$+ \frac{(u-\sigma)(u+k+1)(u+m)(u+m-1)}{u(2u+1)}\left|u-1,m-1\right\rangle',$$

$$L_{34}\left|u,m\right\rangle' = i\left(\frac{m\sigma(k+1)}{u(u+1)}\left|u,m\right\rangle' + \frac{(u+\sigma+1)(u-k)}{(u+1)(2u+1)}\left|u+1,m\right\rangle' + \frac{(u-\sigma)(u+k+1)(m^2-u^2)}{u(2u+1)}\left|u-1,m\right\rangle'\right).$$

The states $|u,m\rangle'$ will be used in the next section.

2. GENERATING FUNCTIONS

Let us consider the multiplier representation induced by the operators J_3, J^+, and J^- of Eqs. (1.4). We want to calculate the expression

$$e^{b'J^+}e^{c'J^-}e^{\tau'J_3}f(w,t). \tag{2.1}$$

The complex parameters b', c', and τ' are related to the representation g of $SL(2,c)$,

$$g=\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad-bc=1, \tag{2.2}$$

by[11]

$$\exp(\tau'/2)=d^{-1}, \quad b'=-b/d, \quad c'=-cd. \tag{2.3}$$

The action of the group element $e^{\tau'J_3}$ is obtained by solving the differential equations

$$\frac{dt(\tau')}{d\tau'}=t(\tau'), \quad t(0)=t, \tag{2.4}$$

giving

$$t(\tau')=te^{\tau'}=t/d^2\,. \tag{2.5}$$

The action of the group element $e^{b'J^+}$ is obtained by solving the differential equations[2]

$$\frac{dw(b')}{db'}=-t(b')w(b'), \quad \frac{dt(b')}{db'}=t^2(b'), \quad \frac{d\nu(b')}{db'} = -\sigma t(b')\nu(b'), \quad w(0)=w, \quad t(0)=t, \quad \nu(0)=1 \tag{2.6}$$

The solutions of the above equations are

$$t(b')=\frac{t}{1-b't}=\frac{dt}{d+bt}\,, \quad w(b')=(1-b't)w=\frac{d+bt}{d}w,$$

$$\nu(b')=(1-b't)^\sigma=\left(\frac{d+bt}{d}\right)^\sigma, \tag{2.7}$$

where Eqs. (2.3) have been used. Finally, the action of the element $e^{c'J^-}$ is found from the equations

$$\frac{dw(c')}{dc'}=-\frac{w(c')-1}{t(c')}\,, \quad \frac{dt(c')}{dc'}=-1, \quad \frac{d\nu(c')}{dc'}=-\frac{\sigma}{t(c')}\nu(c'),$$

$$w(0)=w, \quad t(0)=t, \quad \nu(0)=1. \tag{2.8}$$

Solving the above equations, we get

$$t(c')=t-c'=t+cd, \quad w(c')=\frac{w(t-c')+c'}{t}=\frac{w(t+cd)-cd}{t},$$

$$\nu(c')=\left(1-\frac{c'}{t}\right)^\sigma=\left(1+\frac{cd}{t}\right)^\sigma. \tag{2.9}$$

Therefore, we get

$$[T(g)f](w,t)=e^{b'J^+}e^{c'J^-}e^{\tau'J_3}f(w,t)=(d+bt)^\sigma\left(a+\frac{c}{t}\right)^\sigma \times f\left((d+bt)[(a+c/t)(w-1)+a], \frac{a+c/t}{d+bt}\,t\right) \tag{2.10}$$

The operators J_3, J^+, and J^- form a basis of the complex Lie algebra $sl(2)$. Let $f_{m_0+n}(w,t)$, $n=0, \pm1, \pm2, \cdots$, be the basic vectors of the representation[12] $D(u,m_0)$ of $sl(2)$ defined for all complex u and m_0 such that $m_0 \pm u$ are not integers and $0 \leq \mathrm{Re}\ m_0 < 1$. In this representation the spectrum of the operator J_3 is m_0+n, $n=0, \pm1, \pm2, \cdots$. We have

$$[T(g)f_{m_0+n}](w,t)=\sum_{\lambda=-\infty}^{\infty} A_{\lambda n}(g)f_{m_0+\lambda}(w,t), \tag{2.11}$$

where the matrix elements $A_{\lambda n}(g)$ are given by[13]

$$A_{\lambda n}(g)=a^{u+m_0+\lambda}d^{u-m_0-n}c^{n-\lambda}\frac{\Gamma(u+m_0+n+1)F(-u-m_0-\lambda, -u+m_0+n; n-\lambda+1; bc/ad)}{\Gamma(u+m_0+\lambda+1)\Gamma(n-\lambda+1)}\,. \tag{2.12}$$

We identify $f_{m_0}(w,t)$ with the basic vector $|u,m_0\rangle'$ of Eq. (1.6). Then we get from Eqs. (2.10), (2.11), and (2.12) if we take $n=0$ and introduce new indeces m, μ, and ν by $\mu=\sigma-u$, $\nu=\sigma+u+1$, $m=\sigma-m_0+1$,

$$(d+bt)^{m-1}(a+c/t)^{\mu+\nu-m}\frac{\Gamma(1-m)}{\Gamma(\mu-m+1)}$$
$$\times F\{\mu,\nu;m;(d+bt)[a+c/t)(w-1)+a]\}t^{-m+(\mu+\nu+1)/2}$$
$$=\sum_{\lambda=-\infty}^{\infty}a^{\nu-m+\lambda}d^{m-\mu-1}c^{-\lambda}$$
$$\times\frac{\Gamma(\nu-m+1)\Gamma(\lambda-m+1)}{\Gamma(\nu-m+\lambda+1)\Gamma(-\lambda+1)\Gamma(\mu-m+\lambda+1)}$$
$$\times F(m-\nu-\lambda,\mu-m+1;-\lambda+1;bc/ad)$$
$$\times F(\mu,\nu;m-\lambda;w)t^{\lambda-m+(\mu+\nu+1)/2} \tag{2.13}$$

with the restrictions (coming from the fact that the power series and the hypergeometric series must converge)

$$\left|\frac{bt}{d}\right|<1,\quad\left|\frac{c}{at}\right|<1,\quad\left|(d+bt)\left[\left(a+c/t\right)w-\frac{c}{t}\right]\right|<1,\quad|w|<1. \tag{2.13'}$$

In Eq. (2.13) and in all equations which we shall derive from it μ, ν, and m are arbitrary complex numbers such that $\mu-m$, $\nu-m$, and m are not integers.[14] In Eq. (2.13) the terms corresponding to $\lambda=1,2,3,\cdots$ are well defined because of the relation

$$\lim_{c\to -n}\frac{\Gamma(a,b;c;z)}{\Gamma(c)}=\frac{a(a+1)\cdots(a+n)\,b\,(b+1)\cdots(b+n)}{(n+1)!}$$
$$\times z^{n+1}F(a+n+1,b+n+1;n+2;z),$$
$$n=0,1,2,\cdots. \tag{2.14}$$

If we define τ and y by

$$\tau=c/at,\quad y=bc/ad, \tag{2.15}$$

Eqs. (2.13) and (2.13′) give

$$F\left(\mu,\nu;m;\frac{(1+y\tau^{-1})[(1+\tau)(w-1)+1]}{1-y}\right)$$
$$\times(1+y\tau^{-1})^{m-1}(1+\tau)^{\mu+\nu-m}(1-y)^{-\mu}$$
$$=\sum_{\lambda=-\infty}^{\infty}\tau^{-\lambda}\frac{\Gamma(\nu-m+1)\Gamma(\lambda-m+1)\Gamma(\mu-m+1)}{\Gamma(\nu-m+\lambda+1)\Gamma(-\lambda+1)\Gamma(\mu-m+\lambda+1)\Gamma(1-m)}$$
$$\times F(m-\nu-\lambda,\mu-m+1;-\lambda+1;y)F(\mu,\nu;m-\lambda;w),$$
$$|y|<|\tau|<1,\quad|(1+y\tau^{-1})[(1+\tau)w-\tau]|<|1-y|,\quad|w|<1. \tag{2.16}$$

For $y=0$, Eq. (2.16) gives,

$$F[\mu,\nu;m;(1+\tau)w-\tau](1+\tau)^{\mu+\nu-m}$$
$$=\sum_{\lambda=0}^{\infty}\tau^{\lambda}\frac{\Gamma(\nu-m+1)\Gamma(-\lambda-m+1)}{\Gamma(\nu-m-\lambda+1)\Gamma(\lambda+1)}$$
$$\times\frac{\Gamma(\mu-m+1)}{\Gamma(\mu-m-\lambda+1)\Gamma(1-m)}F(\mu,\nu;m+\lambda;w),$$
$$|\tau|<1,\quad|(1+\tau)w-\tau|<1,\quad|w|<1. \tag{2.17}$$

If Re $(\mu-m+1)>0$, Eq. (2.17) can be written in the form

$$F[\mu,\nu;m;(1+\tau)w-\tau](1+\tau)^{\mu+\nu-m}$$
$$=\sum_{\lambda=0}^{\infty}\tau^{\lambda}\binom{\nu-m}{\lambda}F(-\lambda,-\mu;-\lambda-m+1;1)$$
$$\times F(\mu,\nu;m+\lambda;w),$$
$$|\tau|<1,\quad|(1+\tau)w-\tau|<1,\quad|w|<1 \tag{2.18}$$

Let us set $y=x\tau$ and then take the limit $\tau\to 0$. If we use Eq. (2.14) and the relation

$$\Gamma(z)\Gamma(1-z)=\pi/\sin\pi z, \tag{2.19}$$

we get from (2.16)

$$F[\mu,\nu;m;(1+x)w](1+x)^{m-1}$$
$$=\sum_{\lambda=0}^{\infty}(-x)^{\lambda}\binom{\lambda-m}{\lambda}F(\mu,\nu;m-\lambda;w)$$
$$=\sum_{\lambda=0}^{\infty}x^{\lambda}\binom{m-1}{\lambda}F(\mu,\nu;m-\lambda;w),$$
$$|x|<1,\quad|(1+x)w|<1,\quad|w|<1. \tag{2.20}$$

For $w'=(1+\tau)w$, (2.17) becomes

$$F(\mu,\nu;m;w'-\tau)(1+\tau)^{\nu-m}=\sum_{\lambda=0}^{\infty}\tau^{\lambda}\binom{\nu-m}{\lambda}$$
$$\times\frac{\Gamma(-\lambda-m+1)\Gamma(\mu-m+1)}{\Gamma(\mu-m-\lambda+1)\Gamma(1-m)}F\left(\mu,\nu;m+\lambda;\frac{w'}{1+\tau}\right)(1+\tau)^{-\mu},$$
$$|\tau|<1,\quad|w'-\tau|<1,\quad|w'|<|1+\tau|. \tag{2.21}$$

Using the relation[2]

$$F\left(\mu,\nu;m+\lambda;\frac{w'}{1+\tau}\right)(1+\tau)^{-\mu}$$
$$=\sum_{\rho=0}^{\infty}(-\tau)^{\rho}\binom{\mu+\rho-1}{\rho}F(\mu+\rho,\nu;m+\lambda;w'),$$
$$|\tau|<1,\ |w'|<|1+\tau|,\quad|w'|<1, \tag{2.22}$$

we get

$$F(\mu,\nu;m;w'-\tau)(1+\tau)^{\nu-m}=\sum_{\lambda=0}^{\infty}\sum_{\rho=0}^{\infty}(-1)^{\rho}\tau^{\rho+\lambda}\binom{\nu-m}{\lambda}\binom{\mu+\rho-1}{\rho}$$
$$\times\frac{\Gamma(-\lambda-m+1)\Gamma(\mu-m+1)}{\Gamma(\mu-m-\lambda+1)\Gamma(1-m)}$$
$$\times F(\mu+\rho,\nu;m+\lambda;w')$$
$$=\sum_{\lambda=0}^{\infty}\sum_{\rho=0}^{\infty}(-1)^{\rho}\tau^{\rho+\lambda}\binom{\nu-m}{\lambda}\binom{\mu+\rho-1}{\rho}$$
$$\times F(-\lambda,-\mu;-\lambda-m+1;1)$$
$$\times F(\mu+\rho,\nu;m+\lambda;w'),$$
$$|\tau|<1,\quad|w'-\tau|<1,\quad|w'|<|1+\tau|,\ |w'|<1. \tag{2.23}$$

The second equality above hold if Re $(\mu - m + 1) > 0$. Similarly, hypergeometric functions of the form $F[\mu, \nu; m; (1+\psi)z - \tau]$ and $F[\mu, \nu; m;\ z/(1+\psi) - \tau]$ can be expressed in terms of hypergeometric functions with argument z.

For $m = -u+n$, $n = 0, 1, 2, \cdots$, $2u \neq 0, 1, 2, \cdots$, the states $|u, -u+n\rangle$ of Eq. (1.6) are the basic states of a local multiplier representation of the type $\dagger u$ in Miller's notation,[2] induced by the operators J_3, $J^{\pm}$ of Eqs. (1.4). In this case we have

$$[T(g)f_{-u+n}](w,t) = \sum_{\lambda=0}^{\infty} B_{\lambda n}(g) f_{-u+\lambda}(w,t), \tag{2.24}$$

where[2]

$$B_{\lambda n}(g) = a^{\lambda} d^{2u-n} c^{n-\lambda} \frac{\Gamma(n+1) F(-\lambda, -2u+n; n-\lambda+1;\ bc/ad)}{\Gamma(\lambda+1)\Gamma(n-\lambda+1)},$$

$$\lambda, n \geq 0. \tag{2.25}$$

Proceeding as in the case of the representation $D(u, m_0)$, we get

$$F\left(\mu, \nu; \nu - n; \frac{(1+y\tau^{-1})[(1+\tau)(w-1)+1]}{1-y}\right)$$
$$\times (1+y\tau^{-1})^{\nu-n-1}(1+\tau)^{\mu+n}(1-y)^{-\mu}$$
$$= \sum_{\lambda=0}^{\infty} \tau^{n-\lambda} \frac{\Gamma(n+1)\Gamma(\lambda-\nu+1)\Gamma(n+\mu-\nu+1)}{\Gamma(\lambda+1)\Gamma(n-\lambda+1)\Gamma(\lambda+\mu-\nu+1)\Gamma(n-\nu+1)}$$
$$\times F(-\lambda, n+\mu-\nu+1; n-\lambda+1; y)$$
$$\times F(\mu, \nu; \nu-\lambda; w),$$
$$|y| < |\tau| < 1, \quad |(1+y\tau^{-1})[(1+\tau)w - \tau]| < |1-y|, \quad |w| < 1. \tag{2.26}$$

The condition $2u \neq 0, 1, 2, \cdots$ is now translated into $\nu - \mu \neq 1, 2, 3, \cdots$ and is imposed to (2.26) and to all equations derived from it. If we write $y = x\tau$ and take the limit $\tau \to 0$, we get from (2.26)

$$F[\mu, \nu; \nu-n; (1+x)w](1+x)^{\nu-n+1}$$
$$= \sum_{\lambda=n}^{\infty} (-x)^{\lambda-n} \binom{\lambda-\nu}{\lambda-n} F(\mu, \nu; \nu-\lambda; w)$$
$$= \sum_{\lambda=0}^{\infty} x^{\lambda} \binom{\nu-n-1}{\lambda} F(\mu, \nu; \nu-n-\lambda, w),$$
$$|x| < 1, \quad |(1+x)w| < 1, \quad |w| < 1. \tag{2.27}$$

For $y = 0$ (2.26) gives

$$F[\mu, \nu; \nu-n; (1+\tau)w - \tau](1+\tau)^{\mu+n}$$
$$= \sum_{\lambda=0}^{n} \tau^{n-\lambda} \binom{n}{\lambda} \frac{\Gamma(\mu+n-\nu+1)\Gamma(\lambda-\nu+1)}{\Gamma(\lambda+\mu-\nu+1)\Gamma(n-\nu+1)}$$
$$\times F(\mu, \nu; \nu-\lambda; w),$$
$$|\tau| < 1, \quad |(1+\tau)w - \tau| < 1, \quad |w| < 1, \tag{2.28}$$

which, if $\mathrm{Re}(\mu+n-\nu+1) > 0$, $\lambda - \nu + 1 \neq 0, -1, -2, \cdots$, can be written in the form

$$F[\mu, \nu; \nu-n; (1+\tau)w - \tau](1+\tau)^{\mu+n}$$
$$= \sum_{\lambda=0}^{n} \tau^{n-\lambda} \binom{n}{\lambda} F(\lambda-n, -\mu; \lambda-\nu+1; 1)$$
$$\times F(\mu, \nu; \nu-\lambda; w),$$
$$|\tau| < 1, \quad |(1+\tau)w - \tau| < 1, \quad |w| < 1. \tag{2.29}$$

We shall also consider the multiplier representation induced by the operator L_{34}, whose matrix elements are known. In the case of the physical Lorentz group the rotation in the $x_3 - x_4$ plane generated by this operator, corresponds to an accelaration, or boost, in the direction x_3. The action of the group element $e^{sL_{34}}$ is obtained from the solution of the equations

$$\frac{dt(s)}{ds} = i\,t(s),$$
$$\frac{dw(s)}{ds} = 2i[w(s) - 1]w(s),$$
$$\frac{d\nu(s)}{ds} = i[2(\sigma - k)w(s) + k - \sigma]\nu(s), \tag{2.30}$$
$$t(0) = t, \quad w(0) = w, \quad \nu(0) = 1.$$

Solving the above system of differential equations, we get

$$t(s) = t\,e^{is},$$
$$w(s) = w/[w - e^{2is}(w-1)],$$
$$\nu(s) = e^{i(\sigma-k)s}/[w - e^{2is}(w-1)]^{\sigma-k}. \tag{2.31}$$

The matrix elements of the boost operator in the direction x_3 corresponding to finite rotations, have been calculated in the canonical basis $|u, m; \sigma, k\rangle$. If we denote this operator by F_3, we get[7,8]

$$F_3 \left| u, m; \sigma, k \right\rangle$$
$$= i\left(\frac{(u^2-m^2)(u^2-\sigma^2)[u^2-(k+1)^2]}{u^2(4u^2-1)}\right)^{1/2}$$
$$\times \left| u-1, m; \sigma, k \right\rangle - \frac{im\sigma(k+1)}{u(u+1)} \left| u, m; \sigma, k \right\rangle$$
$$- i\left(\frac{[(u+1)^2-m^2][(u+1)^2-\sigma^2][(u+1)^2-(k+1)^2]}{(u+1)^2[4(u+1)^2-1]}\right)^{1/2}$$
$$\times \left| u+1, m; \sigma, k \right\rangle. \tag{2.32}$$

Comparing the above expression with the last of Eqs. (1.7), we get

$$F_3 = -L_{34}. \tag{2.33}$$

Let us define the operators $\tilde{L}_{\mu\nu}$ by

$$\tilde{L}_{\mu\nu} = -L_{\mu\nu}. \tag{2.34}$$

If m is the eigenvalue of L_{12}, the eigenvalue of $\tilde{L}_{12}$ is $-m$. The basic functions $|u, m; \sigma, k\rangle$ in the canonical basis corresponding to generators $\tilde{L}_{\mu\nu}$ are obtained from expressions (1·6) by the replacement $m \to -m$, apart

for a normalization, i.e.,

$$|u,m;\sigma,k\rangle \propto |u,-m\rangle'. \tag{2.35}$$

To find the normalization we make in the last of Eqs. (1.7), the substitutions $L_{34} \to -\tilde{L}_{34}$, and $m \to -m$ and compare with Eq. (2.32). In this way we find

$$\left|u,m;\sigma,k\right\rangle = \left(\frac{\Gamma(u+\sigma+1)\Gamma(u-k)\Gamma(u+m+1)}{\Gamma(u-\sigma+1)\Gamma(u+k+2)\Gamma(u-m+1)}(u+\tfrac{1}{2})\right)^{1/2}$$

$$\times \frac{\Gamma(-m-u)}{\Gamma(-m-\sigma)}\left|u,-m\right\rangle' \equiv N_{u,m}\left|u,-m\right\rangle'. \tag{2.36}$$

The matrix elements of the boost operator F_3, corresponding to finite rotations, i.e., the expressions[15]

$$\langle u',m';\sigma,k\,|\,e^{iaF_3}\,|u,m;\sigma,k\rangle \equiv D^{\sigma,k}_{um,u'm'}(a), \tag{2.37}$$

have been calculated before[3-6] in the case of unitary representations of the Lorentz group, because of their physical interest. The unitary and, therefore, infinite-dimensional representations of the Lorentz group are classified into the following two series[7,8]:

(a) The principal series for which we have in our notation:

$$\sigma = 0, \tfrac{1}{2}, 1, \cdots, k+1\text{: pure imaginary},$$

$$u = \sigma, \sigma+1, \sigma+2, \cdots, \quad m = -u, -u+1, \cdots, u-1, u. \tag{2.38}$$

(b) The supplementary series for which we have

$$\sigma = 0, \quad 0 \leq k+1 \leq 1,$$

$$u = 0, \tfrac{1}{2}, 1, \cdots, \quad m = -u, -u+1, \cdots, u-1, u. \tag{2.39}$$

In the case of the principal series the expression for $D^{\sigma,k}_{um,u'm'}(a)$ can be obtained, for example, from the expression $D^{mp}_{J\mu,J'\mu'}(\epsilon)$ of Ref. 3 by the replacements

$$J \to u, \quad J' \to u', \quad \mu \to m, \quad \mu' \to m', \quad \nu_0 \to \sigma, \quad i\rho/2 \to k+1. \tag{2.40}$$

We get for $\tau = \epsilon^{-4}$

$$D^{\sigma k}_{um,u'm'}(a) = (-1)^{u+u'}\frac{\delta_{m,m'}}{\Gamma(u+u'+2)}\Big((2u+1)(2u'+1)\Gamma(u-\sigma+1)\Gamma(u+\sigma+1)\Gamma(u-m+1)$$

$$\times\Gamma(u+m+1)\Gamma(u'-\sigma+1)\Gamma(u'+\sigma+1)\Gamma(u'-m+1)\Gamma(u'+m+1)\frac{\Gamma(u'+k+2)\Gamma(u-k)}{\Gamma(u'-k)\Gamma(u+k+2)}\Big)^{1/2}$$

$$\times\Big(\sum_{d,d'}(-1)^{d+d'}\frac{\Gamma(u+u'-d-d'-m-\sigma+1)}{\Gamma(d+1)\Gamma(d'+1)\Gamma(u-m-d+1)\Gamma(u'-m-d'+1)\Gamma(\sigma+m+d+1)\Gamma(\sigma+m+d'+1)}$$

$$\times\frac{\Gamma(d+d'+m+\sigma+1)}{\Gamma(u-\sigma-d+1)\Gamma(u'-\sigma-d'+1)}F(u-k,\,d+d'+m+\sigma+1;u+u'+2;1-\tau)\tau^{(2d+m+\sigma-k)/2}\Big), \tag{2.41}$$

where the relation

$$F(a,b;c;z) = (1-z)^{-b}F(c-a,b;c;\,z/(z-1)), \tag{2.42}$$

was used. The above expression coincides with the corresponding expression of Ström's paper,[5] for a proper choice of a phase.

From (2.37) we get

$$e^{-iaL_{34}}\left|u,m;\sigma,k\right\rangle = \sum_{u',m'} D^{\sigma k}_{um,u'm'}(a)\left|u',m';\sigma,k\right\rangle \tag{2.43}$$

or using Eq. (2.36)

$$e^{-iaL_{34}}\left|u,-m\right\rangle' = \sum_{u',m'} D^{\sigma k}_{um,u'm'}(a)\frac{N_{u',m'}}{N_{u,m}}\left|u',-m'\right\rangle'. \tag{2.44}$$

The left-hand side of the above equation is easily obtained from Eqs. (1.6) and (2.31). We get

$$e^{-iaL_{34}}\left|u,-m\right\rangle' = \frac{\Gamma(-m-\sigma)}{\Gamma(-m-u)}\frac{e^{(\sigma-k-m)a}}{[w-e^{2a}(w-1)]^{\sigma-k}}$$

$$\times F\Big(\sigma-u,\sigma+u+1;\sigma+m+1;\frac{w}{w-e^{2a}(w-1)}\Big)\,t^{-m}. \tag{2.45}$$

Therefore, for $\tau = e^{-2a}$, Eqs. (1.6), (2.36), (2.41), (2.44), and (2.45) give the generating function

$$F\Big(\sigma-u,\sigma+u+1;\sigma+m+1;\frac{\tau w}{1+(\tau-1)w}\Big)$$

$$\times[1+(\tau-1)w]^{k-\sigma} = \sum_{u'=\sigma}^{\infty}(-1)^{u+u'}(2u'+1)$$

$$\times F(\sigma-u',\sigma+u'+1;\sigma+m+1;w)\frac{\Gamma(u-m+1)\Gamma(u'+m+1)}{\Gamma(u+u'+2)}$$

$$\times\Big[\sum_{d,d'}(-1)^{d+d'}\binom{u-\sigma}{d}\binom{u'+\sigma}{u'-m-d'}$$

$$\times\binom{d+d'+\sigma+m}{d'}\binom{u+u'-\sigma-m-d-d'}{u-m-d}$$

$$\times F(u-k,d+d'+\sigma+m+1;u+u'+2;1-\tau)\tau^{d}\Big], \tag{2.46}$$

$$\left|\frac{\tau w}{1+(\tau-1)w}\right| < 1, \quad |w|<1, \quad |1-\tau|<1,$$

where the range of the parameters σ, k, u, and m are given in (2.38).

The matrix elements of the finite rotation e^{iaF_3} in the case of the supplementary series of unitary representations of the Lorentz group have been calculated in Ref. 6. In this case $\sigma = 0$ and the expression for the matrix elements are simpler that the corresponding expres-

sions of the previous case. The matrix elements in our notation are obtained from those of Ref. 6 by the replacements

$$l \to u, \quad l' \to u', \quad n \to m, \quad s \to -k-1. \tag{2.47}$$

Since $0 \leq s \leq 1$ we have

$$1 \leq -k \leq 2. \tag{2.48}$$

We get for the supplementary series

$$D^{k}_{um,u'm'}(a) = (-1)^{u+u'} \frac{\delta_{m,m'}}{\Gamma(u+u'+2)}$$

$$\times \left[(2u+1)(2u'+1) \left(\frac{\Gamma(u'+k+2)\Gamma(-k)}{\Gamma(k+2)\Gamma(u'-k)} (1-\delta_{u',0}) + \delta_{u',0} \right) \right.$$

$$\times \left(\frac{\Gamma(u-k)\Gamma(k+2)}{\Gamma(-k)\Gamma(u+k+2)} (1-\delta_{u,0}) + \delta_{u,0} \right)$$

$$\left. \times \frac{\Gamma(u+m+1)\Gamma(u'+m+1)}{\Gamma(u-m+1)\Gamma(u'-m+1)} \right]^{1/2}$$

$$\times \left[\sum_{d,d'} (-1)^{d+d'} \binom{u-m}{d} \binom{u'-m}{d'} \right.$$

$$\times \frac{\Gamma(u+1)\Gamma(u'+1)\Gamma(u+u'-m-d-d'+1)\Gamma(m+d+d'+1)}{\Gamma(u-d+1)\Gamma(u'-d'+1)\Gamma(m+d+1)\Gamma(m+d'+1)}$$

$$\left. \times F(u-k, m+d+d'+1; u+u'+2, 1-\tau)\tau^{(2d+m-k)/2} \right], \tag{2.49}$$

where again $\tau = e^{-2a}$.

From Eqs. (2.36), (2.44), (2.45), and (2.49), we get

$$F\left(-u, u+1; m+1; \frac{\tau w}{1+(\tau-1)w}\right)$$

$$\times [1+(\tau-1)w]^{k} = \sum_{u'=0}^{\infty} (-1)^{u+u'}(2u'+1)F(-u', u'+1; m+1; w)$$

$$\times \frac{\Gamma(u'+m+1)\Gamma(u-m+1)}{\Gamma(u+u'+2)} \left[\sum_{d,d'} (-1)^{d+d'} \binom{u}{d} \binom{u'}{u'-m-d'} \right.$$

$$\times \binom{d+d'+m}{d'} \binom{u+u'-m-d-d'}{u-m-d}$$

$$\left. \times F(u-k, d+d'+m+1; u+u'+2; 1-\tau)\tau^{d} \right],$$

$$\left| \frac{\tau w}{1+(\tau-1)w} \right| < 1, \quad |w| < 1, \quad |1-\tau| < 1, \tag{2.50}$$

where the range of the parameters k, u, and m are given in (2.39).

3. WEISNER'S METHOD

Relations involving the hypergeometric functions can be derived by applying Weisner's[9,10] method. The method in general is the following: If C is the Casimir operator of a group G with generators $L_{\mu\nu}$ and C_E its eigenvalue, we consider the simultaneous partial differential equations

$$(C - C_E)f = 0, \quad (\tfrac{1}{2}r_{\mu\nu}L_{\mu\nu} + r_0)f \equiv (Q + r_0)f = 0. \tag{3.1}$$

The coefficients $r_{\mu\nu}$ and r_0 are arbitrary constants. (Of course we cannot have $r_{\mu\nu} = 0$ for all μ and ν, $r_0 \neq 0$.) If S is an element of the group G, since $[C, S] = 0$ we get from the above equations

$$(C - C_E)(Sf) = 0, \quad (SQS^{-1} + r_0)(Sf) = 0. \tag{3.2}$$

Given Q the operators of the form SQS^{-1} constitute its conjugate class. Equations (3.2) imply that it is sufficient to consider only one operator from each class. Each solution Sf of Eqs. (3.2) is a linear combination with constant coefficients of the solutions of equation $(C - C_E)f = 0$.

In our case let us consider the operators J_3, J^+, J^-. We have

$$e^{\tau' J_3} J_{\pm} e^{-\tau' J_3} = e^{\pm\tau'} J_{\pm}, \quad e^{b' J_+} J_- e^{-b' J_+} = -b'^2 J_+ + 2b' J_3 + J_-,$$

$$e^{b' J_+} J_3 e^{-b' J_+} = -b' J_+ + J_3, \quad e^{c' J_-} J_+ e^{-c' J_-} = J_+ - c'^2 J_- - 2c' J_3, \tag{3.3}$$

$$e^{c' J_-} J_3 e^{-c' J_-} = c' J_- + J_3,$$

and we get

$$e^{b' J_+} e^{c' J_-} (\lambda_3 J_3 - m) e^{-c' J_-} e^{-b' J_+} = \lambda_3(-c' b'^2 - b') J_+ + \lambda_3 c' J_-$$
$$+ \lambda_3(2b'c' + 1) J_3 - m. \tag{3.4}$$

Therefore we find that $Q = r_1 J_3 + r_2 J_+ + r_3 J_- + r_0$ is a conjugate of

(i) $\lambda_3 J_3 - m$ if $r_1^2 + 4r_2 r_3 \neq 0$, (3.5)

(ii) $\lambda_+ J_+ - \beta$ (or $\lambda_- J_- - \gamma$) if $r_1^2 + 4r_2 r_3 = 0$. (3.6)

The ratios m/λ_3 and β/λ_+ (or γ/λ_-) are not important and we may choose

$$\lambda_3 = \lambda_+ = \lambda_- = 1. \tag{3.7}$$

We shall apply Weisner's method for case (ii), i.e., we shall consider the simultaneous partial differential equations

$$J^+ g(w,t) = -t\left(w\frac{\partial}{\partial w} - t\frac{\partial}{\partial t} + \sigma \right) g(w,t) = 0, \tag{3.8}$$

$$[C - u(u+1)]g(w,t) = \left(w(w-1)\frac{\partial^2}{\partial w^2} + t\frac{\partial^2}{\partial w \partial t} \right.$$

$$\left. + (\sigma+1)(2w-1)\frac{\partial}{\partial w} + \sigma(\sigma+1) - u(u+1) \right) g(w,t) = 0. \tag{3.9}$$

The solution of Eq. (3.8) is

$$g(w,t) = t^{\sigma} h(wt), \tag{3.10}$$

where h is an arbitrary function of wt. Substituting $g(w,t)$ of Eq. (3.10) into Eq. (3.9) the function h is calculated. In this way we find that the solutions of the simultaneous Eq. (3.8) and (3.9) are

$$g(w,t) = \begin{cases} w^{u-\sigma} t^{u}, \\ w^{-u-\sigma-1} t^{-u-1}. \end{cases} \tag{3.11}$$

We get for $c \neq 0$

$$T(g)\, w^{u-\sigma} t^{u} = \left(a + \frac{c}{t}\right)^{\sigma+u} \left[\left(a + \frac{c}{t}\right) w - \frac{c}{t} \right]^{u-\sigma} t^{u} = \left(1 + \frac{at}{c}\right)^{\sigma+u}$$

$$\times\left[\left(1+\frac{at}{c}\right)w-1\right]^{u-\sigma}t^{-u}c^{2u}. \tag{3.12}$$

Therefore,

$$\left(1+\frac{at}{c}\right)^{\sigma+u}\left[\left(1+\frac{at}{c}\right)w-1\right]^{u-\sigma}t^{-u}c^{2u}$$

$$=\sum_{\lambda=0}^{\infty}h_\lambda F(\sigma-u,\sigma+u+1;\sigma+u-\lambda+1;w)t^{\lambda-u},$$

$$\left|\frac{at}{c}\right|<1,\ \left|\left(1+\frac{at}{c}\right)w\right|<1,\ \ |w|<1, \tag{3.13}$$

which comes from the fact that the left-hand side must be a linear combination of functions of the form $F(\sigma-u,\sigma+u+1;\sigma-m+1;w)t^m$. It follows from the series expansion of the left-hand side of Eq. (3.13) that m must be of the form $m=\lambda-u$ where $\lambda=0,\ 1,\ 2,\cdots$. To calculate the coefficients h_λ we put $w=0$ in Eq. (3.13). We get then

$$(-1)^{u-\sigma}c^{2u}\left(1+\frac{at}{c}\right)^{\sigma+u}=\sum_{\lambda=0}^{\infty}h_\lambda t^\lambda. \tag{3.14}$$

Therefore,

$$h_\lambda=(-1)^{u-\sigma}c^{2u}\binom{\sigma+u}{\lambda}\left(\frac{a}{c}\right)^\lambda, \tag{3.15}$$

and, for $at/c=s$, Eq. (3.13) becomes

$$(1+s)^{\sigma+u}[1-(1+s)w]^{u-\sigma}$$

$$=\sum_{\lambda=0}^{\infty}\binom{\sigma+u}{\lambda}F(\sigma-u,\sigma+u+1;\sigma+u-\lambda+1;w)s^\lambda,$$

$$|s|<1,\ \ |(1+s)w|<1,\ \ |w|<1, \tag{3.16}$$

which can also be written in the form

$$[1-(1+s)w]^{-a}(1+s)^{b-1}$$

$$=\sum_{\lambda=0}^{\infty}\binom{b-1}{\lambda}F(a,b;b-\lambda;w)s^\lambda,$$

$$|s|<1,\ \ |(1+s)w|<1,\ \ |w|<1. \tag{3.17}$$

Therefore the expression $[1-(1+s)w]^{-a}(1+s)^{b-1}$ is a generating function of the hypergeometric functions of the form $F(a,b;b-\lambda;w)$. The solution $g(w,t)=w^{-u-\sigma-1}t^{-u-1}$ gives again Eq. (3.16) with u replaced by $-u-1$.

*Present address.

[1]E. Kyriakopoulos, J. Math. Phys. **15**, 172 (1974).
[2]W. Miller, *Lie Theory and Special Functions* (Academic, New York, 1968).
[3]Dao Vong Duc and Nguyen Van Hieu, Dokl. Akad. Nauk. SSSR **172**, 1281 (1967) [Soviet Phys. -Doklady **12**, 312 (1967)].
[4]Dao Vong Duc and Nguyen Van Hieu, Ann. Inst. H. Poincaré A **6**, 17 (1967).
[5]S. Ström, Ark. Fys. **33**, 465 (1967).
[6]S. Ström, Ark. Fys. **38**, 373 (1968).
[7]M.A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, London, 1964).
[8]I.M. Gel'fand, R.A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon—Macmillan, New York, 1963).
[9]L. Weisner, Pacific J. Math. **5**, 1033 (1955).
[10]L. Weisner, Can. J. Math. **11**, 141,148 (1959).
[11]Reference 2, p. 21.
[12]Reference 2, p. 43.
[13]Reference 2, p. 159.

[14]The first two conditions come from the fact that we are in the case of the representation $D(u,m_0)$.
[15]If the action of the matrix group g on the function $\varphi(x)$ is defined by the right displacement $[T^R(g)\varphi](x)=\varphi(xg)$ or by the left displacement $[T^L(g)\varphi](x)=\varphi(g^{-1}x)$, and if $\tilde{g}$ is the transposed of the matrix g, the matrix elements of unitary representations are related by $\langle a\,|\,T^L(g)\,|\,b\rangle=\langle b\,|\,T^R(\tilde{g})\,|\,\alpha\rangle^*$.

Optical equivalence theorem for unbounded observables

J. R. Klauder

Bell Laboratories, Murray Hill, New Jersey 07974

L. Streit

Fakultät für Physik, Universität Bielefeld, 48 Bielefeld, Germany

The optical equivalence theorem relating c-number and q-number formulations of quantum optics is rigorously extended to cover various unbounded operators, and in particular those operators that directly yield counting rates.

1. INTRODUCTION

The optical equivalence theorem has enjoyed a wide application in quantum optics since its introduction by Sudarshan[1] and subsequent development.[2] This theorem uses the "diagonal" coherent state representation for the density operator ρ, which, for a single degree of freedom system, may be formally expressed as

$$\rho = \pi^{-1} \int \varphi(z) |z\rangle\langle z| d^2z , \tag{1}$$

where $d^2z \equiv d\mathrm{Re}z\, d\mathrm{Im}z$ and the integration is over the entire complex plane. The coherent state $|z\rangle = \exp(a^\dagger z - z^* a)|0\rangle$, where a and $a^\dagger$ are conventional annihilation and creation operators with $a|z\rangle = z|z\rangle$. Although $\varphi(z)$ is a generalized function for a general density matrix ρ we may always find an approximating density matrix for which $\varphi(z) \in \mathcal{S}(R^2)$.[2] More specifically, for a general ρ a sequence ρ_M exists, with the representation (1) and $\varphi_M(z) \in \mathcal{S}(R^2)$, such that $\|\rho - \rho_M\|_1 \to 0$, i.e., $\rho_M \to \rho$ in trace class norm. Consequently, for an arbitrary bounded operator B,

$$\langle B\rangle \equiv \mathrm{Tr}(B\rho) = \lim_{M\to\infty} \mathrm{Tr}(B\rho_M)$$
$$= \lim_{M\to\infty} \pi^{-1} \int \varphi_M(z)\langle z|B|z\rangle d^2z ,$$

where convergence is ensured since

$$|\mathrm{Tr}(B(\rho - \rho_M))| \le \|B\|\, \|\rho - \rho_M\|_1 \to 0 .$$

Indeed convergence is uniform for that class of B with operator norm $\|B\| \le c < \infty$, c fixed.

Although this form of the optical equivalence theorem covers a number of important cases (see Ref. 2, pp. 190–192), there has been some concern[3] since it does not deal with certain unbounded operators, especially those that determine photon counting rates. For a single normal mode the counting rates are given, up to an inessential factor, by

$$\mathrm{Tr}(a^{\dagger m} a^m \rho), \tag{2}$$

which in a formal sense may be expressed as

$$\pi^{-1} \int \varphi(z) |z|^{2m} d^2z .$$

While it is physically plausible that such moments exist, it is by no means true that every density matrix ρ has finite counting rates, and thus any proposed extension of the optical equivalence theorem to unbounded operators cannot apply to an arbitrary ρ. A convenient and physically reasonable subclass of density matrices are those for which $e^{\beta N}\rho$ is in trace class for some $\beta > 0$, where $N = a^\dagger a$ is the number operator. For each such density matrix the moments (2) exist and we seek to calculate them as the limit of a sequence of means based on density operators ρ_M admitting a diagonal representation with $\varphi_M \in \mathcal{S}(R^2)$. Specifically we have proved the following:

Theorem I: Let ρ be a density operator such that $e^{\beta N}\rho$ is in trace class for some $\beta > 0$. Then there is a sequence of functions

$$\varphi_M \in \mathcal{S}(R^2)$$

such that for the trace class operators

$$\rho_M \equiv \frac{1}{\pi} \int \varphi_M(z) |z\rangle\langle z| d^2z$$

the following properties hold:

(i) $\|\rho - \rho_M\|_1 \to 0$ as $M \to \infty$;

(ii) for all operators T for which $Te^{-\beta N}$ and $e^{-\beta N/2} T e^{-\beta N/2}$ are both bounded

$$\mathrm{Tr}(T\rho_M) \xrightarrow[M\to\infty]{} \mathrm{Tr}(T\rho)$$

uniformly on any subset of T for which

$$\|e^{-\beta N/2} T e^{-\beta N/2}\| \le c < \infty, \quad c \text{ fixed};$$

(iii) the approximate density operator ρ_M gives rise to the representation

$$\mathrm{Tr}(T\rho_M) = \frac{1}{\pi} \int \varphi_M(z)\langle z|T|z\rangle d^2z .$$

Application to counting rates

Suppose we deal with a density matrix ρ that fulfills the condition that $e^{\beta N}\rho$ is in trace class for some $\beta > 0$. For

$$T = a^{\dagger m} a^m = N(N-1)(N-2)\cdots(N-[m-1])$$

the two conditions on T are the same, and

$$\|Te^{-\beta N}\| = \max_n (n!/(n-m)!) e^{-\beta n},$$

which is finite for all m. Thus Theorem I assures us that a sequence of functions $\varphi_M(z) \in \mathcal{S}(R^2)$ exists such that, for all m,

$$\mathrm{Tr}(a^{\dagger m} a^m \rho) = \lim_{M\to\infty} \pi^{-1} \int \varphi_M(z) |z|^{2m} d^2z ,$$

which is the desired result.

2. PROOF OF THEOREM

We note first that $e^{\beta N}\rho \in \mathcal{B}_1$ ($\equiv$ trace class operators) for ρ a density operator implies that

$$\rho_\beta \equiv e^{\beta N/2} \rho e^{\beta N/2} \in \mathcal{B}_1 , \tag{3}$$

a fact we establish below in Lemma 2. As such we may

approximate ρ_β in trace class norm by trace class operators

$$\rho_\beta^{(M)} \equiv \frac{1}{\pi}\int d^2z\,\varphi_\beta^{(M)}(z)\,|z\rangle\langle z| \tag{4}$$

where $\varphi_\beta^{(M)} \in \mathcal{S}(R^2)$ for all M.[2] We next introduce a support-controlling function $S(x)$ which is *unity* for $x \leq 0$, *vanishes* for $x \geq 1$ and falls smoothly in between such as

$$S(x) = \exp[-(1-x)^{-2}e^{-x^{-2}}], \quad 0 \leq x \leq 1.$$

We next define weight functions of compact support by

$$\varphi_{\beta,M}(z) \equiv S(|z| - r_M)\varphi_\beta^{(M)}(z) \in \mathcal{S}(R^2), \tag{5}$$

and associated density operators

$$\rho_{\beta,M} \equiv \frac{1}{\pi}\int \varphi_{\beta,M}(z)\,|z\rangle\langle z|\,d^2z, \tag{6}$$

choosing $r_M < \infty$ so that (cf. Lemma 1 below)

$$\|\rho_{\beta,M} - \rho_\beta^{(M)}\|_1 \leq \|\rho_\beta^{(M)} - \rho_\beta\|_1. \tag{7}$$

Finally, we introduce

$$\rho_M \equiv e^{-\beta N/2}\rho_{\beta,M}e^{-\beta N/2} \tag{8}$$

which evidently is again in trace class. Making use of the basic fact that

$$e^{-\beta N/2}|z'\rangle = e^{-\frac{1}{2}(1-e^{-\beta})|z'|^2}|e^{-\beta/2}z'\rangle \tag{9}$$

and a change of variables $e^{-\beta/2}z' \to z$, readily leads to the diagonal representation

$$\rho_M = \frac{1}{\pi}\int \varphi_M(z)\,|z\rangle\langle z|\,d^2z \tag{10}$$

where

$$\varphi_M(z) = e^{\beta}e^{-(e^{\beta}-1)|z|^2}\varphi_{\beta,M}(e^{\beta/2}z) \in \mathcal{S}(R^2).$$

Indeed this function has compact support within the circle $|z| \leq e^{-\beta/2}(r_M+1) < \infty$.

The operators ρ_M constructed above constitute the desired approximating sequence and we now turn to establishing the properties listed in the statement of the Theorem. For property (i) we let $\epsilon > 0$ and choose M such that

$$\|\rho_\beta^{(M)} - \rho_\beta\|_1 < \tfrac{1}{2}\epsilon.$$

Then, in addition, from (7),

$$\|\rho_\beta^{(M)} - \rho_{\beta,M}\|_1 < \tfrac{1}{2}\epsilon,$$

so that we have the estimate

$$\begin{aligned}\|\rho_M - \rho\|_1 &= \|e^{-\beta N/2}(\rho_{\beta,M} - \rho_\beta)e^{-\beta N/2}\|_1 \\ &\leq \|e^{-\beta N/2}\|^2\|\rho_{\beta,M} - \rho_\beta\|_1 \\ &\leq \|\rho_{\beta,M} - \rho_\beta^{(M)}\|_1 + \|\rho_\beta^{(M)} - \rho_\beta\|_1 \\ &< \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon.\end{aligned}$$

Since ϵ was arbitrary, $\rho_M \to \rho$ in trace class.

For property (ii) we first note that $e^{\alpha N}\rho_M$ is a bounded operator for any complex α since $\varphi_M(z)$ has compact support. Because

$$e^{\alpha N}\rho_M = e^{-N}e^{(\alpha+1)N}\rho_M,$$

it follows that $e^{\alpha N}\rho_M$ is in trace class being the product of a trace class operator and a bounded operator. Thus we are assured that

$$T\rho_M = (Te^{-\beta N})(e^{\beta N}\rho_M)$$

and

$$T\rho = (Te^{-\beta N})(e^{\beta N}\rho)$$

are in trace class using the fact that trace class operators form a two-sided ideal of the algebra of all bounded operators.[4]

The trace of such operators may be computed in an arbitrary basis, such as the number operator eigenstate $|n\rangle$. For $T\rho_M$ we find

$$\begin{aligned}\mathrm{Tr}(T\rho_M) &= \sum_{n=0}^{\infty}\langle n|T\rho_M|n\rangle \\ &= \sum_{n=0}^{\infty}\langle n|e^{-\beta N/2}Te^{-\beta N/2}e^{\beta N/2}\rho_M e^{\beta N/2}|n\rangle \\ &= \mathrm{Tr}(e^{-\beta N/2}Te^{-\beta N/2}\rho_{\beta,M}),\end{aligned}$$

and similarly with the index M omitted. As a result

$$\begin{aligned}|\mathrm{Tr}(T\rho_M) - \mathrm{Tr}(T\rho)| &= |\mathrm{Tr}(e^{-\beta N/2}Te^{-\beta N/2}(\rho_{\beta,M} - \rho_\beta))| \\ &\leq \|e^{-\beta N/2}Te^{-\beta N/2}\|\,\|\rho_{\beta,M} - \rho_\beta\|_1\end{aligned}$$

which goes to zero as $M \to \infty$ uniformly for all T with $\|e^{-\beta N/2}Te^{-\beta N/2}\| \leq c < \infty$.

For (iii) we observe that

$$\mathrm{Tr}(T\rho_M) = \mathrm{Tr}(e^{-\beta N/2}Te^{-\beta N/2}\rho_{\beta,M}),$$

while the diagonal representation (6) for $\rho_{\beta,M}$ leads to[2]

$$\begin{aligned}&\mathrm{Tr}(e^{-\beta N/2}Te^{-\beta N/2}\rho_{\beta,M}) \\ &\quad = \frac{1}{\pi}\int d^2z'\varphi_{\beta,M}(z')\langle z'|e^{-\beta N/2}Te^{-\beta N/2}|z'\rangle.\end{aligned}$$

Use of (9), coupled with the same change of variables, establishes that

$$\mathrm{Tr}(T\rho_M) = \frac{1}{\pi}\int \varphi_M(z)\langle z|T|z\rangle d^2z,$$

completing the proof of the Theorem.

Lemmas and additional remarks

In order to justify the statement in (7) we appeal to

Lemma 1: Let $\varphi(z) \in \mathcal{S}(R^2)$ and

$$\varphi_R(z) \equiv S(|z| - R)\varphi(z) \in \mathcal{S}(R^2).$$

Then

$$A_R \equiv \frac{1}{\pi}\int \varphi_R(z)\,|z\rangle\langle z|\,d^2z$$

converges to A_∞ in trace class norm.

Proof: Consider the expression

$$\Delta A \equiv A_\infty - A_R = \frac{1}{\pi}\int \varphi_R'(z)\,|z\rangle\langle z|\,d^2z,$$

where the integrand

$$\varphi_R'(z) \equiv [1 - S(|z| - R)]\varphi(z) \in \mathcal{S}(R^2)$$

vanishes for $|z| \leq R$. The operator $\Delta A \in \mathcal{B}_1$ and fulfills the condition

$$|\langle\psi|\Delta A|\psi\rangle| \leq \langle\psi|\delta A|\psi\rangle$$

for all $|\psi\rangle$, where δA is nonnegative and given by

$$\delta A \equiv \frac{1}{\pi}\int |\varphi_R'(z)|\,|z\rangle\langle z|\,d^2z\,.$$

The positive and negative, Hermitian and skew-Hermitian parts of ΔA may each be so bounded as well leading to

$$\begin{aligned}\|\Delta A\|_1 &\leq 4\|\delta A\|_1 \\ &= \frac{4}{\pi}\sum_n \int |\varphi_R'(z)|\,|\langle z|n\rangle|^2 d^2z \\ &= \frac{4}{\pi}\int_{|z|\geq R} |\varphi_R'(z)|\,d^2z\end{aligned}$$

which evidently vanishes when $R \to \infty$, as desired.

The required property given in (3) is established by

Lemma 2: Let ρ be a density operator such that $e^{\beta N}\rho \in \mathcal{B}_1$ for some $\beta > 0$. Then

$$e^{\beta N/2}\rho e^{\beta N/2} \in \mathcal{B}_1\,.$$

Proof: We may write

$$\rho = \sum_{k=0}^{\infty} a_k |\psi_k\rangle\langle\psi_k|\,,$$

with

$$a_k \geq 0\,, \quad \sum a_k = 1$$

and $\{|\psi_k\rangle\}$ an orthonormal basis. Let $I \equiv \{i : a_i \neq 0\}$. Then for all $i \in I$

$$|\psi_i\rangle \in R(\rho) \subset D(e^{\beta N}) \subset D(e^{\beta N/2})\,.$$

To evaluate the trace of $e^{\beta N}\rho \in \mathcal{B}_1$ we may use any basis,[4] e.g., the basis $\{|\psi_k\rangle\}$. Then

$$\begin{aligned}\mathrm{Tr}(e^{\beta N}\rho) &= \sum_{k=0}^{\infty}\sum_{i\in I} a_i\langle\psi_k|e^{\beta N}|\psi_i\rangle\langle\psi_i|\psi_k\rangle \\ &= \sum_{i\in I} a_i\langle\psi_i|e^{\beta N}|\psi_i\rangle\,,\end{aligned}$$

which clearly converges absolutely. Let

$$\rho_n \equiv \sum_{\substack{i\in I\\ i\leq n}} a_i|\psi_i\rangle\langle\psi_i|$$

and $|\chi\rangle \in D(e^{\beta N/2})$. Then

$$\begin{aligned}e^{\beta N/2}\rho_n e^{\beta N/2}|\chi\rangle &= e^{\beta N/2}\sum_{\substack{i\in I\\ i\leq n}} a_i\langle\psi_i|e^{\beta N/2}|\chi\rangle|\psi_i\rangle \\ &= \sum_{\substack{i\in I\\ i\leq n}} a_i\langle\varphi_i|\chi\rangle|\varphi_i\rangle\,,\end{aligned}$$

where $|\varphi_i\rangle \equiv e^{\beta N/2}|\psi_i\rangle$. Consequently,

$$\rho_n^{\beta} \equiv e^{\beta N/2}\rho_n e^{\beta N/2} \subset \sum_{\substack{i\in I\\ i\leq n}} a_i|\varphi_i\rangle\langle\varphi_i| \in \mathcal{B}_1$$

being an operator of finite rank. These operators and their differences are semidefinite so trace norms equal traces.[4] For $m > n > n_0(\epsilon)$,

$$\begin{aligned}\|\rho_m^{\beta} - \rho_n^{\beta}\|_1 &= \mathrm{Tr}(\rho_m^{\beta} - \rho_n^{\beta}) \\ &= \sum_{\substack{i\in I\\ n<i\leq m}} a_i\langle\varphi_i|\varphi_i\rangle \\ &= \sum_{\substack{i\in I\\ n<i\leq m}} a_i\langle\psi_i|e^{\beta N}|\psi_i\rangle < \epsilon\end{aligned}$$

because of the absolute convergence of $\mathrm{Tr}(e^{\beta N}\rho)$.[4] We conclude that there exists

$$\lim_{n\to\infty}\rho_n^{\beta} = \lim_{n\to\infty} e^{\beta N/2}\rho_n e^{\beta N/2} = \sum_{i\in I} a_i|\varphi_i\rangle\langle\varphi_i| \in \mathcal{B}_1\,.$$

Does this limit agree with $e^{\beta N/2}\rho e^{\beta N/2}$? Let $|\chi\rangle \in D(e^{\beta N/2})$ and introduce

$$|\theta_n\rangle \equiv \rho_n e^{\beta N/2}|\chi\rangle \in D(e^{\beta N/2})\,.$$

It follows that

$$\text{s-lim}\,|\theta_n\rangle = \rho e^{\beta N/2}|\chi\rangle \equiv |\theta\rangle\,,$$

since we can estimate

$$\begin{aligned}\|\,|\theta_n\rangle - |\theta\rangle\,\| &= \|(\rho_n - \rho)e^{\beta N/2}|\chi\rangle\,\| \\ &\leq \|\rho_n - \rho\|_1\,\|e^{\beta N/2}|\chi\rangle\,\|\,.\end{aligned}$$

By a similar estimate it follows that

$$\text{s-lim}\,e^{\beta N/2}|\theta_n\rangle = \Big(\sum_{i\in I} a_i|\varphi_i\rangle\langle\varphi_i|\Big)|\chi\rangle\,.$$

Since $e^{\beta N/2}$ is a closed operator,

$$\begin{aligned}\text{s-lim}\,e^{\beta N/2}|\theta_n\rangle &= e^{\beta N/2}\,\text{s-lim}\,|\theta_n\rangle \\ &= e^{\beta N/2}|\theta\rangle \\ &= e^{\beta N/2}\rho e^{\beta N/2}|\chi\rangle\,.\end{aligned}$$

Consequently,

$$e^{\beta N/2}\rho e^{\beta N/2} \subset \sum_{i\in I} a_i|\varphi_i\rangle\langle\varphi_i| \in \mathcal{B}_1\,,$$

establishing the desired result.

Remark 1: The preceding proof makes no special use of the fact that N is the number operator. Consequently, it follows, for any density operator ρ, if $A^2\rho \in \mathcal{B}_1$ then $A\rho A \in \mathcal{B}_1$ for an arbitrary self-adjoint operator $A \geq 1$. Indeed it is a trivial extension to allow ρ to be a general element of $\mathcal{B}_1$, which in turn, under the stated conditions, implies that $\rho A^2 \in \mathcal{B}_1$.

Remark 2: While only one condition on ρ is necessary, the two conditions imposed on T are *not* redundant as may be seen in the following examples:

(a) Let

$$T = \sum_{n=0}^{\infty} |n\rangle\, e^{2\beta n}\langle 2n|\,.$$

Then

$$Te^{-\beta N} = \sum_{n=0}^{\infty} |n\rangle\langle 2n| \quad \text{(bounded)}\,,$$

$$e^{-\beta N/2}Te^{-\beta N/2} = \sum_{n=0}^{\infty} |n\rangle e^{\beta n/2}\langle 2n| \quad \text{(unbounded)}\,.$$

(b) Let

$$T = \sum_{n=0}^{\infty} |2n\rangle e^{3\beta n/2}\langle n|\,.$$

Then

$$Te^{-\beta N} = \sum_{n=0}^{\infty} |\, 2n\rangle e^{\beta n/2}\langle n| \quad \text{(unbounded)},$$

$$e^{-\beta N/2} T e^{-\beta N/2} = \sum_{n=0}^{\infty} |\, 2n\rangle\langle n| \quad \text{(bounded)}.$$

ACKNOWLEDGMENT

One of us (LS) would like to thank Professor Paul Urban for his hospitality at the Institut für Theoretische Physik, Universität Graz, where part of this work was done.

[1]E.C.G. Sudarshan, Phys. Rev. Lett. 10, 277 (1963).
[2]J.R. Klauder and E.C.G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968), Chaps. 7 and 8. We follow the notation of this text rather closely.
[3]R.J. Glauber, *Physics of Quantum Electronics*, edited by P.C. Kelley, B. Lax, and P.E. Tannewald (McGraw-Hill, New York, 1966), p. 788; *Quantum Optics*, Proceedings of the International School of Physics, "Enrico Fermi", Course XLII, edited by R.J. Glauber (Academic, New York, 1969), p. 15.
[4]Details regarding operator properties may be found, e.g., in F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Unger, New York, 1955); I.M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, translated by A. Feinstein (Academic, New York, 1964); K. Yosida, *Functional Analysis* (Springer-Verlag, New York, 1968).

Duality of observables and generators in classical and quantum mechnics*

Emile Grgin and Aage Petersen

Belfer Graduate School of Science, Yeshiva University, New York, New York 10033
(Received 28 June 1971; final revised manuscript received 26 December 1973)

In classical and in quantum mechanics physical variables play a dual role as observables and as generators of infinitesimal transformations in the invariance groups. We show that if the Lie algebra of generators is central simple, the observable–generator duality restricts the structure of the algebra of observables to two cases: a commutative, associative algebra as in classical mechanics, or a central simple special Jordan algebra as in quantum mechanics.

I. INTRODUCTION

It is a peculiar feature of both classical and quantum mechanics that the physical variables of these theories play a dual role as observables and as generators of infinitesimal transformations in the invariance groups. (For example, one and the same variable plays the role of an observable, called the energy, and of a generator, called the Hamiltonian). The fundamental significance of this pointwise identification of two sets of conceptually different objects manifests itself in the description of the measurement process.[1] In fact, the observable–generator duality is at the root of the Bohr–Heisenberg principle of equivalence between definability and measurability in physics, a principle which has played a fundamental role in the discussions of the foundations of quantum mechanics.[2]

In this paper we study the identification of observables and generators from an algebraic point of view.[3] Each of the sets of observables and generators is an algebra, and the observable–generator duality manifests itself as a map from the space of observables to the space of generators. This map interrelates the two algebras, imposing restrictions on their structures. Our purpose is to investigate these restrictions.

In Sec. II we construct an abstract algebraic structure that has the observable–generator duality as a fundamental property. We show that such a structure is a system $\{H, \tau, \alpha\}$, where H is a linear space over a field $\mathcal{F}$, equipped with two algebraic products, τ and α, such that the product α is a Lie product and its distribution law with respect to the product τ is the derivation rule. The duality map is the canonical projection of the Lie algebra $\{H, \alpha\}$ onto its algebra of inner derivations.

In Sec. III we investigate the important class of structures $\{H, \tau, \alpha\}$ for which the kernel of the duality map is minimal, i.e., $\{0\}$ or the field $\mathcal{F}$, and the image of this map is a central simple Lie algebra. Since objects of this type first appeared in physics in Hamilton's formulation of classical mechanics, we call such structures *Hamilton algebras*.

One sees that if $\{H, \tau, \alpha\}$ is a Hamilton algebra, and if $\tau = \sigma + \pi$, where the products σ and π are, respectively, the symmetric and antisymmetric parts of the product τ, then the structures $\{H, \sigma, \alpha\}$ and $\{H, \pi, \alpha\}$ are also Hamilton algebras.

We show that there exists no nondegenerate antisymmetric Hamilton algebra $\{H, \pi, \alpha\}$. We further prove that in symmetric Hamilton algebras $\{H, \sigma. \alpha\}$, there is a relation between the association properties of the products σ and α. Specifically, according to this *association relation* the associator with respect to the product σ is proportional to the corresponding associator with respect to the product α, the proportionality constant being an element of the field of scalars, $\mathcal{F}$. If this constant vanishes, we call the structure $\{H, \sigma, \alpha\}$ a *classical Hamilton algebra*. Thus, in a classical Hamilton algebra the substructure $\{H, \sigma\}$ is an associative and commutative algebra. If the proportionality constant in the association relation does not vanish, we call the structure $\{H, \sigma, \alpha\}$ a *quantum Hamilton algebra*. We show that in a quantum Hamilton algebra the substructure $\{H, \sigma\}$ is a central simple special Jordan algebra.

II. ALGEBRAS WITH DUALITY

In the standard formulation of classical mechanics the observables are real functions on a phase space and the product of observables is the commutative and associative product of these functions. In the standard formulation of quantum mechanics the observables are self-adjoint operators on a Hilbert space and the product of observables is the commutative, but not associative Hermitian product (i.e., the anticommutator) of these operators.

In addition to the commutative and associative product of functions, the space of classical physical variables is equipped with a second bilinear operator, the Poisson bracket $\{,\}$, which is a Lie product. For every classical physical variable $f(p,q)$, the symbol $\{f,\}$ represents a linear operator which acts on the algebra of observables and produces in that algebra an infinitesimal canonical transformation:

$$h(p,q) \to h(p,q) + \epsilon \cdot \{f,h\}(p,q),$$

where ϵ is an infinitesimal real parameter. Similarly, in addition to the Hermitian product, the space of self-adjoint operators of quantum mechanics is equipped with a second bilinear operation, the commutator $[\,,\,]/i\hbar$, which is a Lie product. For every quantum physical variable A, the symbol $[A,\,]/i\hbar$ represents a linear operator which acts on the algebra of observables and produces in that algebra an infinitesimal unitary transformation:

$$B \to B + \epsilon \circ [A,B]/i\hbar$$

Thus, in both theories, each variable plays a dual role as an observable and as a generator of a transformation belonging to the invariance group of the theory.

We shall now construct an abstract algebraic structure in which the observable–generator duality is incorporated. As a first step, we give a formal definition of this duality. Let $\{\mathcal{H}, \tau\}$ be an algebra, i.e., $\mathcal{H}$ is a linear space over a field $\mathcal{F}$, and τ is the symbol for the product in $\mathcal{H}$:

$$\tau: \mathcal{H} \otimes \mathcal{H} \to \mathcal{H}.$$

For our present purpose it is unnecessary to specify the field of scalars $\mathcal{F}$. Of course, the scalars of the algebras of observables in classical and quantum mechanics are the real numbers. Let $\mathcal{E}'$ denote the associative algebra of linear operators acting on the linear space $\mathcal{H}$, and let $\mathcal{E}'_\tau$ be the subset of $\mathcal{E}'$ consisting of the infinitesimal automorphisms of the algebra $\{\mathcal{H}, \tau\}$. Every element of the set $\mathcal{E}'_\tau$ has the form $I' + \epsilon D'$, where I' is the identity operator, and D' some linear operator, both belonging to $\mathcal{E}'$. The automorphism condition reads

$$(I' + \epsilon D')(f \tau g) = ((I' + \epsilon D')f)\, \tau\, ((I' + \epsilon D')g)$$

where $f, \ldots, \in \mathcal{H}$. This implies

$$D'(f \tau g) = (D'f) \tau g + f \tau (D'g),$$

i.e., the operator D' is a derivation in the algebra $\{\mathcal{H}, \tau\}$. The set $\mathcal{D}'$, $\mathcal{D}' \subset \mathcal{E}'$, of derivations in $\{\mathcal{H}, \tau\}$ is a Lie algebra $\{\mathcal{D}', [\,,\,]\}$ with the commutator as Lie product.

If we think of the linear space $\mathcal{H}$ as the space of physical variables, and of the algebra $\{\mathcal{H}, \tau\}$ as the algebra of observables, then the observable–generator duality is a requirement that every element of the linear space $\mathcal{H}$, i.e., every physical variable, in addition to being an element of the algebra $\{\mathcal{H}, \tau\}$, i.e., an observable, also be able to uniquely define an infinitesimal automorphism of the algebra $\{\mathcal{H}, \tau\}$. Thus, to every element of $\mathcal{H}$ there is associated a unique element of the derivation algebra $\{\mathcal{D}'[\,,\,]\}$, i.e., a generator of an infinitesimal automorphism. In other words, the observable–generator duality amounts to the existence of a map

$$\varphi': \mathcal{H} \to \mathcal{D}'.$$

The map φ', which shall be called the *duality map*, is to be a fundamental structure of the mathematical object we are constructing. This means that the other structures of the object must be compatible with it, and that the object must be so constructed that the duality map is "natural," i.e., that no choices have to be made in order to exhibit it.

To bring out the structural conditions which the duality map φ' imposes on the system consisting of the algebras $\{\mathcal{H}, \tau\}$ and $\{\mathcal{D}', [\,,\,]\}$, we separate the operator structure from the algebraic structure of the set $\mathcal{D}'$. Let $\mathcal{L}$ be a linear space isomorphic with $\mathcal{L}'$, and γ be a map

$$\gamma: \mathcal{L} \otimes \mathcal{H} \to \mathcal{H}$$

defined by the requirement that for every operator $U' \in \mathcal{E}'$, and the corresponding vector $U \in \mathcal{L}$, the identity $U'f \equiv U d l \gamma f \in \mathcal{H}$ should hold for every $f \in \mathcal{H}$. In short, $U' = U\gamma$. Further, let $\mathcal{D} \subset \mathcal{L}$ denote the abstract vector space which corresponds to the space $\mathcal{D}'$ of derivations. A Lie product μ is now defined in $\mathcal{D}$ by the requirement that the Lie algebras $\{\mathcal{D}, \mu\}$ and $\{\mathcal{D}', [\,,\,]\}$ be isomorphic. Specifically, the product μ in $\mathcal{D}$ is related to the commutator in $\mathcal{D}'$ by the identity

$$(U \mu V)\gamma f = [U\gamma, V\gamma] f = U\gamma(V\gamma f) - V\gamma(U\gamma f) \tag{1}$$

for all $U, V \in \mathcal{D}$ and $f \in \mathcal{H}$. We shall denote by φ the map from $\mathcal{H}$ into $\mathcal{D}$ corresponding to the duality map φ', and by $\mathcal{B}$ the image space of φ, i.e., $\mathcal{B} = \varphi(\mathcal{H})$. The elements of $\mathcal{B}$ will be denoted by the letters $F, G, H, \cdots$ where $F = \varphi(f)$ for $f \in \mathcal{H}$, etc.

Every element of $\mathcal{D}$ generates two infinitesimal automorphisms, one in the algebra $\{\mathcal{H}, \tau\}$ via the map γ, and one in the Lie algebra $\{\mathcal{D}, \mu\}$ itself via the product μ. Clearly, for the map φ to be compatible with the other structures, its image space $\mathcal{B}$ must be invariant with respect to the infinitesimal automorphisms generated by elements of $\mathcal{B}$ itself, i.e., $\mathcal{B}\mu\mathcal{B} \subseteq \mathcal{B}$. Hence, $\{\mathcal{B}, \mu\}$ is a Lie subalgebra of the derivation algebra $\{\mathcal{D}, \mu\}$.

The duality map φ induces a new algebraic product, α, in its domain space via the diagram

$$\begin{array}{ccc} \mathcal{H} \otimes \mathcal{H} & \overset{\alpha}{\searrow} & \\ \varphi \downarrow & \| & \mathcal{H} \\ \mathcal{B} \otimes \mathcal{H} & \overset{\gamma}{\nearrow} & \end{array}$$

which is equivalent to the identity

$$f \alpha g = F \gamma g. \tag{2}$$

Thus, the two linear operators on $\mathcal{H}$, $f\,\alpha$ and $F\gamma = \varphi(f)\gamma$ are identical:

$$f \alpha = F\gamma \tag{3}$$

The condition that the product α be preserved by the infinitesimal transformations in $\mathcal{H}$ generated by the operators $H\gamma$, $H \in \mathcal{B}$, reads

$$H\gamma(f \alpha g) = (H\gamma f)\alpha g + f \alpha (H\gamma g).$$

With identity (3), this relation becomes

$$h\alpha(f \alpha g) = (h \alpha f)\alpha g + f \alpha (h \alpha g) \tag{4}$$

From relation (3) one obtains, for the commutator, the expression

$$[H\gamma, F\gamma] g = [h\alpha, f\alpha] g = h\alpha(f\alpha g) - f\alpha(h\alpha g)$$

which, with Eq. (4), yields

$$[h\alpha, f\alpha] = (h\alpha f)\alpha.$$

The antisymmetry of the commutator implies the same property for the product α:

$$f\alpha h = -h\alpha f. \tag{5}$$

The antisymmetry condition (5) and the derivation rule (4) yield the Jacobi identity

$$f\alpha(g\alpha h) + g\alpha(h\alpha f) + h\alpha(f\alpha g) = 0. \tag{6}$$

Hence, the algebra $\{\mathcal{H}, \alpha\}$ is a Lie algebra.

The diagram shows that $(\mathrm{Ker}\ \varphi)\alpha\mathcal{H} = \{0\}$. Thus, $\mathrm{Ker}\varphi \subseteq \mathcal{C}$, where $\mathcal{C}$ is the center of the algebra $\{\mathcal{H}, \alpha\}$. This further implies the condition $\mathcal{B}\gamma\,\mathrm{Ker}\varphi = \{0\}$. One also sees that the image spaces of the linear maps γ and α coincide, i.e., $\mathrm{Im}\alpha = \mathrm{Im}\gamma$.

The duality map also induces a product, ν, in its image space $\mathcal{B}$ via the diagram

$$\begin{array}{ccc} \mathcal{B}\otimes\mathcal{H} & \xrightarrow{\gamma} & \mathcal{H} \\ \| & \downarrow\varphi & \downarrow\varphi \\ \mathcal{B}\otimes\mathcal{B} & \xrightarrow{\nu} & \mathcal{B} \end{array}$$

which is equivalent to the identity

$$\varphi(F\gamma g)=F\nu G. \tag{7}$$

This diagram is possible only if $\mathcal{B}\gamma\mathrm{Ker}\varphi\subseteq\mathrm{Ker}\varphi$, but since the previous diagram required $\mathcal{B}\gamma\mathrm{Ker}\varphi=\{0\}$, no new condition is being imposed. One sees that the image of the induced map ν coincides with the image space of the composite map $\varphi\circ\gamma$, i.e., $\mathrm{Im}\nu=\varphi(\mathrm{Im}\gamma)$.

Identities (2) and (7) yield the relation

$$\varphi(f\alpha g)=\varphi(F\gamma g)=F\nu G.$$

Thus, the diagram

$$\begin{array}{ccc} \mathcal{H}\otimes\mathcal{H} & \xrightarrow{\alpha} & \mathcal{H} \\ \varphi\downarrow\quad\varphi\downarrow & & \varphi\downarrow \\ \mathcal{B}\otimes\mathcal{B} & \xrightarrow{\nu} & \mathcal{B} \end{array}$$

is commutative, i.e., the algebra $\{\mathcal{B},\nu\}$ is a homomorphic image of the Lie algebra $\{\mathcal{H},\alpha\}$ under the duality map φ. Consequently, Ker φ is an ideal in $\{\mathcal{H},\alpha\}$ and $\{\mathcal{B},\nu\}$ is isomorphic to the Lie algebra $\{\mathcal{H},\alpha\}/\mathrm{Ker}\varphi$.

From the previous identities one obtains the relations

$$[f\alpha,g\alpha]=(f\alpha g)\alpha=(\varphi(f\alpha g))\gamma=(F\nu G)\gamma,$$

$$[f\alpha,g\alpha]=[F\gamma,G\gamma]=(F\mu G)\gamma,$$

which imply $F\nu G=F\mu G$ for all $F,\ G\in\mathcal{B}$, i.e., the products ν and μ are identical.

The set $\mathcal{H}\alpha$ of operators $f\alpha$ is the algebra of inner derivations of the Lie algebra $\{\mathcal{H},\alpha\}$. It is isomorphic to the quotient algebra $\{\mathcal{H},\alpha\}/\mathcal{C}$. Since, due to relation (3), all operators in the set $\mathcal{B}\gamma$ are inner derivations in $\{\mathcal{H},\alpha\}$, one has $\mathcal{B}\gamma\subseteq\mathcal{H}\alpha$. On the other hand, $\mathrm{Ker}\varphi\subseteq\mathcal{C}$ implies

$$\{\mathcal{H},\alpha\}/\mathcal{C}\subseteq\{\mathcal{H},\alpha\}/\mathrm{Ker}\varphi,$$

i.e., $\mathcal{H}\ \alpha$ is isomorphic with a subalgebra of $\{\mathcal{B},\mu\}$, so that $\mathcal{B}\gamma\supseteq\mathcal{H}\alpha$. Thus, $\mathrm{Ker}\varphi=\mathcal{C}$, i.e., the kernel of the duality map is the center of the algebra $\{\mathcal{H},\alpha\}$. In other words, the duality map is the canonical projection of the Lie algebra $\{\mathcal{H},\alpha\}$ onto its algebra of inner derivations.

In summary, the system consisting of an algebra $\{\mathcal{H},\tau\}$ a subalgebra $\{\mathcal{B}',[\ ,\ ]\}$ of the algebra of derivations in $\{\mathcal{H},\tau\}$, and a duality map of $\mathcal{H}$ onto $\mathcal{B}'$, is equivalent to an algebra $\{\mathcal{H},\tau,\alpha\}$ with two products such that the distribution law of the Lie product α with respect to the product τ is the derivation rule, and the Lie algebra $\mathcal{H}\alpha$ of inner derivations in $\{\mathcal{H},\alpha\}$ is isomorphic to the Lie algebra $\{\mathcal{B}',[\ ,\ ]\}$.

III. HAMILTON ALGEBRAS

Having shown that a structure with duality is equivalent to an algebraic system with two products $\{\mathcal{H},\tau,\alpha\}$, we now investigate the restrictions that are imposed on the product τ and the Lie product α by the conditions that they exist in a common underlying space and be related by the derivation rule. We do not study this question for general algebras $\{\mathcal{H},\tau,\alpha\}$, but restrict ourselves to the family of algebras $\{\mathcal{H},\tau,\alpha\}$ for which the kernel of the duality map is minimal, and its image is a central simple Lie algebra. We call such two-product systems Hamilton algebras.

Definition: *A Hamilton algebra* $\{\mathcal{H},\tau,\alpha\}$ is a linear space $\mathcal{H}$ over a field $\mathcal{J}$, equipped with two bilinear operations: a product τ and a Lie product α, satisfying the following conditions (where $f,\ g,\ h,\ \cdots\in\mathcal{H}$):

(1) Lie conditions:

$$f\alpha g=-g\alpha f \tag{8}$$

$$(f\alpha g)\alpha h+(g\alpha h)\alpha f+(h\alpha f)\alpha g=0. \tag{9}$$

(2) Derivation condition:

$$f\alpha(g\tau h)=(f\alpha g)\tau h+g\tau(f\alpha h). \tag{10}$$

(3) Minimality condition: The center $\mathcal{C}$ of the Lie algebra $\{\mathcal{H},\alpha\}$ is minimal, i.e., $\mathcal{C}=\mathcal{J}e$ if the algebra $\{\mathcal{H},\tau\}$ has a unit element e, and $\mathcal{C}=\{0\}$ if $\{\mathcal{H},\tau\}$ has no unit element.

(4) Simplicity condition: The Lie algebra $\mathcal{H}\alpha\simeq\{\mathcal{H},\alpha\}/\mathcal{C}$ is central simple. That is, (a) $\mathcal{H}\alpha$ is simple, i.e., if $\mathcal{G}$ is an ideal in $\mathcal{H}\alpha$, then either $\mathcal{G}=\{0\}$ or $\mathcal{G}=\mathcal{H}\alpha$; (b) any algebra obtained from $\mathcal{H}\ \alpha$ by extension of the base field $\mathcal{J}$ is simple.

This definition of a Hamilton algebra is purely algebraic. If $\mathcal{H}$ is a topological linear space and $\{\mathcal{H},\alpha\}$ a topological Lie algebra, the corresponding definition of a topological Hamilton algebra imposes itself, but we do not consider it in this paper.

If $\{\mathcal{H},\alpha\}$ is a central simple Lie algebra, the structure $\{\mathcal{H},r\alpha,\alpha\}$, where $r\in\mathcal{J}$, is a Hamilton algebra. Such a Hamilton algebra shall be called *degenerate*, or *trivial* if $r=0$.

Let $\{\mathcal{H},\tau,\alpha\}$ be a Hamilton algebra. Since the linear operators $f\alpha$, $f\in\mathcal{H}$, are derivations with respect to the product τ, they are also derivations with respect to the opposite product τ' defined by the identity $g\tau' h=h\tau g$, and hence, by linearity, they are derivations with respect to the symmetric and antisymmetric parts of τ defined by the relations

$$\sigma=(\tau+\tau')/2,\quad \pi=(\tau-\tau')/2.$$

In other words, if $\{\mathcal{H},\tau,\alpha\}$ is a Hamilton algebra, so are the structures $\{\mathcal{H},\sigma,\alpha\}$ and $\{\mathcal{H},\pi,\alpha\}$, which shall be called, respectively, *symmetric* and *antisymmetric* Hamilton algebras. Since an algebra $\{\mathcal{H},\pi\}$ with π antisymmetric has no unit, the algebras $\{\mathcal{H},\alpha\}$ and $\mathcal{H}\alpha$ are isomorphic in an antisymmetric Hamilton algebra.

In investigating the association properties of the product τ we shall need the concept of the associator. The *associator* $[f,g,h]$ of the product τ in $\mathcal{H}$ is a linear map

$$[\ ,\ ,\ ]\colon\ \mathcal{H}\otimes\mathcal{H}\otimes\mathcal{H}\to\mathcal{H}$$

defined by the identity

$$[f,g,h]=(f\tau g)\tau h-f\tau(g\tau h).$$

It "measures" the deviation from associativity of the product τ. One verifies that if the product τ is either symmetric, i.e., $\tau=\sigma$, or antisymmetric, $\tau=\pi$, the

associator is antisymmetric in its exterior variables, i.e.,

$$[f,g,h]=-[h,g,f].$$

Furthermore, if $\tau=\sigma$ or $\tau=\pi$, the associator satisfies the identity

$$[f,h\alpha g,g]+[h,f\alpha g,g]\equiv 0. \tag{11}$$

For $\tau=\sigma$, identity (11) is obtained by expanding each term of the identity

$$(f\sigma h)\,\alpha(g\sigma g)+(g\sigma g)\,\alpha(f\sigma h)\equiv 0$$

by means of the derivation rule (10). For $\tau=\pi$, it is obtained in a similar manner from the identity

$$(f\pi g)\,\alpha(g\pi h)+(g\pi h)\,\alpha(f\pi g)\equiv 0.$$

We shall now show that relation (11) leads to the following:

Lemma: Let $\{\mathcal{H},\tau,\alpha\}$ be a symmetric or antisymmetric Hamilton algebra. Then there exists in $\mathcal{H}$ a product

$$\rho:\ \mathcal{H}\otimes\mathcal{H}\to\mathcal{H}$$

such that

$$f\rho g=-g\rho f, \tag{12}$$

$$f\alpha(g\rho h)=(f\alpha g)\rho h+g\rho(f\alpha h), \tag{13}$$

$$g\rho(f\alpha h)=[f,g,h]. \tag{14}$$

Proof: For any $g\in\mathcal{H}$ let R_g denote the right multiplication operator defined by the identity $R_g f=f\alpha g$. Relation (11) implies $[k,f\alpha g,g]=0$ for every $k\in\ker R_g$. Thus, the kernel of the linear map $A:\mathcal{H}\to\mathcal{H}$ defined by $A=[\ ,f\alpha g,g]$ contains $\mathrm{Ker}R_g$. This guaranties the existence of a map $B:\mathcal{H}\to\mathcal{H}$ such that $A=B\circ R_g$, i.e.,

$$[h,f\alpha g,g]=B(h\alpha g).$$

Since the associator in this relation is linear in the variable $f\alpha g$, while the right-hand side is linear in $h\alpha g$, one can introduce a new map, M_g, defined by the identity

$$M_g(h\alpha g,f\alpha g)=[h,f\alpha g,g]. \tag{15}$$

It follows from relation (11) that the map

$$M_g:\ (\mathcal{H}\alpha g)\otimes(\mathcal{H}\alpha g)\to\mathcal{H}$$

is antisymmetric.

We shall now prove that there exists a map $M:\mathcal{H}\otimes\mathcal{H}\to\mathcal{H}$ such that, for every $g\in\mathcal{H}$, the map M_g is a restriction of the map M. First, we note that, for any given $g\in\mathcal{H}$, there is an extension of M_g to a map whose domain is $\mathcal{H}\otimes\mathcal{H}$. Such an extension of M_g shall also be denoted by M_g, and hence, from now on, $M_g:\mathcal{H}\otimes\mathcal{H}\to\mathcal{H}$. In substituting tg for g, where $t\in\mathcal{J}$, in relation (15), one verifies that the map M_g is the same for all points belonging to the same ray through the point g in $\mathcal{H}$, i.e., $M_{tg}=M_g$. By linearity, substitution of $tg+k$ for g in relation (15), t being a variable over $\mathcal{J}$ and k an arbitrary element of $\mathcal{H}$, now yields the equation

$$t^2\{M_g(h\alpha g,f\alpha g)-M_{g+k/t}(h\alpha g,f\alpha g)\}+t\{[h,f\alpha g,k]$$

$$+[h,f\alpha k,g]-M_{tg+k}(h\alpha g,f\alpha k)-M_{tg+k}(h\alpha k,f\alpha g)\}$$

$$+\{M_k(h\alpha k,f\alpha k)-M_{tg+k}(h\alpha k,f\alpha k)\}\equiv 0.$$

Since this relation must be satisfied for all values of t, the coefficients must vanish. The vanishing of the first coefficient implies $M_g=M_{g+k/t}$. Since k is arbitrary and does not appear in the arguments of these maps, the indices g and $g+k/t$ are independently arbitrary points of $\mathcal{H}$ and it follows that M_g does not depend on g. Thus, there is indeed only one map $M:\mathcal{H}\otimes\mathcal{H}\to\mathcal{H}$. The vanishing of the third coefficient yields the same result. The vanishing of the second coefficient implies the identity

$$[h,f\alpha g,k]-M(h\alpha k,f\alpha g)$$

$$=-[h,f\alpha k,g]+M(h\alpha g,f\alpha k).$$

In the antisymmetric case, $\tau=\pi$, the product α is onto. One can thus write $f\alpha g=m\in\mathcal{H}$, and consider m as an arbitrary point. The left-hand side of the above identity is then linear in m, while the right-hand side is not a function of this variable. This implies the vanishing of both sides, i.e.,

$$M(h\alpha k,m)=[h,m,k].$$

By writing the antisymmetric bilinear map M in the product symbolism one obtains relation (14). In the symmetric case, $\tau=\sigma$, the product α may be into, and the domain of the variable m may be only the complement of the α-center $\mathcal{C}$ in $\mathcal{H}$. In this case, one can extend the domain of the map M from $\mathcal{H}\otimes\mathcal{H}_\alpha$, where $\mathcal{H}_\alpha=\mathcal{H}\alpha\mathcal{H}$ is the image space of α, to $\mathcal{H}\otimes\mathcal{H}$ by setting $M(h\alpha k,m)=0$ if $m\in\mathcal{C}$. The derivation relation (13) is obtained by applying the derivation law (10), with $\tau=\sigma$ or $\tau=\pi$, to relation (14) in which the associator is first replaced by its definition. This completes the proof of the lemma.

We next investigate the two cases, $\tau=\pi$ and $\tau=\sigma$, separately.

The antisymmetric case

Let $\{\mathcal{H},\alpha\}$ be a central simple Lie algebra and let Π denote the set of all possible products π on $\mathcal{H}$ for which $\{\mathcal{H},\pi,\alpha\}$ is an antisymmetric Hamilton algebra. Since every $\pi\in\Pi$ is a linear map, $\pi:\mathcal{H}\otimes\mathcal{H}\to\mathcal{H}$, and since the derivation rule (10) is linear, the set Π is a linear space. The origin of the space Π corresponds to the trivial Hamilton algebra $\{\mathcal{H},0,\alpha\}$, while the existence of the degenerate algebra $\{\mathcal{H},\alpha,\alpha\}$ implies $\alpha\in\Pi$. Furthermore, according to the previous lemma, for every $\pi\in\Pi$ there exists an element $\rho\in\Pi$ defined by relation (14).

It is easy to see that all maps $\pi\in\Pi$, except the origin $\pi=0$, are onto. For any $\pi\in\Pi$, let $\mathcal{H}_\pi$ denote its image space. Obviously, since $\mathcal{H}\pi\mathcal{H}_\pi\subseteq\mathcal{H}_\pi$, the space $\mathcal{H}_\pi$ is an ideal in the algebra $\{\mathcal{H},\pi\}$. The derivation rule (10) implies

$$\mathcal{H}\alpha\mathcal{H}_\pi\subseteq\mathcal{H}_\pi,$$

i.e., $\mathcal{H}_\pi$ is an ideal in the algebra $\{\mathcal{H},\alpha\}$. Since $\{\mathcal{H},\alpha\}$ is simple, it follows that either $\mathcal{H}_\pi=\{0\}$, i.e., π is the zero map, or $\mathcal{H}_\pi=\mathcal{H}$, i.e., π is onto.

Consider now relation (14), which, written full, reads

$$(f\pi g)\pi h-f\pi(g\pi h)=g\rho(f\alpha h). \tag{16}$$

By subtracting from this identity its two cyclic transforms and using the derivation rule (13), one obtains the relation

$$2g\,\pi(h\pi f)=g\rho(f\alpha h)+g\alpha(f\rho h). \tag{17}$$

Let $\mathcal{J}_\pi$ be an ideal of the algebra $\{\mathcal{H}, \pi\}$ and let g be a point in this ideal. Relation (16) then implies that $\mathcal{H}\rho\mathcal{J}_\pi = \mathcal{J}_\pi$, i.e., $\mathcal{J}_\pi$ is also a ρ-ideal. With this result, relation (17) then implies that $\mathcal{J}_\pi$ is also an α-ideal. Hence, all algebras $\{\mathcal{H}, \pi\}$, $\pi \in \Pi$ and $\pi \neq 0$, are simple. We can now prove the following:

Theorem: There are no nondegenerate antisymmetric Hamilton algebras.

Proof: Since $\alpha \in \Pi$ (corresponding to the degenerate Hamilton algebra $\{\mathcal{H}, \alpha, \alpha\}$), the linear space Π is at least one-dimensional. The theorem states that it is exactly one-dimensional. We shall prove this by deriving a contradiction from the assumption dim $\Pi > 1$.

Consider a linear basis in Π consisting of the linearly independent points α and π_i, $i \in I$, where I is an index set. Every point $\pi \in \Pi$ is then of the form $\pi = \pi' + s\alpha$, where s is in $\mathcal{J}$ and $\pi' \in \Pi$ has no α component, i.e., π' is of the form

$$\pi' = \sum_i r^i \pi_i$$

the coefficients r^i being in $\mathcal{J}$. Since the map ρ corresponding to π is also in Π, as proved earlier, a set of coefficients, $a^i = a^i(r, s)$, $b = b(r, s)$, functions of r^i, s, exists such that

$$\rho = \sum_i a^i \pi_i + b\alpha.$$

By substituting these expressions into relation (17), one obtains

$$2g\pi'(h\pi' f) = g\rho'(f\alpha h) + g\alpha(f\rho' h), \tag{18}$$

where

$$\rho' = \sum_j (a^j + 2s\, r^j)\pi_j + (b + s^2)\alpha.$$

Since $g\pi'(h\pi' f)$ is independent of the variable s, ρ' must also be independent of s, which implies

$$a^j(r, s) = a^j(r, 0) - 2sr^j \quad \text{for all } j \in I,$$

$$b(r, s) = b(r, 0) - s^2.$$

Furthermore, since $g\pi'(h\pi' f)$ is a bilinear function of the coefficients r^i, identity (18) implies that ρ' must also be such a function, i.e.,

$$a^j(r, 0) = \sum_{i,k} P^j_{ik} r^i r^k,$$

$$b(r, 0) = \sum_{i,k} Q_{ik} r^i r^k,$$

where P^j_{ik}, Q_{ik} are some constant coefficients. Thus

$$a^j(r, s) = \sum_{i,k} P^j_{ik} r^i r^k - 2sr^j,$$

$$b(r, s) = \sum_{i,k} Q_{ik} r^i r^k - s^2.$$

Consider now the following system of quadratic equations:

$$\sum_{i,k} P^j_{ik} r^i r^k - 2sr^j = 0,$$

$$\sum_i r^i r^i = 1.$$

Since there are as many unknowns r^k, s, as equations, the fundamental theorem of algebra guaranties the existence of solutions for r^i, s. These solutions are in general in the algebraic closure of $\mathcal{J}$. Let r^i_0, s_0 be a set of solutions, and $\pi_0 = \sum_i r^i_0 \pi_i + s_0\alpha$ the corresponding product π. The equation $\sum_i r^i r^i = 1$ prevents the vanishing of all coefficients r^i, and hence π_0 can not be degenerate or trivial. Since $a^i(r_0, s_0) = 0$, the product ρ corresponding to π_0 is proportional to α, say $\rho = -b_0\alpha$. Equation (17) then implies

$$f\pi_0(g\pi_0 h) = b_0 f\alpha(g\alpha h). \tag{19}$$

The coefficient b_0 cannot vanish, since this would imply $\pi_0 = 0$. Since $g\alpha h$ can be an arbitrary element k of $\mathcal{H}$, identity (19) implies that $g\pi_0 h$ is a linear function of $g\alpha h$:

$$g\pi_0 h = L(g\alpha h) \tag{20}$$

where $L: \mathcal{H} \to \mathcal{H}$. The same relation further implies the following identity involving L and α only:

$$L(f\alpha(Lk)) = b_0 f\alpha k, \tag{21}$$

f and k being arbitrary elements of $\mathcal{H}$. From the derivation condition,

$$f\alpha(g\pi_0 h) = (f\alpha g)\pi_0 h + g\pi_0(f\alpha h),$$

one gets the identity

$$f\alpha(Lk) = L(f\alpha k) \tag{22}$$

which, substituted into relation (21), yields

$$L^2(f\alpha k) = b_0 f\alpha k.$$

Since the point $f\alpha k$ can be chosen arbitrarily, one has

$$L^2 = b_0 I \tag{23}$$

where I is the identity map in $\mathcal{H}$. Relation (23) implies that the linear space $\mathcal{H}$ is a direct sum of the two eigenspaces of the operator L, i.e., $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$. In $\mathcal{H}_+$, $L = \sqrt{b_0} I$, and in $\mathcal{H}_-$, $L = -\sqrt{b_0} I$. Identity (22) implies that the points k and $f\alpha k$ must belong to the same eigenspace for all $f \in \mathcal{H}$, i.e., $\mathcal{H}\alpha\mathcal{H}_+ \subseteq \mathcal{H}_+$ and $\mathcal{H}\alpha\mathcal{H}_- \subseteq \mathcal{H}_-$. Hence, the eigenspaces $\mathcal{H}_+$ and $\mathcal{H}_-$ are ideals in the algebra $\{\mathcal{H}, \alpha\}$. Since the algebra $\{\mathcal{H}, \alpha\}$ is central simple, this implies that $\mathcal{H}_+$ and $\mathcal{H}_-$ are either $\{0\}$ or $\mathcal{H}$. Thus, there can be only one eigenspace. In other words, the operator L is a multiple of the identity, which implies that π_0 is degenerate. This contradicts the previous conclusion that it is not. Hence, Π is one-dimensional. This completes the proof of the theorem.

The symmetric case

Let $\{\mathcal{H}, \sigma, \alpha\}$ be a symmetric Hamilton algebra. According to relation (14), the product σ satisfies the identity

$$(f\sigma g)\sigma h - f\sigma(g\sigma h) = g\rho(f\alpha h), \tag{24}$$

where ρ is an antisymmetric product with respect to which the product α is a derivation. One verifies that the associator of a symmetric product satisfies the identity

$$[f, g, h] + [g, h, f] + [h, f, g] = 0.$$

By substituting into this identity Eq. (24), one obtains the relation

$$g\rho(f\alpha h) + h\rho(g\alpha f) + f\rho(h\alpha g) = 0$$

which, with the derivation rule of α with respect to ρ, yields

$$f\rho(h\alpha g) = f\alpha(h\rho g).$$

This relation implies that $h\rho g$ is a linear function of $h\alpha g$, and by arguments similar to those used in proving relation (23) one shows that the product ρ is proportional to α. Hence, relation (24) reads

$$[f,g,h] = ag\alpha(f\alpha, h) \tag{25}$$

where $a \in \mathcal{J}$. We shall call this identity the *association relation*. Using the Jacobi identity, one can also write the association relation in the form

$$[f,g,h]_\sigma = a[f,g,h]_\alpha. \tag{26}$$

One sees that in a Hamilton algebra $\{\mathcal{H}, \sigma, \alpha\}$ the properties of the product σ depend crucially on whether the constant a is zero or different from zero. If $a=0$, we shall say that the Hamilton algebra is *classical*. In a classical Hamilton algebra, the algebra $\{\mathcal{H}, \sigma\}$ is both associative and commutative. A symmetric Hamilton algebra for which $a \neq 0$ shall be called a *quantum* Hamilton algebra. It follows from the association relation that in a quantum Hamilton algebra the product σ satisfies the identity $[f^2, g, f] = 0$. This identity, together with the symmetry of σ, defines a so-called Jordan algebra. A Jordan algebra whose product can be expressed as the anticommutator of an associative product is called a special Jordan algebra. In order to check whether the Jordan algebra $\{\mathcal{H}, \sigma\}$ is special, we consider a product $\beta = \sigma + b\alpha$, and require $[f,g,h]_\beta = 0$, which implies the relation

$$[f,g,h]_\sigma = -b^2[f,g,h,]_\alpha. \tag{27}$$

Comparison of relations (27) and (26) shows that the product $\beta = \sigma + \sqrt{-a}\ \alpha$ is associative. Hence, in a quantum Hamilton algebra $\{\mathcal{H}, \sigma, \alpha\}$, the algebra $\{\mathcal{H}, \sigma\}$ is a special Jordan algebra. Let $\mathcal{J}_\sigma$ be an ideal in the algebra $\{\mathcal{H}, \sigma\}$, and let $g \in \mathcal{J}_\sigma$. The association relation (25), with $a \neq 0$, implies that $\mathcal{J}_\sigma$ is also an α-ideal. Hence, in a quantum Hamilton algebra $\{\mathcal{H}, \sigma, \alpha\}$, the algebra $\{\mathcal{H}, \sigma\}$ is a central simple special Jordan algebra. Since a simple special Jordan algebra has a unit element e, the center $\mathcal{C}$ of the algebra $\{\mathcal{H}, \alpha\}$ in a quantum Hamilton algebra $\{\mathcal{H}, \sigma, \alpha\}$ is $\mathcal{J}e$.

In quantum mechanics, the base field $\mathcal{J}$ is the field R of real numbers, the linear space $\mathcal{H}$ is the space of Hermitian operators on a Hilbert space, and the special Jordan algebra $\{\mathcal{H}, \sigma\}$ is the algebra of these operators under the Hermitian product. The associative product β is the product of operators on the Hilbert space, and the product α is the commutator divided by $\sqrt{-1}\hbar$. Thus, the constant a is related to the quantum of action, $\hbar \in R$, by the equation $a = (\hbar/2)^2$.

ACKNOWLEDGMENTS

We wish to thank Michel Cahen, Debendranath Sahoo, and Stanley Sawyer for helpful discussions.

*This work was supported by the USAFOSR Contract numbers F 44620-71-C-0013 and 68-1524 and by the NSF Contract number GP-14803.

[1]See, e.g., J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, 1955), Chap. VI and D. Bohm, *Quantum Theory* (New York, 1951), Chap. 22.

[2]Y. Aharonov and A. Petersen, in *Quantum Mechanics and Beyond*, edited by T. Bastin (Cambridge U.P., Cambridge, 1970), p. 135.

[3]For other algebraic studies of classical and quantum mechanics, see, for example, P. Jordan, J. von Neumann, and E. Wigner, Ann. Math. 35, 29 (1934); G. Falk, Math. Ann. 123, 379 (1951); T.F. Jordan and E.C.G. Sudarshan, Rev. Mod. Phys. 33, 518 (1961); C.L. Mehta, J. Math. Phys. 5 (1964); R.T. Prosser, Mem. Am. Math. Soc. No. 61 (1966); E. Grgin and A. Petersen, Phys. Rev. D 5, 300 (1972).

Cancellation of the Green's function in the generation of continuum bound states by nonlocal potentials*

Thomas O. Krause and Bernard Mulligan

Department of Physics, The Ohio State University, Columbus, Ohio 43210
(Received 10 December 1973)

Continuum bound states (CBS) are known to appear in the spectra of some nonlocal scattering equations. We give a simple derivation of the presence of these states consistent with the requirement that such states occur for zeros of the Fredholm determinant. Examination of the form of the nonlocal potential necessary to the generation of a CBS shows that CBS solutions appear only when the potential has the effect of cancelling the Green's function in the kernel of the integral equation. Several examples from the literature are cited to demonstrate this characteristic feature of CBS.

1. INTRODUCTION

Nonlocal potentials have been the subject of considerable investigation since their introduction into the nuclear problem by Yamaguchi.[1] The work of Feshbach[2] and of Perey and Buck[3] has been important in extending the concept of the nonlocal potential to the nucleon–nucleus interaction. It was pointed out by Gourdin and Martin[4,5] that one feature of the separable nonlocal potential is the possibility of wavefunctions at positive energy which behave asymptotically like bound states. These states have been studied by various authors[6–9] who have labeled them positive energy bound states, positive energy degenerate states, spurious bound states, or continuum bound states (CBS). In this paper we demonstrate the occurrence of CBS solutions in a particularly simple manner. We then show that such solutions can be generated only by nonlocal potentials which have the property of cancelling the effect of the Green's function in the scattering equation.

2. CONTINUUM BOUND STATES

Consider the integral equation which defines the physical solutions to the scattering equation

$$\psi^{\pm}(k,r)=\sin(kr)+\int_0^\infty\int_0^\infty G^{\pm}(k,r,r')V(r',s)\psi^{\pm}(k,s)\,ds\,dr' \tag{1}$$

and the associated homogeneous equation

$$\chi^{\pm}(k,r)=\int_0^\infty\int_0^\infty G^{\pm}(k,r,r')V(r',s)\chi^{\pm}(k,s)\,ds\,dr', \tag{2}$$

where

$$G^{\pm}(k,r,r')=-(1/k)\exp(\pm ikr_>)\sin(kr_<) \tag{3}$$

for both $k^2>0$ and $k^2<0$.

It is well known from Fredholm theory that if (2) has a solution for some value of k, then the complete solution to (1) at that value of k will include an arbitrary amount of the solution to (2). By using (2) and (3), we can find an expression for the asymptotic form of the homogeneous solution:

$$\chi^{\pm}(k,r)\underset{r\to\infty}{\longrightarrow}-\frac{\exp(\pm ikr)}{k}\int_0^\infty\int_0^\infty\sin(kr')V(r',s)\chi^{\pm}(k,s)\,ds\,dr'. \tag{4}$$

In general, therefore, we would expect positive energy solutions to (2) (i.e., k real) to oscillate at infinity while negative energy solutions (k imaginary) would decay like $\exp(-\epsilon r)$, where $\epsilon^2=-k^2>0$. However, we are interested in the special condition in which both (1) and (2) have solutions for the same real value of k. In this case, we can further examine Eq. (4) by extension of a procedure used by Martin.[5]

Multiply (1) by $\chi^{\pm}(k,t)V(r,t)$ and integrate t and r over $(0,\infty)$. Similarly, multiply (2) by $\psi^{\pm}(k,t)V(t,r)$ and integrate r and t over $(0,\infty)$. The results are

$$\begin{aligned}\int_0^\infty\int_0^\infty\psi^{\pm}(k,r)\chi^{\pm}(k,t)V(r,t)\,dt\,dr&=\int_0^\infty\int_0^\infty\sin(kr)\chi^{\pm}(k,t)V(r,t)\,dt\,dr\\&+\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty G^{\pm}(k,r,r')V(r',s)\psi^{\pm}(k,s)\chi^{\pm}(k,t)\\&\times V(r,t)\,ds\,dr'\,dt\,dr,\end{aligned} \tag{5}$$

$$\begin{aligned}\int_0^\infty\int_0^\infty\chi^{\pm}(k,r)\psi^{\pm}(k,t)V(t,r)\,dr\,dt&=\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty G^{\pm}(k,r,r')V(r',s)\chi^{\pm}(k,s)\psi^{\pm}(k,t)\\&\times V(t,r)\,ds\,dr'\,dr\,dt.\end{aligned} \tag{6}$$

Subtracting (6) from (5) and making use of the assumed symmetry of $V(r,r')$ and of $G^{\pm}(k,r,r')$ leaves

$$\int_0^\infty\int_0^\infty\sin(kr)\chi^{\pm}(k,t)V(r,t)\,dr\,dt=0. \tag{7}$$

With (7), the asymptotic form in Eq. (4) becomes

$$\chi^{\pm}(k,r)\underset{r\to\infty}{\longrightarrow}0. \tag{8}$$

This means we can have positive energy solutions of (2) which vanish at infinity. These are continuum bound states (CBS).

The homogeneous equation (2) will have solutions for some real value of k only when the Fredholm determinant of the integral equations (1) and (2) vanishes for that value of k. In that case, the solution of the inhomogeneous equation (1) contains an arbitrary amount of the solutions to (2). We have demonstrated that these solutions to (2) at that value of k will vanish at infinity even for $k^2>0$ and are therefore CBS solutions.

3. CANCELLATION OF THE GREEN'S FUNCTION

We wish to examine the form which a potential must have in order to produce CBS. Local potentials cannot produce CBS[10] because the Fredholm determinant for

the scattering equation cannot vanish for real values of k, except possibly at $k=0$; the derivation of the asymptotic behavior of CBS which was given in the previous section fails for a local potential. However, nonlocality in a potential is not sufficient to guarantee CBS solutions. To generate CBS requires that the form of the potential be such that Eq. (2) as well as Eq. (1) have a solution (or, equivalently, that the Fredholm determinant vanish) at some value of k. In this paper, we show that this will occur only if the potential cancels the effect of the Green's function in the scattering equation.

An attempt to solve directly the problem of what is required of a nonlocal potential to produce CBS could be made by writing out the Fredholm determinant, setting it equal to zero, and inverting to find the potential. This approach is not fruitful because of the complexity of the determinant. Additional insight into the requisite form of the potential can be gained, however, by reconsidering the asymptotic behavior of the homogeneous solutions. The form of the Green's function (3) causes the solutions to the homogeneous equation to appear to behave at infinity like $\exp(\pm ikr)$, as indicated by Eq. (4). But, as we have shown, solutions need not behave in this way. For CBS solutions, it is only necessary that they vanish at infinite distances. In order for this to happen, the effect of the Green's function in the integral equation must disappear. This can be the case only when the Green's function is cancelled by the potential. Let us see how this might occur.

In three dimensions, the Schrödinger equation is

$$(\nabla_r^2+k^2)\Psi(\mathbf{k},\mathbf{r})=\int V(\mathbf{r},\mathbf{r}')\Psi(\mathbf{k},\mathbf{r}')\,d\mathbf{r}'.$$

We can formally investigate the possibility of a homogeneous solution to this equation by bringing the differential operator to the right-hand side

$$\Psi(\mathbf{k},\mathbf{r})=\frac{1}{\nabla_r^2+k^2}\int V(\mathbf{r},\mathbf{r}')\Psi(\mathbf{k},\mathbf{r}')\,d\mathbf{r}'. \tag{9}$$

Of course, this form is most inconvenient. Since division by a differential operator is equivalent to multiplication by an integral operator, Eq. (9) is usually rewritten in integral form with the aid of a Green's function $\mathcal{G}$ which takes account of the boundary conditions

$$\Psi(\mathbf{k},\mathbf{r})=\int \mathcal{G}(\mathbf{k},\mathbf{r},\mathbf{r}')\int V(\mathbf{r}',\mathbf{s})\Psi(\mathbf{k},\mathbf{s})\,d\mathbf{s}\,d\mathbf{r}'.$$

The effect under discussion can, however, be more clearly seen from Eq. (9). Suppose the potential has the separable symmetric form

$$V(\mathbf{r},\mathbf{r}')=\lambda(\nabla_r^2+\kappa^2)\Phi(\mathbf{r})(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}'), \tag{10}$$

where κ is a constant. Then at that energy where $k=\kappa$ the integral equation (9) becomes

$$\Psi(\kappa,\mathbf{r})=\lambda\Phi(\mathbf{r})\int[(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')]\,\Psi(\kappa,\mathbf{r}')\,d\mathbf{r}'. \tag{11}$$

This is a "pure" integral equation. It no longer contains a Green's function, and it can behave asymptotically in a manner other than $\exp(\pm ikr)$. The form of the solution to (11) is clearly

$$\Psi(\kappa,\mathbf{r})=N\Phi(\mathbf{r}), \tag{12}$$

where N is arbitrary since (9) is homogeneous in $\Psi(\kappa,\mathbf{r})$. If a potential can be written in the form (10), then it can be seen immediately that there may be a continuum bound state $E_{\text{c.m.}}=\hbar^2\kappa^2/2\mu$, and that the form of the CBS wavefunction would be $\Phi(\mathbf{r})$.

There is an additional condition which must be met by the simple separable potential (10) in order that a solution of the form (12) exist. The solution (12) must be consistent with Eq. (11). That is,

$$\Psi(\kappa,\mathbf{r})=N\Phi(\mathbf{r})=\lambda\Phi(\mathbf{r})\int[(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')]\Psi(\kappa,\mathbf{r}')\,d\mathbf{r}'.$$

Therefore

$$N=\lambda\int[(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')]\Psi(\kappa,\mathbf{r}')\,d\mathbf{r}'. \tag{13}$$

Substituting (12) into (13), we have

$$N=\lambda\int[(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')]N\Phi(\mathbf{r}')\,d\mathbf{r}'.$$

or

$$\lambda\int[(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')]\,d\mathbf{r}'=1. \tag{14}$$

Condition (14) is equivalent to the requirement derived by Gourdin and Martin[4,5] for the generation of a CBS by a single term separable potential although they approached the problem from a completely different point of view, that of phase shift behavior.

The concept that cancellation of the Green's function is necessary to the generation of CBS is not restricted to the simple separable potential (10). Consider the potential

$$V(\mathbf{r},\mathbf{r}')=\lambda(\nabla_r^2+\kappa^2)\Phi(\mathbf{r})(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')+\sum_{i=2}^{N}\lambda_i F_i(\mathbf{r})F_i(\mathbf{r}'). \tag{15}$$

Substitution into (9) yields (at $k=\kappa$)

$$\Psi(\kappa,\mathbf{r})=\lambda\Phi(\mathbf{r})\int(\nabla_{r'}^2+\kappa^2)\Phi(\mathbf{r}')\Psi(\kappa,\mathbf{r}')\,d\mathbf{r}'+\frac{1}{\nabla_r^2+\kappa^2}\sum_{i=2}^{N}\lambda_i F_i(\mathbf{r})\int F_i(\mathbf{r}')\Psi(\kappa,\mathbf{r}')\,d\mathbf{r}'. \tag{16}$$

The effect of the Green's function has been cancelled from the first term of (16) but not from other terms. At positive energy the oscillatory behavior of the wavefunction due to these terms will dominate at infinity. Thus, this potential can produce CBS solutions only if

$$\sum_{i=2}^{N}\lambda_i F_i(\mathbf{r})\int F_i(\mathbf{r}')\Psi(\kappa,\mathbf{r}')\,d\mathbf{r}'=0. \tag{17}$$

If the constraint (14) is also satisfied by the potential (15), then a CBS solution will be produced.

This method of approaching the existence of a CBS is consistent with the requirements for a CBS presented in Sec. 2. That is, it can be shown by direct calculation that a potential of the form (15) with constraints (14) and (17) will result in a vanishing Fredholm determinant at $k=\kappa$.

Ghirardi and Rimini[11] have written a potential which generates N chosen wavefunctions $|\psi_i\rangle$ at N energies E_i,

$$V(\mathbf{r},\mathbf{r}')=-\sum_{j,k=1}^{N}X_{jk}\langle\mathbf{r}|W_j\rangle\langle W_k|\mathbf{r}'\rangle, \tag{18}$$

where $\langle\mathbf{r}|W_j\rangle=(E_j-T)|\psi_j\rangle$ and T is the single particle kinetic energy operator. Examination of this potential

shows that it is a sum of terms where each term has the form of Eq. (10). The method used by Ghirardi and Rimini to pick the coefficients X_{jk} is equivalent to requiring the constraint (17). This N term G—R potential, therefore, can be used to produce an orthogonal set of N CBS solutions at N required energies.

The concept of cancellation of the Green's function is considerably easier to use in k-space since the Green's function is the simple function $(\kappa^2 - k^2)^{-1}$, where $\hbar^2\kappa^2/2\mu$ is the center-of-mass energy. Tabakin[12] has written a single term separable k-space potential

$$V(k,k') = \frac{\hbar^2}{2\mu}\frac{1}{2\pi^2}\, g(k)g(k'), \tag{19}$$

where

$$g(k) = \alpha(k_c^2 - k^2)\,\frac{k^2+d^2}{k^2+b^2}\,\frac{1}{k^4+a^4}\,. \tag{20}$$

From the foregoing discussion it can be seen immediately that this potential generates a CBS at $E_{c.m} = \hbar^2 k_c^2/2\mu$. The form of the CBS wavefunction can be seen at once by comparing (19) with (10) and (12). The k-space representation of this wavefunction is

$$\phi(k) = N\,\frac{k^2+d^2}{k^2+b^2}\,\frac{1}{k^4+a^4}\,,$$

where N is a normalization constant. These results are the same as those which Tabakin[12] has derived from a consideration of phase shift behavior. The equation which Tabakin derives as a necessary condition for a CBS [Eq. (15) of Ref. 12] is equivalent to our constraint (14). The example of Tabakin's potential in particular demonstrates some of the advantages of considering the CBS problem through the cancellation of the Green's function.

4. A MORE GENERAL POTENTIAL

The potentials given in the previous section explicitly contain a term which cancels the Green's function. Indeed, for a simple separable potential, the inverse Green's function *must* appear explicitly as there is only one way to write a one term potential. For multiterm potentials the appearance of this inverse Green's function need not be explicit. However, any separable potential which produces a CBS, although complicated in appearance, is such that it can be written in a form like (15) with constraints (14) and (17).

As an example which does not contain an inverse Green's function explicitly, but which can have a CBS for a proper choice of its parameters, consider the following k-space potential:

$$V(k,k') = \frac{V_A^2}{(k^2+a^2)(k'^2+a^2)} - \frac{V_B^2}{(k^2+b^2)(k'^2+b^2)}\,. \tag{21}$$

Mongan[13] has shown that potentials of this form fit the two nucleon phase shifts over a range of energies. We shall show that it is possible to choose parameters in (21) (different from those selected by Mongan) such that the potential will have a square-integrable solution at positive energy. For each such choice of parameters, the solution will be the result of the cancellation of the Green's function in the integral equation for the wavefunction.

We obtain the one-dimensional k-space Schrödinger equation by Fourier transforming the three-dimensional configuration space Schrödinger equation and integrating out the angular dependence. The result is

$$(-k^2+\kappa^2)\phi(k) = \frac{2\mu}{\hbar^2}\int_0^\infty V(k,k')\phi(k')k'^2\,dk', \tag{22}$$

where the energy dependence has been written

$$\kappa^2 = 2\mu E/\hbar^2.$$

As before, we formally obtain the homogeneous integral equation for $\phi(k)$ by substituting (21) into (22) and dividing by $(-k^2+\kappa^2)$.

$$\phi(k) = \frac{2\mu}{\hbar^2}\frac{1}{\kappa^2-k^2}\int_0^\infty \frac{V_A^2\phi(k')k'^2\,dk'}{(k^2+a^2)(k'^2+a^2)} - \frac{2\mu}{\hbar^2}\frac{1}{\kappa^2-k^2}\int_0^\infty \frac{V_B^2\phi(k')k'^2\,dk'}{(k^2+b^2)(k'^2+b^2)}\,. \tag{23}$$

With the definitions

$$\alpha = \int_0^\infty \frac{\phi(k')k'^2dk'}{(k'^2+a^2)}, \tag{24}$$

$$\beta = \int_0^\infty \frac{\phi(k')k'^2dk'}{(k'^2+b^2)}, \tag{25}$$

Eq. (23) becomes

$$\phi(k) = \frac{2\mu}{\hbar^2}\frac{V_A^2\alpha}{(\kappa^2-k^2)(k^2+a^2)} - \frac{2\mu}{\hbar^2}\frac{V_B^2\beta}{(\kappa^2-k^2)(k^2+b^2)}\,. \tag{26}$$

Equation (26) represents the most general form for a solution to the homogeneous integral equation (23) with this potential.

The constants α and β must satisfy certain self-consistency conditions if (26) is to be a solution. Specifically, we can solve for α and β by substituting the wavefunction (26) into Eqs. (24) and (25). This will yield two consistency equations involving α and β which must be satisfied simultaneously if (26) is to be a solution to (23).

We are looking for solutions to (23) of the CBS type which go to zero at infinity. The appearance of the Green's function $(\kappa^2 - k^2)^{-1}$ in the wavefunction (26) would seem to exclude the possibility of CBS solutions because this factor, when Fourier transformed back into configuration space, will produce oscillatory functions which will dominate the behavior at infinity. However, the wavefunction (26) can be a CBS solution if we pick the parameters such that the Green's function is cancelled. We now show how this may be accomplished.

Let us put the wavefunction (26) over a common denominator

$$\phi(k) = \frac{2\mu}{\hbar^2}\,\frac{V_A^2\alpha(k^2+b^2) - V_B^2\beta(k^2+a^2)}{(\kappa^2-k^2)(k^2+a^2)(k^2+b^2)}\,. \tag{27}$$

We are looking for a wavefunction of the form

$$\phi(k) = \frac{2\mu}{\hbar^2}\,\frac{c}{(k^2+a^2)(k^2+b^2)}, \tag{28}$$

where the Green's function dependence has specifically canceled out. The requirement for writing (27) in the form (28) is

$$V_A^2\alpha(k^2+b^2) - V_B^2\beta(k^2+a^2) = c(\kappa^2 - k^2)$$

or

$$V_A^2\alpha - V_B^2\beta = -c$$

and

$$V_A^2\alpha b^2 - V_B^2\beta a^2 = c\kappa^2.$$

Eliminating c, we have

$$\frac{\beta}{\alpha} = \frac{V_A^2}{V_B^2}\frac{b^2+\kappa^2}{a^2+\kappa^2}. \tag{29}$$

This condition on α and β must be met in order to produce a wavefunction of the form (28) without Green's function dependence. Condition (29) is independent of the consistency conditions on α and β which must be satisfied separately in order that the potential produce any solution at all. This is analogous to the discussion of Sec. 3 in which the consistency condition (14) must be met independently of the conditions for cancellation of the Green's function.

In this example we have chosen to cancel the Green's function from the wavefunction (26) rather than to put the potential (21) into a form in which the cancellation would be explicit. This is strictly a matter of convenience. The potential is a function of two variables, the wavefunction only one. It is behavior of the wavefunction, after all, which is our ultimate interest.

5. CONCLUSIONS

The study of scattering involves the inhomogeneous wave equation

$$\Psi^{\pm}(\mathbf{k},\mathbf{r}) = \Psi_0 + \int\int \mathcal{G}^{\pm}(\mathbf{k},\mathbf{r},\mathbf{r}')V(\mathbf{r}',\mathbf{s})\Psi^{\pm}(\mathbf{k},\mathbf{s})\,d\mathbf{s}\,d\mathbf{r}'.$$

This equation has solutions at all real k for any potential. For those particular values of k at which the Fredholm determinant for this equation vanishes, the solutions to this inhomogeneous equation will include arbitrary amounts of the solutions to the associated homogeneous equation (which has solutions only at these values of k). We have shown that these solutions to the homogeneous equation are therefore continuum bound states. When the Fredholm determinant vanishes, then, the complete solution to the inhomogeneous equation will include CBS wavefunctions in addition to the regular scattering wave functions.

The presence of the Green's function would be expected to force the solutions of the homogeneous equation to oscillate at infinity. Since CBS solutions do not oscillate at infinity, they can be produced only by potentials which cancel the effect of the Green's function. In single term potentials which produce CBS, the inverse Green's function which performs the cancellation must appear explicitly. If a single term potential contains an inverse Green's function and also meets the independent consistency condition (14), then the energy of the CBS and the form of the CBS wavefunction can be obtained by inspection.

For potentials of two or more terms, one term with an explicit inverse Green's function may appear for each CBS produced. If the potential has this form it is again possible to determine the CBS energies and wavefunctions by inspection. In this case the sum of all terms which do not generate a particular CBS must be orthogonal to that CBS wavefunction. Although terms with explicit Green's function dependence may appear in a potential which generates CBS, such explicit dependence is not necessary. It is possible, however, to rewrite in this form any potential which produces CBS. Also, regardless of the form of the potential, there exist consistency conditions which must be met independently of the cancellation of the Green's function if CBS solutions are to be generated.

ACKNOWLEDGMENT

We are grateful to L. G. Arnold for helpful discussions.

*This work has been presented by one of us (T. O. K.) as part of a Ph. D. thesis at The Ohio State University.
[1]Y. Yamaguchi, Phys. Rev. 95, 1628 (1954); Y. Yamaguchi and Y. Yamaguchi, Phys. Rev. 95, 1635 (1954).
[2]H. Feshbach, Ann. Phys. (N.Y.) 5, 357 (1958); 19, 287 (1962).
[3]F. Perey and B. Buck, Nucl. Phys. 32, 353 (1962).
[4]M. Gourdin and A. Martin, Nuovo Cimento 6, 757 (1957); 8, 699 (1958).
[5]A. Martin, Nuovo Cimento 7, 607 (1958).
[6]M. Bertero, G. Talenti, and G. Viano, Nuovo Cimento 46, 337 (1966); Comm. Math. Phys. 6, 128 (1967); Nucl. Phys. A 113, 625 (1968); Nucl. Phys. A 115, 395 (1968).
[7]Kh. Chadan, Nuovo Cimento 10, 892 (1958); 47, 510 (1967).
[8]M. Bolsterli, Phys. Rev. 182, 1095 (1969).
[9]P. Beregi, Nucl. Phys. A 206, 217 (1973).
[10]L. G. Arnold and A. MacKellar, Phys. Rev. C 3, 1095 (1971).
[11]G. Ghirardi and A. Rimini, J. Math. Phys. 5, 722 (1964).
[12]F. Tabakin, Phys. Rev. 174, 1208 (1968).
[13]T. R. Mongan, Phys. Rev. 175, 1260 (1968); Phys. Rev. 178, 1597 (1969).

Motion of charged particles in homogeneous electromagnetic fields

Eli Honig*† and Engelbert L. Schucking*

Department of Physics, New York University, New York, New York 10003

C. V. Vishveshwara

Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 8 August 1973)

An invariant geometrical description of the world lines of charged particles in arbitrary homogeneous electromagnetic fields is presented. This is accomplished through the combined use of the Frenet–Serret equations and the Lorentz equation. The results apply to flat as well as Riemannian space–time. The intrinsic scalars associated with these curves (i.e., their curvatures and first and second torsions) are found to be constants of the motion when they are well defined. Moreover, they form simple relationships with the field invariants as well as with the energy and momentum densities of the rest frame fields. When they are evaluated in the instantaneous rest frame of the particle, the Frenet vectors lend themselves to simple physical interpretation. It is shown that one cannot distinguish in an intrinsic geometrical manner between the curves of positive and negative charges. The same is true for positive and negative magnetic monopoles if they exist. In such a case, however, one would be able to distinguish intrinsically between ordinary and magnetic charges. The effect of duality rotations of the field tensor on the Frenet scalars is studied. A physical realization of the Frenet frame is obtained by considering the classical description of spin precession. Finally the Frenet formalism is applied to timelike Killing trajectories. These are shown to closely resemble the world lines of charged particles in homogeneous electromagnetic fields.

I. INTRODUCTION

The main purpose of the present work is to investigate the intrinsic geometrical features of the world line of a charged particle moving under the influence of a homogeneous electromagnetic field. Our definition of a homogeneous field is equivalent to the field tensor being covariantly constant. The motion of the particle is then governed by the Lorenz equation that incorporates this constant electromagnetic field tensor. The conventional way of describing the space—time trajectory of the particle is through a parametric representation taking the initial conditions into account. This task can be accomplished in principle by integrating the Lorentz equation (see the Appendix). Even if this approach may be of practical value, it suffers from two drawbacks, namely it is not coordinate independent and, moreover, does not give direct information about the geometrical characteristics of the world line. These two shortcomings can at once be removed by recourse to the Frenet—Serret formalism.[1] Here a curve is associated at every point with the orthonormal Frenet—Serret tetrad; the members of the tetrad, the first of which is the unit tangent vector to the curve, obey the Frenet—Serret equations. Furthermore, the intrinsic geometry of the curve is uniquely determined by the Frenet scalars, namely the curvature and the first and second torsions defined along the curve. We combine the Lorentz equation and the Frenet—Serret equations by identifying the curve with the timelike world line of the charged particle. The intrinsic geometry of the world line is thereby obtained in a manifestly direct and elegant manner, which is completely independent of coordinates as well as initial conditions. While the Frenet—Serret method provides an excellent framework for our present study, the physical phenomenon we have invoked serves, in turn, to illuminate the basic mechanism inherent to the formalism itself. This is achieved by establishing simple connections between the field quantities and the geometrical entities. For instance, the Frenet scalars, which will be shown to be constants along the world line, are directly related to the Lorentz invariants of the electromagnetic field. In addition, the Frenet vectors, when evaluated in the instantaneous rest frame of the particle, find simple interpretation in terms of the electric and magnetic fields, and the Poynting vector. Further insight is achieved in this direction, if we assume that the charge has spin and magnetic moment with a gyromagnetic ratio of two. We shall see that in this case the spin and hence the magnetic moment will not precess at all with respect to the spatial Frenet triad as it is carried along the world line of the charge. Thus along the path of the particle the frame which has constant components of the magnetic moment is either identical to the Frenet frame or at the most differs from it by a rigid rotation. This then provides for any observer a physical realization of the Frenet vectors which would otherwise remain of essentially mathematical significance.

To summarize, the problem we study in this paper helps to describe the motion of charged particles in homogeneous fields by focussing on the invariant geometrical aspects of their paths, and at the same time clarifies the nature of the Frenet—Serret formalism on the basis of a simple physical model.

The rest of this paper is divided into different parts as follows. Section II is devoted mainly to deriving general expressions for the geometrical quantities in terms of the field tensor. In Sec. III we specialize to the instantaneous rest frame of the particle and obtain simple interpretations for the already derived relations. Section IV examines the effect of duality rotations[2,3] of the field tensor on the Frenet scalars. We point out in Sec. V the formal similarity that exists between the world lines of charged particles and timelike Killing trajectories admitted by Riemannian spaces. Finally in the Appendix we present a covariant solution to the Lorentz equation for

homogeneous electromagnetic fields in flat space–times. This has been included not only for the sake of completeness, but also to contrast this rather cumbersome approach with the elegant Frenet–Serret method.

II. GENERAL RELATIONS

The motion of a charged particle of mass m and charge e in an electromagnetic field is governed by the Lorentz equation

$$\dot{u}^\mu = \frac{e}{mc^2}\overline{F}^\mu{}_\nu u^\nu ,$$

where u^μ is the 4-velocity of the particle and $\overline{F}_{\mu\nu}$ is the electromagnetic field tensor. In our notation Greek indices range from 0 to 3 and italic indices from 1 to 3. The signature of the metric is $(+,-,-,-)$. A dot over any quantity stands for its absolute derivative with respect to the proper time τ along the world line of the particle, e.g.,

$$\dot{A}^{\mu\nu\ldots}_{\alpha\beta\ldots} \equiv \frac{DA^{\mu\nu\ldots}_{\alpha\beta\ldots}}{D\tau} \equiv A^{\mu\nu\ldots}_{\alpha\beta\ldots;\gamma}u^\gamma ,$$

the semicolon representing covariant differentiation. It is convenient to write the Lorentz equation as

$$\dot{u}^\mu = F^\mu{}_\nu u^\nu \tag{1}$$

with $F_{\mu\nu} = \lambda\overline{F}_{\mu\nu}$ and $\lambda = e/mc^2$.

As was mentioned in the introduction, the intrinsic geometrical characteristics of a curve Γ can most elegantly be described by utilizing the Frenet–Serret equations. The curve Γ, which in our case is the world line of the particle, is assumed to be sufficiently smooth (at least of class C^3). Associated with Γ at every point is the Frenet–Serret tetrad consisting of four orthonormal vectors $e^\mu_{(\alpha)}$. The index within the parenthesis is the tetrad index which singles out a particular member of the tetrad. In particular, $e^\mu_{(0)} \equiv u^\mu$ is the timelike unit tangent vector to the curve. The spatial triad $e^\mu_{(i)}$ $(i=1,2,3)$ consist of the normal, the first and the second binormals, respectively. The orthonormality conditions are summarized by $e^\mu_{(\alpha)}e_{(\beta)\mu} = \eta_{\alpha\beta}$, where $\eta_{\alpha\beta}$ is the Lorentz metric: diag(1, −1, −1, −1). The vectors obey the Frenet–Serret equations

$$\begin{bmatrix} \dot{e}^\mu_{(0)} \\ \dot{e}^\mu_{(1)} \\ \dot{e}^\mu_{(2)} \\ \dot{e}^\mu_{(3)} \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 & 0 \\ \kappa & 0 & \tau_1 & 0 \\ 0 & -\tau_1 & 0 & \tau_2 \\ 0 & 0 & -\tau_2 & 0 \end{bmatrix} \begin{bmatrix} e^\mu_{(0)} \\ e^\mu_{(1)} \\ e^\mu_{(2)} \\ e^\mu_{(3)} \end{bmatrix} \tag{2a–2d}$$

The Frenet scalars κ, τ_1, and τ_2, which are termed the curvature, the first and the second torsions, respectively, completely characterize the intrinsic nature of the curve when given as functions of τ. The assignment of the initial configuration of the tetrad would further fix the curve uniquely in space–time. The sign of κ and τ_1 is taken to be positive, whereas that of τ_2 is fixed by the requirement that the triad $e^\mu_{(i)}$ be a right-handed one. We shall assume that all the three scalars are well defined unless otherwise stated (i.e., κ and τ_1 are both nonzero).

We shall now apply the Frenet–Serret formalism to charged particles in electromagnetic fields obeying the condition

$$F_{\mu\nu;\gamma} = 0 .$$

In other words, the field tensor $F_{\mu\nu}$ is assumed to be covariantly constant, which in the special case of flat space–times implies the constancy of electromagnetic field both in space and time. All the results we shall derive require only the weaker condition $\dot{F}_{\mu\nu} = 0$, but this would impose severe restrictions on the curve followed by the particle. We shall therefore assume the stronger, but more general, condition that $F_{\mu\nu}$ is a covariantly constant bivector. We shall refer to this property of $F_{\mu\nu}$ by the equivalent statement that the associated fields are homogeneous. We first proceed to prove the following theorem.

Theorem: The Frenet scalars remain constant along the world line of a charged particle in a homogeneous electromagnetic field. Further each of the Frenet vectors obeys the Lorentz equation.

Proof: From Eqs. (1) and (2a), we have

$$\dot{e}^\mu_{(0)} = F^\mu{}_\nu e^\nu_{(0)} = \kappa e^\mu_{(1)} . \tag{3}$$

Taking the absolute derivative of the above equation with respect to τ, we obtain

$$\dot{\kappa}e^\mu_{(1)} + \kappa\dot{e}^\mu_{(1)} = F^\mu{}_\nu\dot{e}^\nu_{(0)} .$$

Again using Eq. (2a)

$$\dot{\kappa}e^\mu_{(1)} + \kappa\dot{e}^\mu_{(1)} = \kappa F^\mu{}_\nu e^\nu_{(1)} . \tag{4}$$

Forming the scalar product of both sides of the above equation with $e_{(1)\mu}$, gives immediately

$$\dot{\kappa} = 0 ,$$

since

$$\dot{e}^\mu_{(1)}e_{(1)\mu} = 0 , \quad e^\mu_{(1)} \text{ being a unit vector} ,$$

and

$$F_{\mu\nu}e^\mu_{(1)}e^\nu_{(1)} = 0$$

by the antisymmetry of $F_{\mu\nu}$. Equation (4) then reduces to

$$\dot{e}^\mu_{(1)} = F^\mu{}_\nu e^\nu_{(1)} .$$

Thus $\kappa = \text{const}$ along the world line of the particle and $e^\mu_{(1)}$ in fact obeys the Lorentz equation. By repeatedly applying the above procedure to the other Frenet–Serret equations, we can show that τ_1 and τ_2 are constants and that $e^\mu_{(2)}$, $e^\mu_{(3)}$ also obey the Lorentz equation. To summarize,

$$\dot{\kappa} = \dot{\tau}_1 = \dot{\tau}_2 = 0$$

and

$$\dot{e}^\mu_{(\alpha)} = F^\mu{}_\nu e^\nu_{(\alpha)} . \tag{5}$$

This completes the proof of the theorem. We wish to emphasize that the properties of $F_{\mu\nu}$ that entered into the proof of the theorem were its antisymmetry and its constancy along the curve, viz. $\dot{F}_{\mu\nu} = 0$. We shall need this fact later when we draw the analogy between the present case and that of the Killing trajectories.

Next we shall evaluate the Frenet vectors and the scalars κ, τ_1, and τ_2 in terms of $F_{\mu\nu}$ and the 4-velocity $e^\mu_{(0)}$. For convenience we denote

$$(F^n)^\mu{}_\nu \equiv F^\mu{}_\alpha F^\alpha{}_\beta \cdots F^\sigma{}_\nu ,$$

with $F^\gamma{}_\rho$ repeated n times on the right-hand side. Further, it is easy to show that given an arbitrary vector q^μ if we define $p_\mu = (F^n)_{\mu\nu} q^\nu$, then

$$p_\mu p^\mu = (-1)^n (F^{2n})_{\mu\nu} q^\mu q^\nu . \tag{6}$$

We have already the relation

$$\kappa e^\mu_{(1)} = F^\mu{}_\nu e^\nu_{(0)} ,$$

or

$$e^\mu_{(1)} = \frac{1}{\kappa} F^\mu{}_\nu e^\nu_{(0)} . \tag{7a}$$

Squaring and using Eq. (6),

$$\kappa^2 = (F^2)_{\mu\nu} e^\mu_{(0)} e^\nu_{(0)} . \tag{7b}$$

In order to obtain τ_1 and $e^\mu_{(2)}$, we observe that Eqs. (5), (2b), and (7a) together lead to

$$\dot{e}^\mu_{(1)} = F^\mu{}_\nu e^\nu_{(1)} = \frac{1}{\kappa}(F^2)^\mu{}_\nu e^\nu_{(0)} = \kappa e^\mu_{(0)} + \tau_1 e^\mu_{(2)} ,$$

that is

$$\kappa e^\mu_{(0)} + \tau_1 e^\mu_{(2)} = \frac{1}{\kappa}(F^2)^\mu{}_\nu e^\nu_{(0)} .$$

The required expression for $e^\mu_{(2)}$ is obtained directly from the above equation, while by squaring the latter τ_1 can be readily found. So,

$$e^\mu_{(2)} = \frac{1}{\kappa\tau_1}[(F^2)^\mu{}_\nu - \kappa^2 \delta^\mu{}_\nu] e^\nu{}_{(0)} \tag{8a}$$

and

$$\tau_1^2 = \kappa^2 - \frac{1}{\kappa^2}(F^4)^\mu{}_\nu e^\nu_{(0)} e_{(0)\mu} . \tag{8b}$$

Similarly $e^\mu_{(3)}$ and τ_2 can be obtained by combining Eqs. (5), (8a), and (2c):

$$\dot{e}^\mu_{(2)} = F^\mu{}_\nu e^\nu_{(2)}$$

$$= \frac{1}{\kappa\tau_1} F^\mu{}_\nu [(F^2)^\nu{}_\alpha - \kappa^2 \delta^\nu{}_\alpha] e^\alpha_{(0)} = -\tau_1 e^\mu_{(1)} + \tau_2 e^\mu_{(3)} .$$

Equivalently, we can write

$$e^\mu_{(3)} = \frac{1}{\kappa\tau_1\tau_2}[(F^3)^\mu{}_\alpha + (\tau_1^2 - \kappa^2) F^\mu{}_\alpha] e^\alpha_{(0)} . \tag{9a}$$

Squaring and using Eq. (6), we find

$$\tau_2^2 = \frac{1}{\kappa^2\tau_1^2}(F^6)_{\mu\nu} e^\mu_{(0)} e^\nu_{(0)} - \frac{1}{\tau_1^2}(\kappa^2 - \tau_1^2)^2 . \tag{9b}$$

The above expressions can be employed to determine the tetrad components and the Frenet scalars whenever $e^\mu_{(0)}$ and $F_{\mu\nu}$ are known. However, we can simplify the formulas we have already derived and establish some interesting interrelations among these quantities which throw more light on the geometry of the world line. As a preliminary step towards this end we invoke some useful identities involving $F_{\mu\nu}$. First let us define

$$\alpha \equiv \tfrac{1}{2} F^\mu{}_\nu F^\nu{}_\mu = \lambda^2 (E^2 - H^2)$$

and

$$\beta \equiv \tfrac{1}{4} F^\mu{}_\nu \hat{F}^\nu{}_\mu = \lambda^2 (\mathbf{E} \cdot \mathbf{H})$$

which are the Lorentz invariants of the electromagnetic field with E and H denoting the electric and magnetic fields respectively. The tensor $\hat{F}_{\mu\nu}$, which is the dual of $F_{\mu\nu}$, is defined by

$$\hat{F}_{\mu\nu} = \tfrac{1}{2}\sqrt{-g}\,\epsilon_{\mu\nu\alpha\beta} F^{\alpha\beta}$$

and

$$\hat{F}^{\mu\nu} = \tfrac{1}{2}(1/\sqrt{-g})\epsilon^{\mu\nu\alpha\beta} F_{\alpha\beta} ,$$

where $g = \det \| g_{\alpha\beta} \|$ and $\epsilon_{\alpha\beta\gamma\delta}$ is the completely antisymmetric Levi–Civita symbol with $\epsilon^{0123} = 1 = -\epsilon_{0123}$. Further, we note $\hat{\hat{F}}^{\mu\nu} = -F^{\mu\nu}$. Now, two arbitrary bivectors $A_{\alpha\beta}$ and $B_{\alpha\beta}$ along with their duals (denoted by hats above these tensors) defined on a four-dimensional Riemannian space satisfy the identity[4]

$$A^\mu{}_\alpha B^\alpha{}_\nu - \hat{B}^\mu{}_\alpha \hat{A}^\alpha{}_\nu = \tfrac{1}{2}\delta^\mu{}_\nu (A_{\gamma\delta} B^{\delta\gamma}) .$$

Identifying $A_{\mu\nu} = B_{\mu\nu} = F_{\mu\nu}$ gives

$$(F^2)^\mu{}_\nu - (\hat{F}^2)^\mu{}_\nu = \alpha\, \delta^\mu{}_\nu , \tag{10a}$$

whereas making $A_{\mu\nu} = F_{\mu\nu}$ and $B_{\mu\nu} = \hat{F}_{\mu\nu}$ and vice versa leads to

$$F^\mu{}_\alpha \hat{F}^\alpha{}_\nu = \hat{F}^\mu{}_\alpha F^\alpha{}_\nu = \beta \delta^\mu{}_\nu . \tag{10b}$$

With the help of these two identities it is a straightforward process to derive three further identities:

$$(F^3)^\mu{}_\nu - \alpha F^\mu{}_\nu - \beta \hat{F}^\mu{}_\nu = 0 , \tag{10c}$$

$$(F^4)^\mu{}_\nu - \alpha (F^2)^\mu{}_\nu - \beta^2 \delta^\mu{}_\nu = 0 , \tag{10d}$$

$$(F^6)^\mu{}_\nu - (\alpha^2 + \beta^2)(F^2)^\mu{}_\nu - \alpha\beta^2 \delta^\mu{}_\nu = 0 . \tag{10e}$$

We are now ready to apply the above identities to the already derived results. Substituting for $(F^4)^\mu{}_\alpha$ from Eq. (10d) in Eq. (8b), we obtain

$$\tau_1^2 = \kappa^2 - \alpha - \beta^2/\kappa^2 . \tag{11}$$

The curvature κ depends on $F_{\mu\nu}$ as well as on $e^\mu_{(0)}$. Once it has been determined from the rather simple expression of Eq. (7b), τ_1 can be evaluated at once with the help of the invariants α and β which are free from the parameters associated with the curve. This is true in the case of the second torsion τ_2 also. A vastly simplified form for τ_2 results if $(F^6)^\mu{}_\nu$ in Eq. (9b) is substituted for from the identity (10e). Thus we have

$$\tau_2^2 = \frac{1}{\kappa^2\tau_1^2}[(\alpha^2 + \beta^2)(F^2)^\mu{}_\nu + \alpha\beta^2 \delta^\mu{}_\nu] e_{(0)}{}^\nu e_{(0)\mu} - \frac{1}{\tau_1^2}(\kappa^2 - \tau_1^2)^2 .$$

The formulas for κ^2 and $(\kappa^2 - \tau_1^2)$ from Eqs. (7b) and (11) reduce the above equation to the extremely simple form

$$\tau_2^2 = \beta^2/\kappa^2 \quad \text{or} \quad \kappa^2\tau_2^2 = \beta^2 .$$

As we pointed out earlier, while κ and τ_1 are assumed to be positive, the sign of τ_2 is fixed by the demand that $e^\mu_{(i)}$ form a right-handed frame. We shall show in the next section that this requirement implies that

$$\kappa\tau_2 = -\beta . \tag{12}$$

Again from Eqs. (11) and (12) we obtain

$$\kappa^2 - \tau_1^2 - \tau_2^2 = \alpha . \tag{13}$$

Equations (12) and (13) relate the Frenet scalars to the field invariants. Apart from displaying the connection between the geometry of the particle's world line on one hand and the electromagnetic field acting on it on the

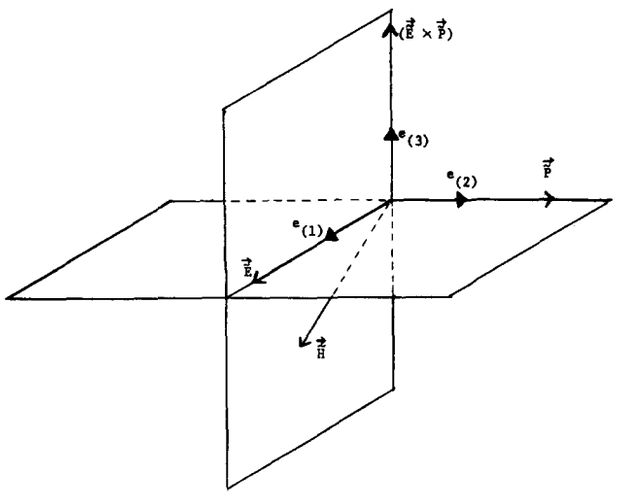


FIG. 1. Rest frame configuration of a positively charged particle ($\lambda = e/mc^2 > 0$). The spatial Frenet vectors are given by $\mathbf{e}_{(1)} = \mathbf{E}/E$, $\mathbf{e}_{(2)} = \mathbf{P}/P$, and $\mathbf{e}_{(3)} = \mathbf{E}\times\mathbf{P}/EP$, where $\mathbf{P} = (c/4\pi) \times (\mathbf{E}\times\mathbf{H})$.

other, the above relations can also serve as a simple means of obtaining τ_1 and τ_2 starting from κ, provided of course κ, τ_1, and τ_2 are all well defined.[5] Another piece of information that elucidates the interplay between the geometry and the physics of the situation is provided by the energy–momentum tensor,

$$T^{\mu}{}_{\nu} = (1/4\pi)[F^{\mu}{}_{\alpha}F^{\alpha}{}_{\nu} - \tfrac{1}{4}(F^{\alpha}{}_{\beta}F^{\beta}{}_{\alpha})\delta^{\mu}{}_{\nu}].$$

By using Eqs. (8a) and (13), we arrive at

$$T^{\mu}{}_{\nu}e^{\nu}_{(0)} = (1/4\pi)[\tfrac{1}{2}(\kappa^2 + \tau_1^2 + \tau_2^2)e^{\mu}_{(0)} + \kappa\tau_1 e^{\mu}_{(2)}]. \tag{14}$$

Clearly, the above equation implies that in the instantaneous rest system of the particle the energy density of the field is given by $\epsilon = (1/8\pi)(\kappa^2 + \tau_1^2 + \tau_2^2)$ and the flux of energy density or the Poynting vector is purely in the direction of $\mathbf{e}_{(2)}$ and has the magnitude $\kappa\tau_1$.

Finally, consider Eq. (9a) for $e^{\mu}_{(3)}$. Replacing $(F^3)^{\mu}{}_{\nu}$ through identity (10c) and further employing Eqs. (12) and (13), we get

$$-\hat{F}^{\mu}{}_{\nu}e^{\nu}_{(0)} = \tau_1 e^{\mu}_{(3)} + \tau_2 e^{\mu}_{(1)} \equiv \omega^{\mu}. \tag{15}$$

Firstly, this equation relates the magnetic field in the instantaneous rest frame of the particle to the Frenet vectors $e^{\mu}_{(3)}$ and $e^{\mu}_{(1)}$. Secondly, a Fermi transported vector undergoes rotation with respect to the Frenet triad with the angular velocity vector $-\omega^{\mu}$ given above.[6] Or since a Fermi transported frame is the relativistic analog of a nonrotating frame in Newtonian physics,[7,8] the Frenet frame can be considered to rotate with respect to a nonrotating frame at a rate given by ω^{μ}; evidently the vector $-\omega^{\mu}$ can be identified with the rest frame magnetic field.

Towards the end of the foregoing discussions we have alluded to the instantaneous rest frame of the particle. We shall present more detailed considerations pertaining to this frame in the next section.

III. SPECIALIZATION TO THE INSTANTANEOUS REST FRAME OF THE PARTICLE

In the instantaneous rest frame of the particle $e^{\mu}_{(0)} = \delta^{\mu}{}_{0}$. Consequently, the members of the triad $e^{\mu}_{(i)}$ will not have time components, since they are orthogonal to $e^{\mu}_{(0)}$. The electric and magnetic fields, $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$, in the rest system are given by the equations

$$F^{\mu}{}_{\nu}e^{\nu}_{(0)} = \tilde{E}^{\mu}, \quad \hat{F}^{\mu}{}_{\nu}e^{\nu}_{(0)} = \tilde{H}^{\mu}. \tag{16}$$

In the above equations, $F_{\mu\nu}$ and $\hat{F}_{\mu\nu}$ are of course evaluated in the rest system. It is evidently a simple and straightforward procedure to translate all the results derived in the last section to their equivalents in the instantaneous rest system of the charged particle.

Remembering that κ and τ_1 have been chosen to be positive, we get directly from Eqs. (7), (11), and (8),

$$\kappa = |\lambda|\tilde{E}, \quad \mathbf{e}_{(1)} = \frac{\lambda}{|\lambda|}\frac{\tilde{\mathbf{E}}}{\tilde{E}},$$

and

$$\tau_1 = |\lambda|\frac{|\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}|}{\tilde{E}}, \quad \mathbf{e}_{(2)} = \frac{\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}}{|\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}|} = \frac{\tilde{\mathbf{P}}}{\tilde{P}} \tag{17}$$

where $\tilde{\mathbf{P}} = (c/4\pi)\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}$ is the Poynting vector. The fact that the Poynting vector is in the direction of $\mathbf{e}_{(2)}$ in the rest system, with its magnitude equal to $\kappa\tau_1$, had been anticipated towards the end of the last section.

We shall now proceed to examine the sign of τ_2. Assuming that $\kappa\tau_2 = -\beta$, we derived Eq. (15). That equation, in the rest system, reads

$$-\lambda\tilde{\mathbf{H}} = \tau_1\mathbf{e}_{(3)} + \tau_2\mathbf{e}_{(1)}. \tag{18}$$

This is indeed consistent with $\mathbf{e}_{(1)}$, $\mathbf{e}_{(2)}$, and $\mathbf{e}_{(3)}$ forming a right-handed system as can be seen by taking the cross product of the above equation with $\lambda\tilde{\mathbf{E}} = \kappa\mathbf{e}_{(1)}$:

$$-\lambda^2\tilde{\mathbf{E}}\times\tilde{\mathbf{H}} = -\kappa\tau_1\mathbf{e}_{(2)}$$

which agrees with Eq. (17). Thus,

$$\tau_2 = -|\lambda|\tilde{\mathbf{E}}\cdot\tilde{\mathbf{H}}/\tilde{E}. \tag{19}$$

Equation (18) shows that $\mathbf{e}_{(3)}$ lies in the plane of $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{H}}$; the rest system magnetic field in properly chosen units (taking into account the coefficient λ) has components $-\tau_2$ and $-\tau_1$ along $\mathbf{e}_{(1)}$ and $\mathbf{e}_{(3)}$, respectively. Similarly in the same units $\tilde{\mathbf{E}}$ has component κ along $\mathbf{e}_{(1)}$. As was pointed out in the last section $-\lambda\tilde{\mathbf{H}}$ coincides with the angular velocity vector that measures the rate of rotation of the Frenet triad with respect to a Fermi transported spatial frame.

The rest frame configuration of a positively charged particle ($\lambda = e/mc^2 > 0$) is shown in Fig. 1. It displays the triad

$$\mathbf{e}_{(1)} = \frac{\lambda}{|\lambda|}\frac{\tilde{\mathbf{E}}}{\tilde{E}}, \quad \mathbf{e}_{(2)} = \frac{\tilde{\mathbf{P}}}{\tilde{P}}, \quad \mathbf{e}_{(3)} = \frac{\lambda}{|\lambda|}\frac{\tilde{\mathbf{E}}\times\tilde{\mathbf{P}}}{\tilde{E}\tilde{P}},$$

$$\tilde{\mathbf{P}} = \frac{c}{4\pi}\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}.$$

Since the particle is initially at rest here, it will get its first impulse in the $\tilde{\mathbf{E}}$ direction, i.e., along $\mathbf{e}_{(1)}$, thereby acquiring an infinitesimal velocity $\mathbf{V}_1$ along $\mathbf{e}_{(1)}$. The second impulse felt by the particle is due to the force $(e/c)\mathbf{V}_1\times\tilde{\mathbf{H}}$ along $\mathbf{e}_{(2)}$. The velocity component thus acquired in the direction of $\mathbf{e}_{(2)}$ couples with $\tilde{\mathbf{H}}$ to impel the particle out of the $\mathbf{e}_{(1)}-\mathbf{e}_{(2)}$ plane.

We summarize below the Frenet scalars evaluated in the rest system:

$$\kappa = |\lambda|\tilde{E}, \quad \tau_1 = |\lambda| \frac{|\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}|}{\tilde{E}}, \quad \tau_2 = -|\lambda| \frac{\tilde{\mathbf{E}}\cdot\tilde{\mathbf{H}}}{\tilde{E}}. \tag{20}$$

The formalism we have developed would equally well apply to magnetic monopoles, if they exist, provided we replace $F_{\mu\nu}$ by its dual $\hat{F}_{\mu\nu}$, and λ by $\hat{\lambda} = \hat{e}/mc^2$ where $\hat{e}$ is the magnetic charge of the monopole. The resulting Frenet scalars in the rest system of the monopole would be

$$\kappa = |\hat{\lambda}|\tilde{H}, \quad \tau_1 = |\hat{\lambda}| \frac{|\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}|}{\tilde{H}}, \quad \tau_2 = |\lambda| \frac{\tilde{\mathbf{E}}\cdot\tilde{\mathbf{H}}}{\tilde{H}}. \tag{21}$$

Comparing the Eqs. (20) and (21), we notice that in both cases the scalars are independent of the sign of the charge concerned. This means that we cannot distinguish intrinsically between the space–time trajectories of positive and negative charges of equal mass. In order to differentiate between these charges we must also specify some external information namely the direction of the arrow of time. However, the motion of a negative charge is the same as that of a positive charge moving backwards in time. This is a reflection of the fact that the Lorentz equation is invariant under the simultaneous reversal in sign of the proper time and the charge. Nevertheless, τ_2 changes sign upon passing from an electric charge to a magnetic monopole, thereby making the intrinsic distinction possible between the two. Evidently, underlying this distinguishability is the tacit assumption that one can define positively oriented reference frames; in other words, space–time is orientable.

We shall now establish the relevance of the Frenet frame to the phenomenon of spin precession. Let us assume that the charged particle of mass m and charge e, has a spin S and a consequent magnetic moment $\mathbf{M} = (ge_0/2mc)\mathbf{S}$, where g is the gyromagnetic ratio of the particle and e_0 is the electronic charge. We shall not go into a detailed discussion of the covariant dynamical description of spin precession at this point.[9] It is sufficient for our purpose to note that, in the instantaneous rest system of the particle, spin precession is described by the familiar equation

$$\frac{d\mathbf{S}}{d\tau} = \frac{ge_0}{2mc^2}\mathbf{S}\times\tilde{\mathbf{H}}. \tag{22}$$

At this stage we have made no assumptions about the spatial frame to which S is referred, nor about the transport of such a frame along the world line. Choosing this frame to be the instantaneous Frenet triad $\mathbf{e}_{(i)}$, we can write

$$\mathbf{S} = S^{(i)}\mathbf{e}_{(i)}. \tag{23}$$

Substituting Eq. (23) in (22), we find

$$\frac{dS^{(i)}}{d\tau}\mathbf{e}_{(i)} + S^{(i)}\frac{d\mathbf{e}_{(i)}}{d\tau} = \frac{ge_0}{2mc^2}S^{(i)}\mathbf{e}_{(i)}\times\tilde{\mathbf{H}}. \tag{24}$$

In the space orthogonal to $e^{\mu}_{(0)} = \delta^{\mu}_{0}$, we have the Frenet equations

$$\frac{d}{d\tau}\begin{bmatrix}\mathbf{e}_{(1)}\\ \mathbf{e}_{(2)}\\ \mathbf{e}_{(3)}\end{bmatrix} = \begin{bmatrix}0 & \tau_1 & 0\\ -\tau_1 & 0 & \tau_2\\ 0 & -\tau_2 & 0\end{bmatrix}\begin{bmatrix}\mathbf{e}_{(1)}\\ \mathbf{e}_{(2)}\\ \mathbf{e}_{(3)}\end{bmatrix}. \tag{25}$$

Substituting the above in Eq. (24) and defining the modified gyromagnetic ratio $g' = (e_0/e)g$, we arrive at the equation

$$\dot{\mathbf{S}} = \boldsymbol{\Omega}\times\mathbf{S}, \quad \dot{\mathbf{S}} = \frac{dS^{(i)}}{d\tau}\mathbf{e}_{(i)} \tag{26}$$

where $\boldsymbol{\Omega} = -[(g'/2)\lambda\tilde{\mathbf{H}} + \boldsymbol{\omega}]$ and $\boldsymbol{\omega} = \tau_2\mathbf{e}_{(1)} + \tau_1\mathbf{e}_{(3)}$.

Assume now that the charged particle is in a homogeneous electromagnetic field and that its motion is governed by the Lorentz equation. We have ignored the fact that the spin-electromagnetic field interaction would in itself result in a deviation from the Lorentz force. Then, as we have seen [Eq. (18)], $\boldsymbol{\omega} = -\lambda\tilde{\mathbf{H}}$ and, consequently, $\boldsymbol{\Omega} = (1 - g'/2)\lambda\tilde{\mathbf{H}}$. We see immediately that if $g' = 2$ then $\boldsymbol{\Omega} = 0$ and hence the spin (consequently, the magnetic moment) does not precess with respect to the Frenet triad.[10] Its components along the Frenet vectors will remain constant. Or, if a frame can be picked that is attached to the moving charge and in which the magnetic moment has constant components, then such a frame could differ from the Frenet frame by at most a constant rotation. Thus, as we stated in the introduction, the phenomenon of spin precession affords a direct physical realization of the Frenet frame up to, of course, a constant rotation of the spatial triad.

We have assumed all along our development of the formalism that κ, τ_1, and τ_2 are well defined, i.e., at least $\kappa \neq 0$ and $\tau_1 \neq 0$. From the expressions for these scalars in the rest system [Eq. (20)], it can be seen that this stipulation can be satisfied by taking $\tilde{\mathbf{E}}\times\tilde{\mathbf{H}} \neq 0$, a fact which we shall need in the next section. A detailed discussion of the bearing of field configurations on the Frenet scalars and hence on the intrinsic geometry of the world lines can be found in Ref. 9b.

IV. DUALITY ROTATIONS

Duality rotations[2,3] can transform a given electromagnetic field to new ones that are related to the former by simple equalities. In this section we briefly explore the effect of such transformations on the geometry of the charged particle trajectories.

Consider electric and magnetic fields, E and H, that are solutions to source free Maxwell's equations. If we define $\mathbf{K} = \mathbf{E} + i\mathbf{H}$, duality rotations are represented by

$$\mathbf{K}' = e^{i\phi}\mathbf{K}, \tag{27}$$

where ϕ is a real constant. Obviously, both K and K′ satisfy the source free Maxwell's equations. The Lorentz field invariants $\mathbf{E}\cdot\mathbf{H}$ and $(E^2 - H^2)$ are not invariant under these transformations. Nevertheless two new invariants are admitted here, namely

$$\mathbf{K}'\cdot\mathbf{K}'^{*} = \mathbf{K}\cdot\mathbf{K}^{*} \quad \text{or} \quad E'^2 + H'^2 = E^2 + H^2 \tag{28a}$$

and

$$\mathbf{K}'\times\mathbf{K}'^{*} = \mathbf{K}\times\mathbf{K}^{*} \quad \text{or} \quad \mathbf{E}'\times\mathbf{H}' = \mathbf{E}\times\mathbf{H}, \tag{28b}$$

where the asterisk indicates complex conjugation. With this preamble we establish the following theorem:

Theorem: Let the initial velocity $\mathbf{v}_0$ of a charged particle in an arbitrary homogeneous electromagnetic field be such that κ, τ_1, and τ_2 are all well defined (i.e., $\kappa \neq 0$, $\tau_1 \neq 0$ or equivalently $\mathbf{E}\times\mathbf{H} \neq 0$ in the rest frame).

If we perform an arbitrary duality rotation on the electromagnetic field, and if the particle has the same initial velocity v_0 in the newly generated field, then the new scalars κ', τ_1', and τ_2' will also be well defined.

Proof: We note that Lorentz transformations and duality rotations acting on fields commute. In what follows we use the same duality transformation (same "angle" ϕ) and the same Lorentz transformation (corresponding to a velocity $-v_0$ that produces the instantaneous rest frame of the charged particle) acting at the initial space–time point of the particle. We first perform the duality rotation ($\mathbf{E}\to\mathbf{E}'$, $\mathbf{H}\to\mathbf{H}'$) followed by the Lorentz transformation ($\mathbf{E}'\to\tilde{\mathbf{E}}'$, $\mathbf{H}'\to\tilde{\mathbf{H}}'$). Reversing the order of the operations we have,

$$(\mathbf{E},\mathbf{H})\to(\tilde{\mathbf{E}},\tilde{\mathbf{H}})\to[(\tilde{\mathbf{E}})',(\tilde{\mathbf{H}})'].$$

Since the two types of transformations commute we conclude $\tilde{\mathbf{E}}'=(\tilde{\mathbf{E}})'$ and $\tilde{\mathbf{H}}'=(\tilde{\mathbf{H}})'$. On the other hand, since $\mathbf{E}\times\mathbf{H}$ is invariant under duality rotations,

$$\tilde{\mathbf{E}}\times\tilde{\mathbf{H}}=(\tilde{\mathbf{E}})'\times(\tilde{\mathbf{H}})'=\tilde{\mathbf{E}}'\times\tilde{\mathbf{H}}'. \tag{29}$$

Thus, $\tilde{\mathbf{E}}'\times\tilde{\mathbf{H}}'\neq 0$ and therefore κ', τ_1', τ_2' are well defined at the initial point and consequently all along the curve since these scalars are constants. This completes the proof.

Equations (29) and (20) yield the relation

$$\kappa'\tau_1'=\kappa\tau_1. \tag{30a}$$

Similarly with the help of Eqs. (28a) and (14), we can show that

$$\kappa'^2+\tau_1'^2+\tau_2'^2=\kappa^2+\tau_1^2+\tau_2^2. \tag{30b}$$

Further relations are obtained by noting that $K'^2=e^{2i\phi}K^2$ which, with the help of Eqs. (12) and (13), leads to

$$\begin{bmatrix}\kappa'^2-\tau_1'^2-\tau_2'^2\\ 2\kappa'\tau_2'\end{bmatrix}=\begin{bmatrix}\cos2\phi & -\sin2\phi\\ \sin2\phi & \cos2\phi\end{bmatrix}\begin{bmatrix}\kappa^2-\tau_1^2-\tau_2^2\\ 2\kappa\tau_2\end{bmatrix}. \tag{30c}$$

Equation (30a) through (30c) can easily be uncoupled so that each of the scalars κ', τ_1', τ_2' is expressed in terms of κ, τ_1, τ_2, and ϕ.

V. KILLING TRAJECTORIES

There exists a striking similarity between the world lines of charged particles in homogeneous fields and Killing trajectories admitted by four-dimensional Riemannian spaces. Consider such a timelike Killing vector ξ^μ. We can define then the 4-velocity along ξ^μ by

$$u^\mu=e^{\psi}\xi^\mu,\quad e^{-2\psi}=\xi^\nu\xi_\nu. \tag{31}$$

It is easy to show that e^{ψ} or, equivalently, ψ is constant along the Killing trajectory, i.e., $\psi_{,\mu}\xi^\mu=0$, from the Killing equation:

$$\xi_{\mu;\nu}+\xi_{\nu;\mu}=0. \tag{32}$$

So

$$\dot{u}^\mu=e^{\psi}\xi^\mu{}_{;\nu}u^\nu. \tag{33}$$

We can identify $e^{\psi}\xi_{\mu;\nu}$ with an antisymmetric tensor $F_{\mu\nu}$ by virtue of Eq. (32),

$$F_{\mu\nu}\equiv e^{\psi}\xi_{\mu;\nu}, \tag{34}$$

so that Eq. (33) is formally the same as the Lorentz equation. Further, since $e^{\psi}=\text{const}$ along ξ^μ, we have

$$\dot{F}_{\mu\nu}=e^{2\psi}\xi_{\mu;\nu;\gamma}\xi^\gamma=0 \tag{35}$$

since $\xi_{\mu;\nu;\gamma}=R_{\mu\nu\gamma\lambda}\xi^\lambda$, where $R_{\mu\nu\gamma\lambda}$ is the Riemann tensor. As we emphasized in Sec. II, the properties $F_{\mu\nu}=-F_{\nu\mu}$ and $\dot{F}_{\mu\nu}=0$ were sufficient to derive all the results of that section. Those results, then, are valid in the case of the Killing trajectories with suitable changes in interpretation whenever necessary.

We concentrate attention on some of the significant results that emerge. To begin with, the Frenet scalars are constants along the Killing trajectories. Equation (15) now becomes

$$\omega^\mu=\tau_1e^\mu_{(3)}+\tau_2e^\mu_{(1)}=-e^{2\psi}\hat{\xi}^{\mu;\nu}\xi_\nu, \tag{36}$$

where the dual $\hat{\xi}^{\mu;\nu}=(1/2\sqrt{-g})\epsilon^{\mu\nu\alpha\beta}\xi_{\alpha;\beta}$ as before.

We define

$$\bar{\omega}^\mu\equiv\hat{\xi}^{\mu;\nu}\xi_\nu=(1/2\sqrt{-g})\epsilon^{\mu\nu\alpha\beta}\xi_\nu\xi_{\alpha;\beta}. \tag{37}$$

In the above equation $\bar{\omega}^\mu$ is the rotation vector of the Killing congruence. And also $e^{2\psi}\hat{\xi}^{\mu;\nu}\xi_\nu=(1/2\sqrt{-g})\epsilon^{\mu\nu\alpha\beta}\times u_\nu u_{\alpha;\beta}$ the rotation vector associated with the congruence of Killing world lines. In the present context these vectors are evaluated along a particular curve belonging to the congruence. We have already pointed out in Sec. II that ω^μ gives the rotation of the Frenet frame with respect to the Fermi transported frame. In the case of Killing trajectories, we have arrived at the interesting result that ω^μ is directly proportional to the rotation vector of the Killing congruence $\bar{\omega}^\mu$. Obviously, in static space–times, that are characterized by $\bar{\omega}^\mu=0$, the Frenet frame is a nonrotating frame defined along the Killing line.

Next consider the Killing analog of Eq. (13) which reads

$$\kappa^2-\tau_1{}^2-\tau_2^2=-(e^{2\psi}/2)\xi_{\mu;\nu}\xi^{\mu;\nu}. \tag{38}$$

In order to interpret this equation, define the family of hypersurfaces,

$$\Sigma:\ \xi^\mu\xi_\mu=\text{const}. \tag{39}$$

The normal to any of these surfaces is given by

$$n_\mu=(\xi^\nu\xi_\nu)_{,\mu}=2\xi_{\nu;\mu}\xi^\nu. \tag{40}$$

By the antisymmetry of $\xi_{\alpha;\beta}$ we find $n_\mu\xi^\mu=0$. This shows that the Killing trajectories having the same length $\xi^\mu\xi_\mu$ (say c_1) lie in the corresponding member of the family Σ characterized by the constant c_1. For the 4-velocities along these trajectories, we have

$$\dot{u}_\mu=e^{2\psi}\xi_{\mu;\nu}\xi^\nu=\kappa e_{(1)\mu}=-\tfrac{1}{2}e^{2\psi}n_\mu. \tag{41}$$

Or,

$$\kappa e_{(1)\mu}=-\tfrac{1}{2}e^{2\psi}n_\mu,\quad \kappa^2=-\tfrac{1}{4}e^{4\psi}n_\mu n^\mu. \tag{42}$$

Further, from Eqs. (36) and (37),

$$-\bar{\omega}^\mu e^{2\psi}=\tau_1e^\mu_{(3)}+\tau_2e^\mu_{(1)},\quad -(\tau_1^2+\tau_2^2)=e^{4\psi}\bar{\omega}^\mu\bar{\omega}_\mu. \tag{43}$$

With the help of Eqs. (42) and (43), Eq. (38) reduces to

$$\bar{\omega}^\mu \bar{\omega}_\mu - \tfrac{1}{4} n_\mu n^\mu = -\tfrac{1}{2}\xi^\nu \xi_\nu (\xi_{\alpha;\beta}\xi^{\alpha;\beta}) . \tag{44}$$

This is an equation that has black hole physics as its usual habitat.[11] It shows that the surface on which ξ_μ becomes null is a null surface (the horizon or the black hole), provided $\bar{\omega}^\mu$ also becomes null. In the case of static metrics ($\bar{\omega}^\mu = 0$), e.g. Schwarzschild metric, this automatically occurs, but not necessarily in stationary space-times ($\bar{\omega}^\mu \neq 0$). In the case of Kerr metric, for instance, the condition is not satisfied for the global time-like ξ^μ (which is time-like at spatial infinity); nevertheless the combination $\xi^\mu + c\eta^\mu$, where η^μ is the axial Killing vector and c is a suitably chosen constant does satisfy the condition making the Kerr black hole in fact a Killing horizon.[12]

Equation (12) translates into

$$\kappa \tau_2 = \tfrac{1}{4}\xi_{\mu;\nu}\hat{\xi}^{\mu;\nu} . \tag{45}$$

If $\tau_2 = 0$ then the above equation shows that $\xi_{\mu;\nu}$ is a simple bivector. We may mention that this holds true in the case of the combined Killing vector $\xi^\mu + c\eta^\mu$ alluded to in connection with the Kerr black hole.

Finally, consider the Lorentz equation for a charged particle in flat space–time

$$\frac{du_\mu}{d\tau} = F_{\mu\nu}\frac{dx^\nu}{d\tau} . \tag{46}$$

If $F_{\mu\nu}$ is constant, Eq. (46) can be directly integrated to give

$$u_\mu(\tau) = F_{\mu\nu}x^\nu(\tau) + \alpha_\mu . \tag{47}$$

Here $x^\nu(\tau)$ is the parametric form of the charge trajectory with the proper time τ as the parameter and α_μ is constant. On the other hand, the most general form of a Killing vector in the flat space–time is given by

$$\xi_\mu = \beta_{\mu\nu}x^\nu + \gamma_\mu , \tag{48}$$

where $\beta_{\mu\nu}$ is constant and antisymmetric ($\beta_{\mu\nu} = -\beta_{\nu\mu}$) and γ_μ is also constant. Therefore, $\xi^\mu/\sqrt{\xi^\nu\xi_\nu}$ can be identified with u^μ, provided $x^\nu = x^\nu(\tau)$ in Eq. (48). This shows that trajectories of charged particles acted upon by constant electromagnetic fields in flat space–time coincide with timelike Killing lines. Probably this is not true in the case of arbitrary Riemannian spaces, although at present we do not possess any theorems that point one way or the other.

After submitting the present paper for publication we have learned that D. M. Eardley has studied space–times admitting constant bivectors.[13]

APPENDIX

The usual method of describing the motion of a charged particle in an electromagnetic field is through a parameter representation of the particle's world line. The initial conditions give rise to eight constants of integration, but the side condition

$$u^\mu u_\mu = 1 \quad \text{where } u^\mu = \frac{dx^\mu}{d\tau} , \quad \mu = 0, \ldots, 3 ,$$

reduces this number to seven. It is preferable to have a representation that displays these transformation properties explicitly. Unfortunately, the usual procedure for treating the motion of charged particles in electromagnetic fields is to specialize the fields.[14] Then the transformation properties of the constants of integration are effectively masked. In the following we avoid this by presenting a general covariant solution to the Lorentz equation for homogeneous electromagnetic fields in flat space–time. This solution, and alternate solutions given elsewhere are quite cumbersome.[15] This is in sharp contrast with the elegant results that are obtained through the Frenet–Serret formalism.

The formal solution of $\dot{u} = Fu$ when $\dot{F} = 0$ in flat space–time is

$$u(\tau) = e^{F\tau}u(0) . \tag{A1}$$

In the above F stands for the matrix $F^\mu{}_\nu$. The series operator $e^{F\tau}$ can be simplified through the recursion relation

$$F^4 - \alpha F^2 - \beta^2 I = 0 \tag{A2}$$

where

$$\alpha \equiv \lambda^2(E^2 - H^2) \quad \text{and} \quad \beta \equiv \lambda^2(\mathbf{E}\cdot\mathbf{H}) .$$

This method, however, is too complicated to use, especially if both α and β are nonzero.

The simplest approach is to operate on u with both sides of (A2). Since $du/d\tau = Fu$ together with $\dot{F} = 0$ imply that $d^n u/d\tau^n = F^n u$ for all integral n, it follows that

$$\dddot{u} - \alpha\ddot{u} - \beta^2 u = 0 . \tag{A3}$$

This is a simple linear equation whose solution is

$$u(\tau) = \frac{1}{\lambda^2|K^2|}\left[\begin{array}{l} [\alpha_2^2\cosh\alpha_1\tau + \alpha_1^2\cos\alpha_2\tau]I + \left[\alpha_2^2\dfrac{\sinh\alpha_1\tau}{\alpha_1} + \alpha_1^2\dfrac{\sin\alpha_2\tau}{\alpha_2}\right]F \\ \\ + [\cosh\alpha_1\tau - \cos\alpha_2\tau]F^2 + \left[\dfrac{\sinh\alpha_1\tau}{\alpha_1} - \dfrac{\sin\alpha_2\tau}{\alpha_2}\right]F^3 \end{array}\right] u(0) , \tag{A4}$$

where $\mathbf{K} = \mathbf{E} + i\mathbf{H}$ and

$$\alpha_1 \equiv \lambda(|K^2| + \mathrm{Re}K^2/2)^{1/2} , \quad \alpha_2 \equiv \lambda(|K^2| - \mathrm{Re}K^2/2)^{1/2} .$$

It follows that

$$x(\tau)=x(0)+\frac{1}{\lambda^2|K^2|}\left[\begin{array}{l}\left[\alpha_2^2\frac{\sinh\alpha_1\tau}{\alpha_1}+\alpha_1^2\frac{\sin\alpha_2\tau}{\alpha_2}\right]I+\left[\alpha_2^2\left(\frac{\cosh\alpha_1\tau-1}{\alpha_1^2}\right)-\alpha_1^2\left(\frac{\cos\alpha_2\tau-1}{\alpha_2^2}\right)\right]F\\+\left[\frac{\sinh\alpha_1\tau}{\alpha_1}-\frac{\sin\alpha_2\tau}{\alpha_2}\right]F^2+\left[\left(\frac{\cosh\alpha_1\tau-1}{\alpha_1^2}\right)+\left(\frac{\cos\alpha_2\tau-1}{\alpha_2^2}\right)\right]F^3\end{array}\right]u(0)\,. \tag{A5}$$

Various limiting cases in which $\mathbf{E}\cdot\mathbf{H}=0$ or $E^2-H^2=0$ or both $\mathbf{E}\cdot\mathbf{H}=E^2-H^2=0$ can be evaluated. In this regard we note that if $\beta=\lambda^2\mathbf{E}\cdot\mathbf{H}=0$, then F obeys

$$F^3-\alpha F=0\,. \tag{A6}$$

Note added in proof: We have discovered that J. L. Synge, Proc. Roy. Irish Acad. A **65**, 27 (1967), has considered some geometrical properties of flat space–time trajectories of charged particles in homogeneous electro-magnetic fields.

ACKNOWLEDGMENTS

We thank Dr. R.D. Greene for helpful discussions. One of us (E.H.) would like also to thank Professors Peter Havas and Arthur Komar for useful comments and suggestions.

*Supported in part by the National Science Foundation Grant No. GP-36316X.

†Present address: Department of Physics, Scarborough College, University of Toronto, Westhill, Ontario, Canada.

[1]J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 10. J. L. Synge, "Relativity Groups and Topology," *Les Houches* 1963, edited by C. deWitt and B. deWitt (Gordon and Breach, New York, 1964), p. 21 ff. F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965), p. 173 ff., has utilized the Frenet–Serret formalism in ordinary 3-space to investigate the trajectories of relativistic charged particles in electromagnetic fields.

[2]G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).

[3]C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525 (1957).

[4]H. S. Ruse, Proc. London Math. Soc. **41**, No. 2, 302 (1936).

[5]In an orthonormal frame at any point of the curve

$$\kappa^2=\lambda^2\gamma^2\left[\left(\mathbf{E}+\frac{\mathbf{v}}{c}\times\mathbf{H}\right)^2-\left(\frac{\mathbf{v}}{c}\cdot\mathbf{E}\right)^2\right],$$

where $\mathbf{v}$ is the velocity of the particle and $\gamma=(1-v^2/c^2)^{-1/2}$. Then τ_1 and τ_2 are obtained simply from

$$\tau_1^2=\kappa^2-\lambda^2(E^2-H^2)-\lambda^4[(\mathbf{E}\cdot\mathbf{H})^2/\kappa^2] \quad\text{and}\quad \tau_2^2=\lambda^4[(\mathbf{E}\cdot\mathbf{H})^2/\kappa^2]\,.$$

Since κ, τ_1, and τ_2 are constants, once we are given the fields and the initial velocity $\mathbf{v}_0$, we can automatically describe the curve.

[6]J. L. Synge, *Relativity: General Theory*, pp. 139–140.

[7]F. A. E. Pirani, Acta Phys. Polonica **15**, 389 (1956).

[8]A. Walker, Proc. Edinb. Math. Soc. **4** (2), 170 (1935).

[9]See, for example, V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959). For further discussions and references on this topic see E. Honig, Ph. D. thesis, New York University, 1973.

[10]M. Fierz and V. L. Telegdi, *Quanta-Essays in Theoretical Physics*, edited by P. G. O. Fruend, C. J. Goebel, and Y. Nambu (University of Chicago Press, Chicago, 1970), pp. 196–208. The authors integrated the covariant spin equation (Bargmann, Michel, Telegdi equation) with respect to a Michel tetrad. This frame consists of four orthonormal vectors one of which is the 4-velocity and all of which satisfy the Lorentz equation. The Frenet frame in a homogeneous electromagnetic field is an example of such a tetrad.

[11]C. V. Vishveshwara, J. Math. Phys. **9**, 1319 (1968).

[12]See Ref. 11. See also B. Carter, Phys. Rev. **174**, 1559 (1968) and R. D. Greene, E. L. Schucking, and C. V. Vishveshwara, "The Rest Frame in Stationary Spacetimes with Axial Symmetry," J. Math. Phys. (to be published).

[13]D. M. Eardley, "Note on Spacetimes that admit Constant Electromagnetic Fields," preprint (1973).

[14]See for example L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, Mass., 1962), p. 57 ff.

[15]See for example J. Plebanski, Bull. Acad. Polon. Sci. Ser. Math. Astr. Phys. **9**, 587, 593 (1961).

Explicit $O(5)$ Wigner coefficients*

M. Hongoh and R. T. Sharp

Physics Department, McGill University, Montreal, Quebec, Canada

D. E. Tilley

Département de physique, Collège militaire royal de Saint-Jean, Saint-Jean, Québec, Canada

(Received 21 January 1974)

The $O(5)$ van der Waerden invariant is given; with its help the general Wigner coefficient in the canonical $SU(2)\times SU(2)$ chain is calculated. Some special Wigner coefficients and Gaunt-type formulas needed for the construction of the general Wigner coefficient are also presented.

1. INTRODUCTION

The group $O(5)$ has been useful in classifying states in several different areas of physics. Perhaps the best known is the seniority model which treats the pairing force between particles in the same nuclear shell.[1] Special representations serve to describe nuclear surfon states[2] and more recently have been used to treat states of atomic electrons interacting with vibrations of neighboring atoms in a solid (Jahn–Teller effect).[3] In its noncompact version $O(3,2)$ it is related to the Coulomb problem.[4] A practical problem in connection with any group is the calculation of its Wigner coefficients which couple states of three irreducible representations (IR's) to a scalar or, equivalently, its Clebsch–Gordan coefficients which couple states of two IR's to a composite state belonging to a third IR. These coefficients are needed for coupling states and tensors, and perhaps more importantly, in connection with the Wigner–Eckart theorem, for calculating matrix elements of physical quantities. Hecht and Pang[5] calculated $O(5)$ Clebsch–Gordan coefficients for special cases. Wong[6] derived those associated with the coupling $(p,q)\times(0,2)$, involving external multiplicity up to two. The most general Wigner coefficient involving no multiplicity was given by Ališauskas and Jucys.[7] Holman[8] calculates the general $O(5)$ coefficient by a factorization scheme which involves embedding in a higher group; a disadvantage is the presence of redundant labels, so that an orthogonalization or some such procedure is needed to retain a complete and independent set of couplings.

In Sec. 2 we construct the general van der Waerden invariant for $O(5)$.

The method of van der Waerden invariants provides a complete nonredundant solution of the external labeling problem which is symmetric in the three IR's and which has the advantage of making no reference to the internal basis states to be used. To calculate Wigner coefficients, one merely expands the van der Waerden invariant in products of states of the three IR's using whatever basis states are convenient for the problem at hand. In this paper we use states classified according to the canonical $SU(2)\times SU(2)$ subgroup of $O(5)$.

Section 3 is devoted to the calculation of certain Gaunt coefficients which arise when a product of two $O(5)$ states in the same variables is expanded in $O(5)$ states.

In Sec. 4 we evaluate certain classes of simple Wigner coefficients.

In Sec. 5 the general $O(5)$ Wigner coefficient is calculated explicitly. It takes the form of a sum of products of $9j$ symbols, Gaunt coefficients, and the simple Wigner coefficients of Sec. 4. It involves 41 internal summations.

2. $O(5)$ VAN DER WAERDEN INVARIANT

The external labeling problem for a compact group is that of specifying all couplings of two IR's to give a third or, equivalently, specifying all couplings of three IR's to give an invariant. Long ago van der Waerden[9] wrote down such a general invariant for $SU(2)$ as a product of powers of certain elementary scalars; similar solutions of the external labeling problem exist for $SU(3)$,[10] $SU(4)$,[11] and $SU(5)$.[12] A general discussion of the method is found in Ref. 12; it is an extension to the external labeling problem of the method of elementary multiplets (elementary permissible diagrams) which has been used to solve internal labeling problems.[13,14]

Examination of couplings of low-lying IR's suggests the following elementary scalars for $O(5)$:

$$A_1=(00,\ 10,\ 10),\quad A_2=(10,\ 00,\ 10),\quad A_3=(10,\ 10,\ 00),$$
$$B_1=(00,\ 01,\ 01),\quad B_2=(01,\ 00,\ 01),\quad B_3=(01,\ 01,\ 00),$$
$$C_1=(01,\ 10,\ 10),\quad C_2=(10,\ 01,\ 10),\quad C_3=(10,\ 10,\ 01),$$
$$D_1=(20,\ 01,\ 01),\quad D_2=(01,\ 20,\ 01),\quad D_4=(01,\ 01,\ 20). \tag{2.1}$$

Explicit expressions for the elementary scalars are given in Sec. 4. They are not independent, for C_iC_j is a linear combination of A_kD_k and $A_iA_jB_k$; D_iD_j is a linear combination of $A_k^2B_iB_j$ and $B_k^2C_k^2$; C_iD_i is a linear combination of $A_jB_kC_k$ and $A_kB_jC_j$ (throughout this paper ijk are 123 in any order). Accordingly C_iC_j, D_iD_j, and C_iD_i are incompatible pairs for the purpose of forming the general van der Waerden invariant. Six types of invariant may be distinguished, each characterized by a product of powers of eight elementary scalars

$$S_{cd}=PA_1^{a_1}A_2^{a_2}A_3^{a_3}B_1^{b_1}B_2^{b_2}B_3^{b_3}C_i^cD_j^d, \tag{2.2}$$

where $i\neq j$ are chosen from 123. Since the invariants differ only by a relabeling of the three IR's, only one will be discussed in detail, namely that with $i=1$, $j=2$. P in (2.2) is a projection operator which instructs us to retain only the part which is stretched in all IR labels. Thus

$$p_1 = a_2 + a_3, \quad p_2 = a_1 + a_3 + c + 2d, \quad p_3 = a_1 + a_2 + c,$$
$$q_1 = b_2 + b_3 + c + d, \quad q_2 = b_1 + b_2, \quad q_3 = b_1 + b_2 + d. \tag{2.3}$$

The eight exponents provide the six representation labels and two multiplicity labels. It is clear that we may use c, d as the multiplicity lables. All couplings are obtained by letting the exponents take all nonnegative integer values.

To show that (2.2) solves the external labeling problem, we invoke Speiser's theorem[16] which makes a correspondence between basis states of the first IR $(p_1 q_1)$ and couplings involving the first IR and the second IR $(p_2 q_2)$.

Basis states of the general IR $(p\,q)$ of $O(5)$ can be characterized by products of powers of variables which are the basis states of the two FIR's (fundamental irreducible representations); the basis states are of degree p in the variables $\alpha\beta\gamma\delta$ of (10) and degree q in the variables $\eta\theta\lambda\xi\zeta$ of (01); the notation for the states of the FIR's is that of Ref. 15. The existence of a symmetric scalar in the direct product $(01)\times(01)$ and of a (10) quartet in the product $(10)\times(01)$ occasions the proviso that the following pairs of variables are incompatible for the purpose of forming states of higher IR's: $\theta\xi$, $\gamma\theta$, $\gamma\lambda$, $\delta\lambda$, $\delta\xi$; the states thus eliminated belong to IR's lower than their degrees would indicate. Five types of internal state may be distinguished, each characterized by a product of powers of six variables; the variables are $\alpha\beta\eta\zeta$ together with one of the following five sets of two: $\gamma\delta$, $\gamma\xi$, $\delta\theta$, $\theta\lambda$, $\lambda\xi$. The indices of the powers of the six variables provide the two representation and four internal labels. This primitive solution of the state labeling problem utilizes no subgroup [apart from the $U(1)$ subgroups corresponding to the two weights]; it is convenient for our purpose because it corresponds to our symmetric solution of the external labelling problem.

In specifying the connection between variables and elementary scalars, we ignore A_1, B_1. By identifying weights we are led to the following correspondence:

$$A_2 \sim \alpha, \quad A_3 \sim \beta, \quad B_2 \sim \eta, \quad B_3 \sim \zeta, \quad C_1 \sim \lambda,$$
$$C_2 \sim \delta, \quad C_3 \sim \gamma, \quad D_2 \sim \xi, \quad D_3 \sim \theta.$$

(α and η are regarded as the heaviest states of their respective IR's). D_1 corresponds to $\gamma\delta$; there is no double counting since $C_2 C_3$ are incompatible. Then internal states correspond one-to-one to external couplings. Our justification of the solution provided by (2.2) is valid only when $p_2 q_2$ are sufficiently large. Its correctness in the general case can be proved by the method discussed in Ref. 11 or 12.

3. GAUNT COEFFICIENTS

In expanding the van der Waerden invariant (2.2) for the evaluation of $O(5)$ Wigner coefficients it is necessary to combine $O(5)$ states in, say, the 1-variables arising from different factors of the invariant into a sum of states in the same variables. The required formulas are analogous to the well-known formula of Gaunt by which a product of two spherical harmonics in the same variables is expanded as a sum of spherical harmonics. Because of the projection operator P in Eq. (2.2), we need keep only that part of the expansion which is stretched in the $O(5)$ Cartan labels p, q; terms belonging to lower IR's will be denoted by "unwanted."

First consider a state belonging to the simple IR of type (pO). According to (3.2) of Ref. 15 it can be written

$$\left| \begin{matrix} pO \\ STMN \end{matrix} \right\rangle = \left| \begin{matrix} S \\ M \end{matrix} \right\rangle \left| \begin{matrix} T \\ N \end{matrix} \right\rangle, \tag{3.1}$$

where $|{}^{S}_{M}\rangle$, $|{}^{T}_{N}\rangle$ are Wigner monomials, i.e., $SU(2)$ basis states, in the $\alpha\beta$ and $\gamma\delta$ variables, respectively; $S+T = \frac{1}{2}p$. The $SU(2)$ Gaunt formula is

$$\left| \begin{matrix} S_1 \\ M_1 \end{matrix} \right\rangle \left| \begin{matrix} S_2 \\ M_2 \end{matrix} \right\rangle = \left| \begin{matrix} S \\ M \end{matrix} \right\rangle \left\langle \begin{matrix} S_1 & S_2 \\ M_1 & M_2 \end{matrix} \middle| \begin{matrix} S \\ M \end{matrix} \right\rangle \{S_1;S_2|S\}, \tag{3.2}$$

where $S = S_1 + S_2$ and the $SU(2)$ Gaunt coefficient is

$$\{S_1;S_2|S\} = [(2S_1 + 2S_2)!/(2S_1)!(2S_2)!]^{1/2}. \tag{3.3}$$

Using (3.3), we obtain the Gaunt formula

$$\left| \begin{matrix} p_1 O \\ S_1 T_1 M_1 N_1 \end{matrix} \right\rangle \left| \begin{matrix} p_2 O \\ S_2 T_2 M_2 N_2 \end{matrix} \right\rangle$$
$$= \left| \begin{matrix} pO \\ STMN \end{matrix} \right\rangle \left\langle \begin{matrix} S_1 S_2 \\ M_1 M_2 \end{matrix} \middle| \begin{matrix} S \\ M \end{matrix} \right\rangle \left\langle \begin{matrix} T_1 T_2 \\ N_1 N_2 \end{matrix} \middle| \begin{matrix} T \\ N \end{matrix} \right\rangle$$
$$\times \left\{ \begin{matrix} p_1 O & p_2 O & pO \\ S_1 T_1 & S_2 T_2 & ST \end{matrix} \right\}, \tag{3.4}$$

where $p = p_1 + p_2$, $S = S_1 + S_2$, $T = T_1 + T_2$, and the $O(5)$ Gaunt coefficient is

$$\left\{ \begin{matrix} p_1 O & p_2 O & pO \\ S_1 T_1 & S_2 T_2 & ST \end{matrix} \right\} = \left(\frac{(2S_1 + 2S_2)!(2T_1 + 2T_2)!}{(2S_1)!(2S_2)!(2T_1)!(2T_2)!} \right)^{1/2}. \tag{3.5}$$

Before deriving the Gaunt coefficient for (Oq) type states we cast the (Oq) states in a new form:

$$\left| \begin{matrix} O\ q \\ SS\,;\,SS \end{matrix} \right\rangle = N'_S \eta^{2S} \lambda^{q-2S} + \text{unwanted} \tag{3.6a}$$

[$T = S$ for (Oq) type states]. Also

$$N'_S = \left(\frac{(2S+1)(2q+2)!}{(q-2S)!(q+2S+2)!(q+1)!\,2^{q-2S}} \right)^{1/2}. \tag{3.6b}$$

It is evident that the state (3.6a) is identical with (3.5) of Ref. 15 apart from normalization. That the normalization (3.6b) is correct is verified by taking the scalar product of (3.6a) with (3.5) of Ref. 15. The state $|{}^{O}_{SS};{}^{q}_{MN}\rangle$ is obtained from (3.6a) by replacing η^{2S} by

$$\eta^{2S} \to [(S+M)!(S-M)!(S+N)!(S-N)!]^{1/2}$$
$$\times \sum_m \frac{\eta^{M+m}\theta^{S-m}\xi^{S-M+N-m}\zeta^{m-N}}{(M+m)!(S-m)!(S-M+N-m)!(m-N)!}. \tag{3.6c}$$

The Gaunt formula for (Oq) states is

$$\left| \begin{matrix} Oq_1 \\ S_1 S_1;M_1 N_1 \end{matrix} \right\rangle \left| \begin{matrix} Oq_2 \\ S_2 S_2;M_2 N_2 \end{matrix} \right\rangle$$
$$= \sum_S \left| \begin{matrix} Oq \\ SS;MN \end{matrix} \right\rangle \left\langle \begin{matrix} S_1 S_2 \\ M_1 M_2 \end{matrix} \middle| \begin{matrix} S \\ M \end{matrix} \right\rangle \left\langle \begin{matrix} S_1 S_2 \\ N_1 N_2 \end{matrix} \middle| \begin{matrix} S \\ N \end{matrix} \right\rangle \tag{3.7}$$

$$\times\begin{Bmatrix} Oq_1 & Oq_2 & Oq \\ S_1S_1 & S_2S_2 & SS \end{Bmatrix} + \text{unwanted},$$

where $q=q_1+q_2$. To determine the Gaunt coefficient, put $M_1=N_1=S_1$. $M_2=N_2=-S_2$ in (3.7), substitute from (3.6a) for the states on the left, and take the scalar product with $|{}^{O}_{SS};{}^{q}_{MN}\rangle$ in the form (3.5) of Ref. 15 with the replacement (3.6c). The result is

$$\begin{Bmatrix} Oq_1 & Oq_2 & Oq \\ S_1S_1 & S_2S_2 & SS \end{Bmatrix} = (-1)^{S_1+S_2-S}\left(\frac{(2S_1+1)(2S_2+1)(2q_1+2)!}{(2q_1+2q_2+2)!(q_1-2S_1)!}\right)^{1/2}$$

$$\times\left(\frac{(2q_2+2)!(q_1+q_2-2S)!(q_1+q_2+1)!(q_1+q_2+2S+2)!}{(q_2-2S_2)!(q_1+1)!(q_2+1)!(2S+1)(q_1+2S_1+2)!(q_2+2S_2+2)!}\right)^{1/2} \tag{3.8}$$

We are now in a position to derive the general Gaunt formula

$$\left|\begin{matrix} p_1q_1 \\ S_1T_1;M_1N_1 \end{matrix}\right\rangle \left|\begin{matrix} p_2q_2 \\ S_2T_2;M_2N_2 \end{matrix}\right\rangle$$

$$=\sum_{ST}\left|\begin{matrix} pq \\ ST;MN \end{matrix}\right\rangle \left\langle\begin{matrix} S_1S_2 \\ M_1M_2 \end{matrix}\middle|\begin{matrix} S \\ M \end{matrix}\right\rangle \left\langle\begin{matrix} T_1T_2 \\ N_1N_2 \end{matrix}\middle|\begin{matrix} T \\ N \end{matrix}\right\rangle$$

$$\times\begin{Bmatrix} p_1q_1 & p_2q_2 & pq \\ S_1T_1 & S_2T_2 & ST \end{Bmatrix} + \text{unwanted}, \tag{3.9}$$

where $p=p_1+p_2$, $q=q_1+q_2$. The general $O(5)$ state can be cast in the form

$$\left|\begin{matrix} pq \\ ST;MN \end{matrix}\right\rangle = \left|\begin{matrix} pq & \underline{S}' & \underline{S}'' \\ ST & \underline{M} & \underline{N} \end{matrix}\right\rangle A'^{\,pq}_{ST} + \text{unwanted}, \tag{3.10a}$$

where [see (4.7) of Ref. 15]

$$A'^{\,pq}_{ST}=[A_{ST}(\underline{S}'\,\underline{S}'')]^{-1}, \tag{3.10b}$$

$$\left|\begin{matrix} pq & \underline{S}'\underline{S}'' \\ ST & \underline{M}\underline{N} \end{matrix}\right\rangle = \sum_{M'N'}\left|\begin{matrix} pO \\ \underline{S}'\,\underline{T}'\,M'N' \end{matrix}\right\rangle \left|\begin{matrix} Oq \\ \underline{S}''\,\underline{S}''\,M''N'' \end{matrix}\right\rangle \left\langle\begin{matrix} \underline{S}' & \underline{S}'' \\ \underline{M}' & \underline{M}'' \end{matrix}\middle|\begin{matrix} S \\ M \end{matrix}\right\rangle \left\langle\begin{matrix} \underline{T}' & \underline{S}'' \\ \underline{N}' & \underline{N}'' \end{matrix}\middle|\begin{matrix} T \\ N \end{matrix}\right\rangle, \tag{3.10c}$$

and $\underline{S}'=\frac{1}{2}(\frac{1}{2}p+S-T)$, $\underline{S}''=\frac{1}{2}(S+T-\frac{1}{2}p)$, $\underline{T}'=\frac{1}{2}p-\underline{S}'$,

$$M''=M-M',\quad N''=N-N' \tag{3.10d}$$

The justification of (3.10) is similar to that of (3.6a).

To evaluate the Gaunt coefficient in (3.9), take the scalar product of both sides with $|{}^{p}_{ST};{}^{q}_{MN}\rangle$. In evaluating the scalar product on the left-hand side, use (3.10) for the states on the right side of the scalar product, (4.1) of Ref. 15 for the state on the left. The result is

$$\begin{Bmatrix} p_1q_1 & p_2q_2 & pq \\ S_1T_1 & S_2T_2 & ST \end{Bmatrix} = A'^{\,p_1q_1}_{S_1T_1}A'^{\,p_2q_2}_{S_2T_2}\begin{Bmatrix} p_1O & p_2O & pO \\ \underline{S}_1'\underline{T}_1' & \underline{S}_2'\underline{T}_2' & S'T' \end{Bmatrix}$$

$$\times[(2S'+1)(2S_1+1)(2S_2+1)(2T'+1)(2T_1+1)(2T_2+1)]^{1/2}$$

$$\times\sum_{S''}(2S''+1)\begin{Bmatrix} Oq_1 & Oq_2 & Oq \\ \underline{S}_1''S_1'' & \underline{S}_2''\underline{S}_2'' & S''S'' \end{Bmatrix}$$

$$\times A_{ST}(S',S'')\begin{Bmatrix} \underline{S}_1' & \underline{S}_2' & S' \\ \underline{S}_1'' & \underline{S}_2'' & S'' \\ S_1 & S_2 & S \end{Bmatrix}\begin{Bmatrix} \underline{T}_1' & \underline{T}_2' & T' \\ \underline{T}_1'' & \underline{T}_2'' & S'' \\ T_1 & T_2 & T \end{Bmatrix}. \tag{3.11}$$

The $9j$ symbols are stretched in their first and second columns and their first rows, and contain one sum each.[17] Thus the Gaunt coefficient (3.11) contains three sums.

Special cases of (3.11) needed later follow:

$$\begin{Bmatrix} p_1q_1 & p_2O & pq_1 \\ S_1T_1 & S_2T_2 & ST \end{Bmatrix}$$

$$=(-1)^p[(2S'+1)(2S_1+1)(2T'+1)(2T_1+1)]^{1/2}A'^{\,p_1q_1}_{S_1T_1}$$

$$\times\begin{Bmatrix} S_1 & S_2 & S \\ S' & \underline{S}_1'' & \underline{S}_1' \end{Bmatrix}\begin{Bmatrix} T_1 & T_2 & T \\ T' & \underline{T}_1'' & \underline{T}_1' \end{Bmatrix}\begin{Bmatrix} p_1O & p_2O & pO \\ \underline{S}_1'\underline{T}_1' & S_2T_2 & S'T' \end{Bmatrix}, \tag{3.12}$$

where $p=p_1+p_2$, $S'=\underline{S}_1'+S_2$. (3.12) involves no sum.

$$\begin{Bmatrix} p_1q_1 & Oq_2 & p_1q \\ S_1T_1 & S_2T_2 & ST \end{Bmatrix}$$

$$=(-1)^{S+S_1+S_2+T+T_1+T_2}\ [(2S_1+1)(2T_1+1)]^{1/2}A'^{\,p_1q_1}_{S_1T_1}$$

$$\times\sum_{S''}(2S''+1)A_{ST}(\underline{S}_1'S'')\begin{Bmatrix} S_1 & S_2 & S \\ S'' & \underline{S}_1' & \underline{S}_1'' \end{Bmatrix}$$

$$\times\begin{Bmatrix} T_1 & T_2 & T \\ T'' & \underline{T}_1' & \underline{T}_1'' \end{Bmatrix}\begin{Bmatrix} O & q_1 & O & q_2 \\ \underline{S}_1'' & \underline{S}_1'' & S_2 & S_2 \end{Bmatrix}\begin{Bmatrix} O & q \\ S'' & S'' \end{Bmatrix}, \tag{3.13}$$

where $q=q_1+q_2$. (3.13) involves one sum:

$$\begin{Bmatrix} pO & Oq \\ S'T' & S''T'' \end{Bmatrix}\begin{Bmatrix} p & q \\ S & T \end{Bmatrix} = A_{ST}(S'S''). \tag{3.14}$$

4. SOME SPECIAL WIGNER COEFFICIENTS

We fix the phase and normalization of the van der Waerden invariant (2.2) by writing explicit expressions for the elementary scalars:

$$
\begin{aligned}
A_i &= \alpha_j\beta_k - \beta_j\alpha_k + \gamma_j\delta_k - \delta_j\gamma_k,\\
B_i &= \eta_j\xi_k + \zeta_j\eta_k - \theta_j\xi_k - \xi_j\theta_k + \lambda_j\lambda_k,\\
C_i &= (\alpha_j\beta_k - \beta_j\alpha_k + \delta_j\gamma_k - \gamma_j\delta_k)\lambda_i\\
&\quad + \sqrt{2}(\delta_j\beta_k - \beta_j\delta_k)\eta_i\\
&\quad + \sqrt{2}(-\gamma_j\beta_k + \beta_j\gamma_k)\theta_i + \sqrt{2}(-\delta_j\alpha_k + \alpha_j\delta_k)\xi_i\\
&\quad + \sqrt{2}(\gamma_j\alpha_k - \alpha_j\gamma_k)\zeta_i, \qquad (4.1)\\
D_i &= (\eta_j\theta_k - \theta_j\eta_k)\beta_i^2\\
&\quad - (\xi_j\theta_k - \theta_j\xi_k + \eta_j\zeta_k - \zeta_j\eta_k)\alpha_i\beta_i\\
&\quad + (\xi_j\zeta_k - \zeta_j\xi_k)\alpha_i^2 + (\eta_j\xi_k - \xi_j\eta_k)\delta_i^2\\
&\quad + (\theta_j\zeta_k - \zeta_j\theta_k)\gamma_i^2\\
&\quad - (\theta_j\xi_k - \xi_j\theta_k + \eta_j\zeta_k - \zeta_j\eta_k)\gamma_i\delta_i\\
&\quad + \sqrt{2}(\eta_j\lambda_k - \lambda_j\eta_k)\beta_i\delta_i\\
&\quad - \sqrt{2}(\theta_j\lambda_k - \lambda_j\theta_k)\beta_i\gamma_i - \sqrt{2}(\xi_j\lambda_k - \lambda_j\xi_k)\alpha_i\delta_i\\
&\quad + \sqrt{2}(\zeta_j\lambda_k - \lambda_j\zeta_k)\alpha_i\gamma_i.
\end{aligned}
$$

The van der Waerden invariant may be expanded as a linear combination of products of states in the 1-, 2-, and 3-variables:

$$
\begin{aligned}
S_{cd} = &\sum_{\substack{S_1S_2S_3\\T_1T_2T_3}} \sum_{\substack{M_1M_2\\N_1N_2}} \left|\begin{matrix} p_1q_1\\ S_1T_1;M_1N_1\end{matrix}\right\rangle_1\\
&\times \left|\begin{matrix} p_2q_2\\ S_2T_2;M_2N_2\end{matrix}\right\rangle_2 \left|\begin{matrix} p_3q_3\\ S_3T_3;M_3N_3\end{matrix}\right\rangle_3 \begin{pmatrix} S_1 & S_2 & S_3\\ M_1 & M_2 & M_3\end{pmatrix}\\
&\times \begin{pmatrix} T_1 & T_2 & T_3\\ N_1 & N_2 & N_3\end{pmatrix} \begin{pmatrix} p_1q_1 & p_2q_2 & p_3q_3\\ S_1T_1 & S_2T_2 & S_3T_3\end{pmatrix}_{cd}. \qquad (4.2)
\end{aligned}
$$

The Wigner coefficient is just the coefficient in this expansion; by the Wigner–Eckart theorem it factors into a product of two $SU(2)$ Wigner coefficients and the reduced $O(5)$ Wigner coefficient

$$\begin{pmatrix} p_1q_1 & p_2q_2 & p_3q_3\\ S_1T_1 & S_2T_2 & S_3T_3\end{pmatrix}_{cd}.$$

In this section we calculate the special reduced Wigner coefficients arising from the special van der Waerden invariants C_1^c, D_2^d, $A_1^{a_1}A_2^{a_2}A_3^{a_3}$, and $B_1^{b_1}B_2^{b_2}B_3^{b_3}$; they are needed later as components of the general $O(5)$ Wigner coefficient. These special invariants involve no external multiplicity and are readily normalized. The Wigner coefficients we derive are the unnormalized ones; in each case the normalization constant is given separately. The invariant PC_1^c can be written

$$
\begin{aligned}
PC_1^c = &\frac{1}{N_c}\left(\frac{6}{(c+1)(c+2)(c+3)}\right)^{1/2}\\
&\times \sum_{S_1M_1N_1} \left|\begin{matrix} Oc\\ S_1S_1;M_1N_1\end{matrix}\right\rangle_1 \left|\begin{matrix} Oc\\ S_1S_1;-M_1,-N_1\end{matrix}\right\rangle_{23} (-1)^{M_1-N_1}
\end{aligned}
\qquad (4.3)
$$

where

$$\left|\begin{matrix} Oc\\ S_1S_1;-M_1,-N_1\end{matrix}\right\rangle_{23}$$

is a composite state formed from two IR's of type (cO) in the 2- and 3-variables. N_c is the normalization factor for the invariant. Such states can be constructed by methods similar to those of Ref. 15; the result, in an obvious notation, is

$$
\begin{aligned}
\left|\begin{matrix} Oc\\ SS;MN\end{matrix}\right\rangle_{23} &= (c+1)!\left(\frac{(c-2S)!(c+2S+2)!}{(2S+1)(2c+2)!}\right)^{1/2}\\
&\times \sum_{S_2S_3} \left|\begin{pmatrix} cO & cO\\ S_2T_2 & S_3T_3\end{pmatrix}\begin{matrix} SS\\ MN\end{matrix}\right\rangle_{23}\\
&\times (-1)^{2T_3}[(S_2+S_3-S)!(S+S_2+S_3+1)!\\
&\times (T_2+T_3-S)!(S+T_2+T_3+1)!]^{-1/2}. \qquad (4.4)
\end{aligned}
$$

Substitution of (4.4) into (4.3) gives the special reduced Wigner coefficient:

$$
\begin{aligned}
&\begin{pmatrix} Oc & cO & cO\\ S_1S_1 & S_2S_2 & S_3T_3\end{pmatrix}\\
&= (-1)^{2T_2}\,c!\left(\frac{2^c(2S_1+1)\{(c+1)!\}^3(c-2S_1)!}{(2c+2)!(S_2+S_3-S_1)!}\right)^{1/2}\\
&\times\left(\frac{(c+2S_1+2)!}{(S_1+S_2+S_3+1)!(T_2+T_3-T_1)!(T_1+T_2+T_3+1)!}\right)^{1/2},
\end{aligned}
\qquad (4.5)
$$

and the normalization factor in Eq. (4.3)

$$N_c = \sqrt{3}\,[c!\,(c+1)!\,(c+2)!\,(2c+3)2^{c-1}]^{-1/2}. \qquad (4.6)$$

The invariant PD_2^d can be written.

$$
\begin{aligned}
PD_2^d = &\frac{1}{N_d}\left(\frac{3}{(d+1)(2d+1)(2d+3)}\right)^{1/2}\\
&\times \sum_{S_2M_2N_2} \left|\begin{matrix} 2d & O\\ S_2T_2;M_2N_2\end{matrix}\right\rangle_2 \left|\begin{matrix} 2d & O\\ S_2T_2;-M_2,-N_2\end{matrix}\right\rangle_{31} (-1)^{M_2+N_2},
\end{aligned}
\qquad (4.7)
$$

where $|{}^{2d}_{S_2T_2;-M_2,-N_2}{}^{O}\rangle_{31}$ is a composite state formed from two IR's of type (Od) in the 3- and 1-variables. Again such states can be constructed by the methods of Ref. 15 with the result

$$
\begin{aligned}
\left|\begin{matrix} 2d & O\\ ST & MN\end{matrix}\right\rangle_{31} &= \left(\frac{4(2d+1)(2d+1)!}{d!(d+1)!(2S+1)(2T+1)}\right)^{1/2}\\
&\times \sum_{S_3S_1} \left|\begin{pmatrix} Od & Od\\ S_3S_3 & S_1S_1\end{pmatrix}\begin{matrix} ST\\ MN\end{matrix}\right\rangle_{31}\\
&\times (-1)^{d-2S_3}\left(\frac{(2S_3+1)(2S_1+1)(S+S_3-S_1)!}{(S_3+S_1-S)!(S_3+S_1-T)!}\right.\\
&\times \left.\frac{(S+S_1-S_3)!(S+S_3+S_1+1)!}{(d-2S_3)!(d+2S_3+2)!(d-2S_1)!}\right)^{1/2}\\
&\times \left(\frac{(T+S_1-S_3)!(T+S_3-S_1)!(T+S_3+S_1+1)!}{(d+2S_1+2)!}\right)^{1/2}.
\end{aligned}
\qquad (4.8)
$$

Substitution of (4.8) into (4.7) gives the special reduced Wigner coefficient.

$$\begin{pmatrix} Od \\ S_1T_1 \end{pmatrix}; \begin{matrix} 2dO & Od \\ S_2T_2 & ; S_3T_3 \end{matrix}\Big) = (-1)^{2S_1}\, 2(2d+1)!$$

$$\times\left(\frac{(2S_1+1)(2S_3+1)(S_2+S_3-S_1)!}{(S_3+S_1-S_2)!(T_3+T_1-T_2)!}\right)^{1/2}$$

$$\times\left(\frac{(S_2+S_1-S_3)!(S_1+S_2+S_3+1)!}{(d-2S_3)!(d+2S_3+2)!}\right.$$

$$\left.\times\frac{(T_2+T_3-T_1)!(T_2+T_1-T_3)!(T_1+T_2+T_3+1)!}{(d-2S_1)!(d+2S_1+2)!}\right)^{1/2} \tag{4.9}$$

and the normalization factor in (4.7)

$$N_d = \sqrt{6}\,[d!\,(d+1)!\,(2d+3)!\,]^{-1/2}.$$

Next consider the invariant $A_1^{a_1}A_2^{a_2}A_3^{a_3}$. The norm of $A_1^{a_1}A_2^{a_2}A_3^{a_3}$ is $(a_1+a_2+1)(a_1+a_2+2)(a_1+a_2+3)/6$ times the norm of the term containing $\alpha_3^{a_1+a_2}$. That term is $\alpha_3^{a_1+a_2}|a_3\rangle$, where $|a_3\rangle = \beta_2^{a_1}\beta_1^{a_2}A_3^{a_3}$. The norm of $|a_3\rangle$ is found from the equation

$$\langle a_3+1|\,A_3\,|a_3\rangle = \langle a_3|\,A_3^{+}\,|a_3+1\rangle; \tag{4.10}$$

A_3^{+} is just A_3 with variables replaced by differentiations. The resulting normalization factor for $A_1^{a_1}A_2^{a_2}A_3^{a_3}$ is

$$N_a = \sqrt{6}\,[a_1!\,a_2!\,a_3!\,(a_1+a_2+a_3+3)!\,]^{-1/2}. \tag{4.11}$$

To expand $A_1^{a_1}A_2^{a_2}A_3^{a_3}$, break each A into its S-spin and T-spin parts and make a binomial expansion; the S-factor and T-factor each form an $SU(2)$ van der Waerden invariant. Expanding these and using (3.2) of Ref. 15, we find the special reduced coefficient

$$\begin{pmatrix} p_1O \\ S_1T_1 \end{pmatrix}; \begin{matrix} p_2O & p_3O \\ S_2T_2 & ; S_3T_3 \end{matrix}\Big)$$

$$=\left(\frac{(S_1+S_2+S_3+1)!}{(S_2+S_3-S_1)!(S_3+S_1-S_2)!(S_1+S_2-S_3)!}\right.$$

$$\left.\times\frac{(T_1+T_2+T_3+1)!}{(T_2+T_3-T_1)!(T_3+T_1-T_2)!(T_1+T_2-T_3)!}\right)^{1/2}$$

$$\times[\tfrac{1}{2}(p_2+p_3-p_1)]!\,[\tfrac{1}{2}(p_3+p_1-p_2)]!\,[\tfrac{1}{2}(p_1+p_2-p_3)]!, \tag{4.12}$$

where $p_1 = a_2+a_3$, etc.

Finally consider the invariant $P\,B_1^{b_1}B_2^{b_2}B_3^{b_3}$. Its normalization constant is evaluated by methods similar to those used in normalizing $A_1^{a_1}A_2^{a_2}A_3^{a_3}$. The result is

$$N_b=\left(\frac{6(2b_1+2b_2+1)!\,!(2b_1+2b_3+1)!\,!}{b_1!\,b_2!\,b_3!(2b_1+1)!\,!(2b_2+1)!\,!(2b_3+1)!\,!}\right.$$

$$\left.\times\frac{(2b_2+2b_3+1)!\,!}{(b_1+b_2+b_3+2)!(2b_1+2b_2+2b_3+3)!\,!}\right)^{1/2} \tag{4.13}$$

$PB_1^{b_1}B_2^{b_2}B_3^{b_3}$ is expanded by first expanding the individual factors according to

$$PB_3^{b_3} = b_3!\sum_{SMN}\left|\begin{matrix} O & b_3 \\ SS\,; & MN \end{matrix}\right\rangle_1\left|\begin{matrix} O & b_3 \\ SS\,; & -M-N \end{matrix}\right\rangle_2 (-1)^{M-N}, \tag{4.14}$$

and then combining the states with the help of the Gaunt formula (3.7). The special reduced Wigner coefficient turns out to be

$$\begin{pmatrix} O\,q_1 \\ S_1S_1 \end{pmatrix}; \begin{matrix} O\,q_2 & O\,q_3 \\ S_2S_2 & ; S_3S_3 \end{matrix}\Big)$$

$$=(-1)^{S_1+S_2+S_3}\,(q_2+q_3-q_1+1)!(q_3+q_1-q_2+1)!(q_1+q_2-q_3+1)!$$

$$\times\left(\frac{8(2S_1+1)(2S_2+1)(2S_3+1)q_1!\,q_2!\,q_3!}{(2q_1+1)!(2q_2+1)!}\right.$$

$$\left.\times\frac{(q_1-2S_1)!(q_2-2S_2)!(q_3-2S_3)!}{(2q_3+1)!}\right)^{1/2}$$

$$\times[(q_1+2S_1+2)!(q_2+2S_2+2)!(q_3+2S_3+2)!\,]^{1/2}$$

$$\times\sum_{S_1'S_2'S_3'}(-1)^{2(S_1'+S_2'+S_3')}$$

$$\times\frac{(2S_1'+1)(2S_2'+1)(2S_3'+1)}{[\frac{1}{2}(q_2+q_3-q_1)-2S_1']!\,[\frac{1}{2}(q_3+q_1-q_2)-2S_2']!}$$

$$\times\frac{\begin{Bmatrix} S_1 & S_2 & S_3 \\ S_1' & S_2' & S_3' \end{Bmatrix}^2}{[\frac{1}{2}(q_1+q_2-q_3)-2S_3']!}$$

$$\times\{[\tfrac{1}{2}(q_2+q_3-q_1)+2S_1'+2\}!\,[\tfrac{1}{2}(q_3+q_1-q_2+2S_2'+2]!$$

$$\times[\tfrac{1}{2}(q_1+q_2-q_3)+2S_3'+2]!\}^{-1}. \tag{4.15}$$

Formula (4.15) contains four sums. The special Wigner coefficients found earlier in this section contain no sums.

5. GENERAL WIGNER COEFFICIENT

To expand the general van der Waerden invariant (2.2), one first expands the separate factors $A_1^{a_1}A_2^{a_2}A_3^{a_3}$, $B_1^{b_1}B_2^{b_2}B_3^{b_3}$, C_1^c, D_2^d using the formulas of Sec. 4, then combines the different factors involving the 1-, 2-, 3-variables respectively with the help of the formulas of Sec. 3. The order in which the factors are combined affects the form but not the numerical value of the final result. We first combined $A_1^{a_1}A_2^{a_2}A_3^{a_3}$ with C_1^c and $B_1^{b_1}B_2^{b_2}B_3^{b_3}$ with D_2^d, then combined the results to arrive at a final formula:

$$\begin{pmatrix} p_1q_1 \\ S_1T_1 \end{pmatrix}; \begin{matrix} p_2q_2 & p_3q_3 \\ S_2T_2 & ; S_3T_3 \end{matrix}\Big)_{cd}$$

$$=\sum_{S\tilde{T}}\begin{Bmatrix} p_1O \\ S_1'T_1' \end{Bmatrix}; \begin{matrix} Oc \\ S_1''T_1'' \end{matrix}\Big|\begin{matrix} p_1 & c \\ \tilde{S}_1 & \tilde{T}_1 \end{matrix}\Big\}\begin{Bmatrix} p_2-c-2d\,O & cO \\ S_2'T_2' & ; S_2''T_2'' \end{Bmatrix}\Big|\begin{matrix} p_2-2d & O \\ \tilde{S}_2 & \tilde{T}_2 \end{matrix}\Big\}$$

$$\times\begin{Bmatrix} p_3-c & O \\ S_3'T_3' & ; \end{Bmatrix}\begin{matrix} c\,O \\ S_3''T_3'' \end{matrix}\Big|\begin{matrix} p_3 & O \\ \tilde{S}_3 & \tilde{T}_3 \end{matrix}\Big\}$$

$$\times\begin{Bmatrix} O\,q_1-c-d \\ S_1^*T_1^* \end{Bmatrix}; \begin{matrix} O\,d \\ S_1^{**}T_1^{**} \end{matrix}\Big|\begin{matrix} O\,q_1-c \\ \bar{S}_1\bar{T}_1 \end{matrix}\Big\}\begin{Bmatrix} O\,q_2 \\ S_2^*T_2^* \end{Bmatrix}; \begin{matrix} 2dO \\ S_2^{**}T_2^{**} \end{matrix}\Big|\begin{matrix} 2d\,q_2 \\ \bar{S}_2\bar{T}_2 \end{matrix}\Big\}$$

$$\times\begin{Bmatrix} O\,q_3-d \\ S_3^*T_3^* \end{Bmatrix}; \begin{matrix} O\,d \\ S_3^{**}T_3^{**} \end{matrix}\Big|\begin{matrix} O\,q_3 \\ \bar{S}_3\bar{T}_3 \end{matrix}\Big\}\begin{Bmatrix} p_1c \\ \tilde{S}_1\tilde{T}_1 \end{Bmatrix}; \begin{matrix} O\,q_1-c \\ \bar{S}_1\bar{T}_1 \end{matrix}\Big|\begin{matrix} p_1\,q_1 \\ S_1\,T_1 \end{matrix}\Big\}$$

$$\times\begin{Bmatrix} p_2-2d\,O \\ \tilde{S}_2\tilde{T}_2 \end{Bmatrix}; \begin{matrix} 2d\,q_2 \\ \bar{S}_2\bar{T}_2 \end{matrix}\Big|\begin{matrix} p_2\,q_2 \\ S_2\,T_2 \end{matrix}\Big\}\begin{Bmatrix} p_3O \\ \tilde{S}_3\tilde{T}_3 \end{Bmatrix}; \begin{matrix} O\,q_3 \\ \bar{S}_3\bar{T}_3 \end{matrix}\Big|\begin{matrix} p_3\,q_3 \\ S_3\,T_3 \end{matrix}\Big\}$$

$$\times\begin{pmatrix} p_1\,0 & p_2-c-2d\,0 & p_3-c\,0 \\ S_1'T_1' & S_2'T_2' & S_3'T_3' \end{pmatrix}\begin{pmatrix} 0\,c & c\,0 & c\,0 \\ S_1''T_1'' & S_2''T_2'' & S_3''T_3'' \end{pmatrix}$$

$$\times\begin{pmatrix} 0\,d & 2d\,0 & 0\,d \\ S_1^{**}T_1^{**} & S_2^{**}T_2^{**} & S_3^{**}T_3^{**} \end{pmatrix}[(2\tilde{S}_1+1)(2\tilde{S}_2+1)]^{1/2}$$

$$\times[(2\tilde{S}_3+1)(2\bar{S}_1+1)(2\bar{S}_2+1)(2\bar{S}_3+1)(2S_1+1)$$
$$\times(2S_2+1)(2S_3+1)(2\tilde{T}_1+1)(2\tilde{T}_2+1)(2\tilde{T}_3+1)]^{1/2}$$
$$\times[(2\bar{T}_1+1)(2\bar{T}_2+1)(2\bar{T}_3+1)(2T_1+1)(2T_2+1)(2T_3+1)]^{1/2}$$

$$\times\begin{Bmatrix} S_1'' & S_2'' & S_3'' \\ S_1' & S_2' & S_3' \\ \tilde{S}_1 & \tilde{S}_2 & \tilde{S}_3 \end{Bmatrix}\begin{Bmatrix} S_1^{**} & S_2^{**} & S_3^{**} \\ S_1^{*} & S_2^{*} & S_3^{*} \\ \bar{S}_1 & \bar{S}_2 & \bar{S}_3 \end{Bmatrix}$$

$$\times\begin{Bmatrix} \bar{S}_1 & \bar{S}_2 & \bar{S}_3 \\ \tilde{S}_1 & \tilde{S}_2 & \tilde{S}_3 \\ S_1 & S_2 & S_3 \end{Bmatrix}$$

$$\times\begin{Bmatrix} T_1'' & T_2'' & T_3'' \\ T_1' & T_2' & T_3' \\ \tilde{T}_1 & \tilde{T}_2 & \tilde{T}_3 \end{Bmatrix}\begin{Bmatrix} T_1^{**} & T_2^{**} & T_3^{**} \\ T_1^{*} & T_2^{*} & T_3^{*} \\ \bar{T}_1 & \bar{T}_2 & \bar{T}_3 \end{Bmatrix}$$

$$\times\begin{Bmatrix} \bar{T}_1 & \bar{T}_2 & \bar{T}_3 \\ \tilde{T}_1 & \tilde{T}_2 & \tilde{T}_3 \\ T_1 & T_2 & T_3 \end{Bmatrix} \tag{5.1}$$

The sum in Eq. (5.1) is over the variables $S_1', S_1'', \tilde{S}_1, S_2', \bar{S}_3, S_2'', S_3', S_3'', S_1^*, S_1^{**}, \bar{S}_1, S_2^*, S_2^{**}, \bar{S}_2, S_3^*, S_3^{**}, \tilde{T}_1, \bar{T}_2$. The other S's and T's are not independent:

$$\tilde{S}_2=S_2'+S_2'',\quad \tilde{S}_3=S_3'+S_3'',\quad T_1'=\tfrac{1}{2}(a_2+a_3)-S_1',$$
$$T_1''=S_1'',\quad T_2'=\tfrac{1}{2}(a_3+a_1)-S_2',\quad T_2''=\tfrac{1}{2}c-S_2'',$$
$$\tilde{T}_2=T_2'+T_2'',\ T_3'=\tfrac{1}{2}(a_1+a_2)-S_3',\quad T_3''=\tfrac{1}{2}c-S_3'',\quad \tilde{T}_3$$
$$=T_3'+T_3'',$$
$$T_1^*=S_1^*,\quad T_1^{**}=S_1^{**},\quad \bar{T}_1=\bar{S}_1,\quad T_2^*=S_2^*,$$
$$T_2^{**}=d-S_2^{**},\quad T_3^*=S_3^*,\quad T_3^{**}=S_3^{**},\quad \bar{T}_3=\bar{S}_3.$$

There are 18 explicit summations in Eq. (5.1) and 23 more in the various $9j$ symbols, Gaunt coefficients, and special reduced Wigner coefficients.

The metric matrix for the unnormalized van der Waerden invariant is discussed in an appendix.

The method of van der Waerden invariants is a powerful technique for obtaining expressions for the Wigner coefficients or Clebsch–Gordan coefficients of low order compact groups commonly utilized in physics. We plan to use the van der Waerden invariant of this paper to evaluate certain classes of Wigner coefficients for the noncanonical chains $O(5)\supset SU(2)\times U(1)$ and $O(5)\supset SU(2)$.

APPENDIX

Two van der Waerden invariants with the same p's and q's but different $c_i d_j$ are in general not orthogonal. We indicate here how the metric matrix (overlaps) may be calculated. Isolating the S_3T_3 multiplet which contains the heaviest state of (p_3q_3), we may write

$$S_{cd}=\sum_{\substack{S_1T_1\\S_2T_2}}\sum_{\substack{M_1N_1\\M_2N_2}}\left|\begin{matrix} p_1\,q_1 \\ S_1T_1;M_1N_1\end{matrix}\right\rangle$$

$$\times\left|\begin{matrix} p_2\,q_2 \\ S_2T_2;M_2N_2\end{matrix}\right\rangle\ \left|\begin{matrix} p_3\,q_3 \\ \frac{1}{2}(p_3+q_3)\,\frac{1}{2}q_3;M_3N_3\end{matrix}\right\rangle$$

$$\times\begin{pmatrix} S_1S_2 & \frac{1}{2}(p_3+q_3) \\ M_1M_2 & M_3\end{pmatrix}\begin{pmatrix} T_1T_2\,\frac{1}{2}q_3 \\ N_1N_2\ N_3\end{pmatrix}$$

$$\times\begin{pmatrix} p_1\,q_1 & p_2\,q_2 & p_3\,q_3 \\ S_1T_1 & S_2T_2 & \frac{1}{2}(p_3+q_3)\,\frac{1}{2}q_3\end{pmatrix}$$

+ contribution of other S_3T_3 multiplets. Since each of the $D_3=(p_3+1)(q_3+1)(p_3+q_3+2)(p_3+2q_3+3)/6$ states of (p_3q_3) contribute equally to the metric matrix element $\langle S_{c'd'}|S_{cd}\rangle$, we may write

$$\langle S_{c'd'}|S_{cd}\rangle=\frac{(p_3+1)(p_3+q_3+2)(p_3+2q_3+3)}{6(p_3+q_3+1)}$$

$$\times\sum_{S_1T_1S_2T_2}\begin{pmatrix} p_1q_1 & p_2q_2 & p_3\,q_3 \\ S_1T_1 & S_2T_2 & \frac{1}{2}(p_3+q_3)\,\frac{1}{2}q_3\end{pmatrix}_{c'd'}$$

$$\times\begin{pmatrix} p_1q_1 & p_2q_2 & p_3\,q_3 \\ S_1T_1 & S_2T_2 & \frac{1}{2}(p_3+q_3)\,\frac{1}{2}q_3\end{pmatrix}_{cd}. \tag{A1}$$

The reduced Wigner coefficients appearing in (A1) contain the heaviest ST multiplet of (p_3q_3) and hence are considerably simpler than the most general ones.

If orthogonalized couplings are desired, one can, of course, use the Schmidt scheme. Alternatives are to diagonalize the metric operator (A1) or to use eigenstates of the mixed Casimir operators.[18]

*Supported in part by the National Research Council of Canada.

[1]B.H. Flowers, Proc. Roy. Soc. (London) A **212**, 248 (1952); A.R. Edmonds and B.H. Flowers, *ibid.* A **214**, 515 (1952); K.T. Hecht, Nucl. Phys. **63**, 177 (1965); J.C. Parikh, *ibid.* **63**, 214 (1965).
[2]K.T. Hecht, Nucl. Phys. **63**, 177 (1965); S.A. Williams and D.L. Pursey, J. Math. Phys. **9**, 1230 (1968).
[3]H.A. Jahn and E. Teller, Proc. Roy. Soc. (London) **161**, 220 (1937); B.R. Judd, Can. J. Phys. (to appear).
[4]M.J. Englefield, *Group Theory and the Coulomb Problem* (Interscience, New York, 1972).
[5]K.T. Hecht and S.C. Pang, J. Math. Phys. **10**, 1571 (1969).

[6]M. K. F. Wong, J. Math. Phys. 12, 1530 (1971); Nucl. Phys. A 186, 177 (1972).
[7]S. J. Ališauskas and A. P. Jucys, J. Math. Phys. 10, 2227 (1969); 12, 594 (1971).
[8]W. J. Holman III, J. Math. Phys. 14, 440 (1973).
[9]B. L. van der Waerden, *Die gruppentheoretische Methode in der Quantenmechanik* (Springer, Berlin, 1932).
[10]Hou Pei-Yu, Scientia Sinica, 367 (1965); G. Ponzano, Nuovo Cimento A 41, 142 (1966); C.-K. Chew and R. T. Sharp, Nucl. Phys. B 2, 697 (1967).
[11]R. T. Sharp and D. Lee, Rev. Mex. Fis. 20, 203 (1971).
[12]M. Hongoh (Ph. D. thesis, to be submitted for publication).
[13]V. S. Devi and T. Venkatarayudu, J. Math. Phys. 9, 1057 (1968); R. T. Sharp and C. S. Lam, J. Math. Phys. 10, 2033 (1969).
[14]V. S. Devi and M. Moshinsky, J. Math. Phys. 10, 455 (1969); V. Bargmann and M. Moshinsky, Nucl. Phys. 23, 177 (1961); V. S. Devi, Nuovo Cimento A 63, 1 (1970).
[15]R. T. Sharp and S. C. Pieper, J. Math. Phys. 9, 663 (1968).
[16]J.-P. Antoine and D. Speiser, J. Math. Phys. 5, 1226, 1560 (1964).
[17]R. T. Sharp, Nucl. Phys. A 95, 222 (1967).
[18]L. O'Raifeartaigh, A. J. Macfarlane, and P. S. Rao, Bull. Am. Phys. Soc. 10, 483 (1965).

On global embedding

G. H. Goedecke

New Mexico State University, Las Cruces, Mew Mexico 88003
(Received 3 December 1973)

A basis vector treatment of tensor calculus in an N-dimensional (pseudo-) Euclidean space is used to obtain new insights into the geometrical properties of curved Riemannian spaces of smaller dimension which are globally embedded in the N-space. In particular, it is shown that, in general for a globally embedded hypersurface, (i) partial derivatives of internal basis vectors with respect to internal coordinates must in general be expressed as a linear combination of external as well as internal basis vectors, (ii) there exist two different geometrical expressions, always equal in value, for the intrinsic curvature tensor, (iii) the geodesic equation contains more terms than does the usual one; the extra terms vanish for Schwarzschild metric embeddings. These points are illustrated by the example of a 2-sphere embedded in Euclidean 3-space.

INTRODUCTION

Very many advanced physics texts make no mention or use of basis vectors in their presentations of tensor calculus and general relativity.[1] However, a great many authors have made good use of a tetrad formulation in the context of general relativity.[2] Orthonormal triads of basis vectors in a Euclidean 3-space are very familiar; perhaps not quite so familiar are treatments of tensor calculus in N-dimensional spaces which make use of covariant basis vectors (tangent to the coordinate curves), and their duals, contravariant basis vectors (normal to surfaces of constant coordinate).[3]

Our main purpose in this paper is to show that such a basis vector formulation of tensor calculus provides certain new insights into the geometrical properties of curved Riemannian spaces which are globally embedded in larger pseudo-Euclidean spaces. To this end, in what follows we develop just enough of the general formalism to treat embedding; we omit the bulk of standard tensor calculus (transformation properties, covariant derivatives, etc.) which could also be derived easily by our methods. The most important point of our treatment is that the change with location in a hypersurface of the basis vectors belonging to the coordinates of the hypersurface must in general partake of basis vectors exterior to the hypersurface. This result leads to two different geometrical expressions, always equal in value, for the conventional curvature tensor in an embedded hypersurface. As an example, we evaluate the curvature tensor, using both expressions, for the surface of a sphere embedded in Euclidean 3-space; the results agree.

We also find that the equation for the geodesics in an embedded hypersurface in general contains more terms than does the usual geodesic equation. For Schwarzschild metric embeddings,[4] these extra terms vanish, so that there is no contradiction of general relativity.

FORMALISM

Consider an N-dimensional real coordinate manifold, and an associated pseudo-Euclidean linear vector space PE_N. Then there exist real-valued coordinates X^i, $i=1, \ldots, N$, such that the position vector $\mathbf{x}$ and the vector separation $d\mathbf{x}$ of infinitesimally near points may be written

$$\mathbf{x}=\mathbf{E}_i X^i, \quad d\mathbf{x}=\mathbf{E}_i\, dX^i, \tag{1}$$

where $d\mathbf{x}=0$ if and only if $dX^i=0$, all i; the $\{\mathbf{E}_i\}$ are stant basis vectors for PE_N. The metric may be taken as

$$G_{ij}\equiv \mathbf{E}_i\cdot\mathbf{E}_j=G_{ji}=(0,\ i\neq j;\pm 1,\ i=j). \tag{2}$$

This metric is constant, since the E_i are. Define

$$\mathbf{E}^i=G^{ij}\mathbf{E}_j, \quad G^{ij}G_{jk}=\delta^i_k=\mathbf{E}^i\cdot\mathbf{E}_k \tag{3}$$

The set $\{\mathbf{E}^i\}$ is also a basis in PE_N. Define the gradient operator:

$$\nabla\equiv\mathbf{E}^i\frac{\partial}{\partial X^i}\cdot \tag{4}$$

Consider a one–one (in general nonlinear) mapping to coordinates x^i:

$$x^i=f^i(X^1,\ldots,X^N)\longleftrightarrow X^i=F^i(x^1,\ldots,x^N). \tag{5}$$

Define

$$\mathbf{e}_i\equiv\partial_i\mathbf{x}=\frac{\partial X^k}{\partial x^i}\,\mathbf{E}_k=\text{covariant basis vector}, \tag{6}$$

$$\mathbf{e}^i\equiv\nabla x^i=\frac{\partial x^i}{\partial X^k}\,\mathbf{E}^k=\text{contravariant basis vector}, \tag{7}$$

associated with the coordinates x^i. We use the notation $\partial_i\equiv\partial/\partial x^i$. The basis vector $\mathbf{e}_i$ at a point is tangent at that point to the coordinate curve $x^k=(\text{const})^k$, all $k\neq i$, which passes through that point. The vector $\mathbf{e}^i$ at a point is normal at that point to the hypersurface $x^i=\text{const}$, which contains that point. These interpretations apply *mutatis mutandis* in all coordinate systems, including the reference pseudo-Euclidean system. Define the metric:

$$g_{ij}\equiv\mathbf{e}_i\cdot\mathbf{e}_j=g_{ji}, \quad g^{ij}=\mathbf{e}^i\cdot\mathbf{e}^j=g^{ji}; \tag{8}$$

then $g^{ij}g_{jk}=\delta^i_k=\mathbf{e}^i\cdot\mathbf{e}_k$, and g_{ij} is a tensor of second rank, transforming as a product of basis vectors.

Coordinate dependence of basis

Consider

$$\partial_j\mathbf{e}_i=\partial_j\left(\frac{\partial X^m}{\partial x^i}\,\mathbf{E}_m\right)=\frac{\partial^2X^m}{\partial x^j\partial x^i}\,\frac{\partial x^q}{\partial X^m}\,\mathbf{e}_q\equiv\Gamma_{ji}{}^q\mathbf{e}_q, \tag{9}$$

where the last equality defines $\Gamma_{ji}{}^q=\Gamma_{ij}{}^q$. Then

$$\mathbf{e}_k\cdot\partial_j\mathbf{e}_i=g_{kq}\Gamma_{ji}{}^q\equiv[k,\ ji];\ \mathbf{e}_i\cdot\partial_j\mathbf{e}_k=[i,\ jk];$$

adding these and manipulating, we find

$$[i,\ jk]=\tfrac{1}{2}[\partial_jg_{ki}+\partial_kg_{ji}-\partial_ig_{jk}]=g_{im}\Gamma_{jk}{}^m, \tag{10}$$

the Christoffel symbol of the first kind; then Γ_{ij}^{k} is a Christoffel symbol of the second kind, equal to the (negative of) the affine connection.[1] In our treatment there is no need to consider parallel transport or arbitrarily defined affine connections.

Consider the second derivatives:

$$\partial_k\partial_j\mathbf{e}_i=\frac{\partial^3 X^m}{\partial x^k\partial x^j\partial x^i}\,\mathbf{E}_m=(\partial_k\Gamma_{ji}^{q}+\Gamma_{ij}^{m}\Gamma_{km}^{q})\mathbf{e}_q, \tag{11}$$

$$\partial_j\partial_k\mathbf{e}_i=\frac{\partial^3 X^m}{\partial x^j\partial x^k\partial x^i}\,\mathbf{E}_m=(\partial_j\Gamma_{ki}^{q}+\Gamma_{ik}^{m}\Gamma_{jm}^{q})\mathbf{e}_q. \tag{12}$$

These are manifestly equal, so their difference vanishes:

$$e^l\cdot\partial_{[k}\partial_{j]}\mathbf{e}_i=R^l{}_{ikj}=0, \tag{13}$$

where

$$R^l{}_{ikj}\equiv\partial_{[k}\Gamma_{j]i}^{l}+\Gamma_{i[j}^{m}\Gamma_{k]m}^{l} \tag{14}$$

is the curvature tensor; naturally this is zero in a flat space. We use the notation

$$a_{[i}b_{j]}\equiv a_ib_j-a_jb_i. \tag{15}$$

Note that we do not use the $\frac{1}{2}$ usually associated with index antisymmetrization brackets.

GLOBAL EMBEDDING

Consider the M-dimensional space obtained by placing $(N-M)$ constraints on the curvilinear coordinates $x^{M+1},\ldots,x^{N}$: Let

$$x^a=(\text{const})^a,\quad a=M+1,\ldots,N, \tag{16}$$

define the M-dimensional hypersurface Σ_M. The remaining coordinates x^μ, $\mu=1,\ldots,M$, are coordinates for Σ_M. Here and in what follows, Greek indices range and sum over $1,\ldots,M$; Latin indices $(a, b, c, \ldots)$ from the first part of the alphabet range and sum over $M+1,\ldots,N$; Latin indices $(i, j, k, l, \ldots)$ from the later part of the alphabet range and sum over $1,\ldots,N$. Now notice that

$$\partial_\nu\mathbf{e}_\mu=\Gamma_{\nu\mu}^{l}\mathbf{e}_l=\Gamma_{\nu\mu}^{\lambda}\mathbf{e}_\lambda+\Gamma_{\nu\mu}^{a}\mathbf{e}_a; \tag{17}$$

that is, the change of a basis vector with respect to a change of internal coordinate partakes not only of the basis vectors in Σ_M, but also of those outside Σ_M, since $\Gamma_{\nu\mu}^{a}\neq 0$ in general.

For the curvature tensor with all interior indices, we have from Eq. (14),

$$R^\alpha{}_{\beta\gamma\delta}=0=\{\partial_{[\gamma}\Gamma_{\delta]\beta}^{\alpha}+\Gamma_{\beta[\delta}^{\lambda}\Gamma_{\gamma]\lambda}^{\alpha}\}-\Gamma_{\beta[\gamma}^{a}\Gamma_{\delta]a}^{\alpha}, \tag{18}$$

where all quantities are evaluated on the hypersurface Σ_M. Now

$$\Gamma_{\delta\beta}^{\alpha}=g^{\alpha i}[i,\delta\beta]=g^{\alpha\lambda}[\lambda,\delta\beta]+g^{\alpha a}[a,\delta\beta]={}^{M}\Gamma_{\delta\beta}^{\alpha}+\Lambda_{\delta\beta}^{\alpha}, \tag{19}$$

where ${}^{M}\Gamma_{\delta\beta}^{\alpha}$, $\Lambda_{\delta\beta}^{\alpha}$ are defined by the first and second terms in the last equality, respectively. The Christoffel symbol ${}^{M}\Gamma_{\delta\beta}^{\alpha}$ belongs entirely to the hypersurface Σ_M. From Eqs. (14), (18), (19), we have on the hypersurface Σ_M

$$\begin{aligned}{}^{M}R^\alpha{}_{\beta\gamma\delta}=\Gamma_{\beta[\gamma}^{a}\Gamma^{\alpha}{}_{\delta]a}-\partial_{[\gamma}\Lambda_{\delta]\beta}^{\alpha}-{}^{M}\Gamma_{\beta[\delta}^{\lambda}\Lambda_{\gamma]\lambda}^{\alpha}\\ -\Lambda_{\beta[\delta}^{\lambda}\,{}^{M}\Gamma_{\gamma]\lambda}^{\alpha}-\Lambda_{\beta[\delta}^{\lambda}\Lambda_{\gamma]\lambda}^{\alpha},\end{aligned} \tag{20}$$

where the conventional curvature tensor in the M-space is defined by

$${}^{M}R^\alpha{}_{\beta\gamma\delta}\equiv\partial_{[\gamma}{}^{M}\Gamma_{\delta]\beta}^{\alpha}+{}^{M}\Gamma_{\beta[\delta}^{\lambda}\,{}^{M}\Gamma_{\gamma]\lambda}^{\alpha}. \tag{21}$$

Equations similar to Eq. (20), and to its three independent companions which arise from

$$R^\alpha{}_{\beta\gamma d}=0,\quad R^\alpha{}_{\beta cd}=0,\quad R^a{}_{bcd}=0,$$

have been derived by a different method by Szekeres.[5] If the coordinates are such that $(g^{\alpha a})=0$ on Σ_M, the $(\Lambda_{\delta\beta}^{\alpha})=0$ on Σ_M, and Eq. (20) reduces considerably.

We see that our formalism, with global embedding, restricts us to Riemannian curved spaces with symmetric affine connections; but such spaces are the ones pertinent to Einstein's theory of gravitation, and to successful modifications thereof. Theories with nonsymmetric metric and/or nonsymmetric connections have been attempted, but none have been successful enough to replace Einstein's theory.

GEODESICS

In the curvilinear coordinates x^i in our N-space, a curve lying entirely in our embedded M-space is given by $x^i=z^i(s)$, $i=1,\ldots,N$, subject to the constraints $x^a=z^a(s)=k^a=(\text{const})^a$, $a=M+1,\ldots,N$. The N-vector velocity is $\dot{\mathbf{z}}=\mathbf{e}_i\dot{z}^i=\mathbf{e}_\mu\dot{z}^\mu$, since $\dot{z}^a=0$; the dot means d/ds. We take ds to be the (positive) element of arc length along the curve, whereby $ds^2=\pm(d\mathbf{z}\cdot d\mathbf{z})$; the $\pm$ signs are taken so as to make $ds^2>0$. (We omit discussion of null curves.) Therefore $\dot{\mathbf{z}}\cdot\dot{\mathbf{z}}=\pm 1$. A (nonnull) geodesic in our embedded M-space is a curve of extremal arc length between any two points; its equation is the Euler–Lagrange equation

$$0=\frac{\partial L}{\partial z^k}-\frac{d}{ds}\frac{\partial L}{\partial\dot{z}^k}, \tag{22}$$

with

$$L=(\pm g_{ij}\dot{z}^i\dot{z}^j)^{1/2}-\lambda_a(z^a-k^a), \tag{23}$$

where the λ_a are Lagrange undetermined multipliers, $a=M+1,\ldots,N$. Combining the above equations and constraints, we get

$$\Gamma_{\alpha\beta}^{a}\dot{z}^\alpha\dot{z}^\beta\pm g^{ab}\lambda_b=0,\quad a=M+1,\ldots,N, \tag{24}$$

$$(\ddot{z}^\mu+{}^{M}\Gamma_{\alpha\beta}^{\mu}\dot{z}^\alpha\dot{z}^\beta)+(\Lambda_{\alpha\beta}^{\mu}\dot{z}^\alpha\dot{z}^\beta\pm g^{\mu b}\lambda_b)=0,\quad \mu=1,\ldots,M, \tag{25}$$

where we have used Eq. (19). Equation (24) evaluates the Lagrange multipliers. If we had followed the usual development in an *a priori* curved Riemannian M-space, then instead of Eq. (25), we would have the usual geodesic equation

$$\ddot{z}^\mu+{}^{M}\Gamma_{\alpha\beta}^{\mu}\dot{z}^\alpha\dot{z}^\beta=0. \tag{26}$$

In general Eqs. (25) and (26) disagree; but for the Schwarzschild metric embeddings,[4] $g^{\alpha a}=0$ on Σ_M, implying that $\Lambda_{\mu\nu}^{\lambda}=0$ on Σ_M. Since tests of general relativity are really tests of the (exterior) Schwarzschild metric, there is no contradiction. [Note that the extra terms in Eq. (25) always vanish if $g^{\mu a}=0$ on Σ_M, which is the case for any Σ_M if the x^i are orthogonal coordinates.]

SPHERICAL 2-SURFACE

Consider an example in a background Euclidean 3-space E_3, with coordinates X^i, metric $G_{ij}=\delta_{ij}$. Transform to spherical polar coordinates x^i, $(x^1, x^2, x^3) \equiv(\theta, \phi, r)$:

$$X^1=r\sin\theta\cos\phi, \quad X^2=r\sin\theta\sin\phi, \quad X^3=r\cos\theta; \quad (27)$$

with $0\leq r\leq\infty$, $0\leq\phi<2\pi$, $0\leq\theta\leq\pi$. We ignore singular points. Then

$$dX_i dX^i=g_{ij}dx^i dx^j=r^2(d\theta)^2+(r^2\sin^2\theta)(d\phi)^2+(dr)^2,$$

so the metric g_{ij} is diagonal, with

$$g_{11}=r^2=1/g^{11}, \quad g_{22}=r^2\sin^2\theta=1/g^{22}, \quad g_{33}=1=g^{33}. \quad (28)$$

By using Eq. (10), the only nonzero Christoffel symbols are easily found to be

$$\Gamma_{13}^1=\Gamma_{31}^1=1/r, \quad \Gamma_{22}^1=-\sin\theta\cos\theta, \quad \Gamma_{12}^2=\Gamma_{21}^2=\cot\theta,$$
$$\Gamma_{23}^2=\Gamma_{32}^2=1/r, \quad \Gamma_{11}^3=-r, \quad \Gamma_{22}^3=-r\sin^2\theta. \quad (29)$$

A sphere of radius a centered at the origin has the equation $x^3=a=r$, so that here $(x^1, x^2)=(\theta, \phi)$ are interior coordinates, and $x^3=r$ is the exterior coordinate.

By symmetry, the only independent non-zero component of ${}^M R^\alpha{}_{\beta\gamma\delta}$ is

$${}^M R^1{}_{212}=\partial_1\Gamma_{22}^1-\partial_2\Gamma_{12}^1+\Gamma_{22}^\lambda\Gamma_{1\lambda}^1-\Gamma_{12}^\lambda\Gamma_{2\lambda}^1=\partial_1\Gamma_{22}^1-\Gamma_{12}^2\Gamma_{22}^1$$
$$=-\cot\theta(-\sin\theta\cos\theta)+\frac{\partial}{\partial\theta}(-\sin\theta\cos\theta)=\sin^2\theta. \quad (30)$$

Using the right-hand side of Eq. (20), we find the expected agreement:

$$\Gamma_{12}^3\Gamma_{23}^1-\Gamma_{22}^3\Gamma_{13}^1=-(-a\sin^2\theta)(1/a)=\sin^2\theta=R^1{}_{212}. \quad (31)$$

It is trivial to check the zero-valued components of Eq. (20). Notice that

$$\partial_1\mathbf{e}_1=\Gamma_{11}^i\mathbf{e}_i=-a\mathbf{e}_3,$$
$$\partial_1\mathbf{e}_2=\Gamma_{12}^i\mathbf{e}_i=\cot\theta\,\mathbf{e}_2=\partial_2\mathbf{e}_1,$$
$$\partial_2\mathbf{e}_2=\Gamma_{22}^i\mathbf{e}_i=-\sin\theta\cos\theta\,\mathbf{e}_1-a\sin^2\theta\,\mathbf{e}_3.$$

So, in general, interior derivatives of interior basis vectors involve the exterior basis vector $\mathbf{e}_3$.

DISCUSSION

We should briefly discuss embedded vs conventional *a priori* curved Riemannian spaces, from the point of view of our basis vector formalism.

In the *a priori* formulation, the set of M coordinates $\{x^\alpha\}$ is considered to be the whole coordinate manifold. In a basis vector treatment, the vector separation of infinitesimally near points would not be a perfect differential. We would write $d\mathbf{x}=\mathbf{e}_\alpha dx^\alpha$, and we might be tempted to write

$$\partial_\alpha\mathbf{e}_\beta={}^M\tilde{\Gamma}_{\alpha\beta}^\lambda\mathbf{e}_\lambda, \quad (32)$$

where the ${}^M\tilde{\Gamma}_{\alpha\beta}^\lambda$ are proportionality factors to be determined. For a space without torsion, we would then arbitrarily take ${}^M\tilde{\Gamma}_{\alpha\beta}^\lambda={}^M\tilde{\Gamma}_{\beta\alpha}^\lambda$. Then the definitions above would yield ${}^M\tilde{\Gamma}_{\alpha\beta}^\lambda={}^M\Gamma_{\alpha\beta}^\lambda$, (intrinsic) Christoffel symbols of second kind, as usual. But then we could not take $\partial_\alpha\partial_\beta\mathbf{e}\gamma=\partial_\beta\partial_\alpha\mathbf{e}_\gamma$, since, if we did so, the curvature tensor would vanish, a contradiction. Yet it seems strange that these mixed derivatives would have to be unequal, since $\mathbf{e}_\gamma$ is merely a vector-valued function of the coordinates. Our embedding formalism provides insight here: For all curved spaces which may be globally embedded in a pseudo-Euclidean space of higher dimension, the relation (32) above is clearly wrong; it should be our Eq. (17), $\partial_\alpha\mathbf{e}_\beta=\Gamma_{\alpha\beta}^\lambda\mathbf{e}_\lambda+\Gamma_{\alpha\beta}^a\mathbf{e}_a$, which leads to a nonzero intrinsic curvature tensor even through $\partial_\alpha\partial_\beta\mathbf{e}_\gamma=\partial_\beta\partial_\alpha\mathbf{e}_\gamma$. Since it has been shown that all definite metric Riemannian spaces without torsion,[6] and a great many interesting indefinite metric ones,[4] may be globally embedded in larger pseudo-Euclidean spaces, this insight is of some significance.

It is clear that suitable contractions of Eq. (20) yield equations formally identical with Einstein's field equations, where the stress—energy tensor occurring on the right-hand side of the latter is here a purely geometrical object; it is an attractive speculation that coordinates "exterior" to space—time may be associated with "internal" coordinates of the fundamental particles.[7] Doubtless it is this possibility of total geometrization which sustains physicists' interest in the embedding problem.

[1]For example, R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).
[2]For example, see the article by G.C. Debney, R.P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[3]For example, A. Lichnerowicz, *Tensor Calculus* (Methuen, London, 1962).
[4]C. Fronsdal, Phys. Rev. **116**, 778 (1958); J. Rosen, Rev. Mod. Phys. **37**, 204 (1965).
[5]P. Szekeres, Nuovo Cimento **43**, 1062 (1966).
[6]For reference to the original work on embedding, see the papers in Rev. Mod. Phys. **37**, 201 (1965), *Seminar on the Embedding Problem*, esp. the paper by A. Friedman, p. 201.
[7]For a discussion of this idea, see the paper by Y. Ne'emann, Rev. Mod. Phys. **37**, 227 (1965).

On stress-energy tensors

G. H. Goedecke

New Mexico State University, Las Cruces, New Mexico 88003
(Received 3 December 1973)

An old but not well-known formal method is detailed and used to obtain the symmetrized stress–energy tensor from Noether's theorem applied to Poincaré-covariant Lagrangian field theories. A variation in a standard method is presented and used to obtain the corresponding canonical stress–energy tensor, valid in arbitrary curvilinear coordinates, in the limit of Lorentz metric. These two tensors are shown to be equal, in each case, for scalar, Maxwell, Dirac spinor, coupled Maxwell–Dirac, and vector (non-Maxwell) Lagrangian field theories.

INTRODUCTION

In the Maxwell theory in Lorentz metric it is well known that a straightforward application of Noether's theorem, under uniform space–time translation form invariance of the Lagrangian, yields an asymmetric stress–energy tensor,[1] which may be symmetized either by educated guesswork, or by application of a formal scheme.[2] The formal procedure may be followed in general for all Lagrangian field theories in Lorentz metric.

On the other hand, there is the canonical stress–energy tensor,[2] manifestly symmetric, with zero divergence in an arbitrary curvilinear (or curved space) Riemannian metric. This tensor is believed to always reduce to the symmetrized Noether stress–energy tensor in the limit of Lorentz metric. To our knowledge, this presumed equality of the canonical tensor with the Noether tensor has never been proven in general. We have not yet been able to give a general proof.

In what follows, we detail a formal symmetrization method for the Noether tensor, and a formal method for obtaining the canonical tensor in the limit of Lorentz metric. Then we derive these tensors, and show that they are indeed equal, for several Lagrangian field theories: scalar, Maxwell, Dirac spinor, coupled Maxwell–Dirac and vector (non-Maxwell).

NOETHER'S THEOREM

We work in Lorentz metric, $\eta_{\mu\nu}=\mathrm{diag}(111-1)$, coordinates $x^\mu=(\mathbf{x},t)$, natural units, notation $\partial_\mu\equiv\partial/\partial x^\mu$; summation convention over all indices.

Consider fields $\phi_a(x)$ in a Lagrangian field theory with Lagrangian density $\mathcal{L}(\phi_a,\partial_\mu\phi_a)$. Here the index a runs over different components of, say, a 4-vector field, and/or over different fields. If $\mathcal{L}$ is a form-invariant functional of its arguments under a transformation of coordinates and/or fields, then there exists a quantity with zero 4-divergence.[3] For example, if $\mathcal{L}$ is form-invariant under uniform space–time translations, then $\partial_\lambda t^{\lambda\nu}=0$, where

$$t_{\lambda\nu}=\eta_{\lambda\nu}\,\mathcal{L}-(\partial_\nu\phi_a)\frac{\partial\mathcal{L}}{\partial(\partial^\lambda\phi_a)} \tag{1}$$

is identified as the (unsymmetrized) stress–energy (SE) tensor. This tensor is asymmetric in general.

If $\mathcal{L}$ is form-invariant under Lorentz rotations, then $\partial_\mu j^\mu_{\lambda\nu}=0$, where $j_{\mu\lambda\nu}$ is the angular momentum tensor:

$$j_{\mu\lambda\nu}\equiv l_{\mu\lambda\nu}+s_{\mu\lambda\nu}, \tag{2}$$

$$l_{\mu\lambda\nu}\equiv x_\lambda t_{\mu\nu}-x_\nu t_{\mu\lambda}, \tag{3}$$

$$s_{\mu\lambda\nu}\equiv\frac{\partial\mathcal{L}}{\partial(\partial^\mu\phi_a)}(I_{\nu\lambda})_a{}^b\phi_b, \tag{4}$$

and the $(I_{\nu\lambda})=-(I_{\lambda\nu})$ are matrix representations of the generators of Lorentz rotations. Note $s_{\mu\lambda\nu}=-s_{\mu\nu\lambda}$.

$\mathcal{L}$ may also be form-invariant under certain gauge groups, leading to conserved currents, which do not concern us here.

SYMMETRIZED STRESS-ENERGY TENSOR

We sketch the development given by Belinfante.[2] Define

$$\theta_{\lambda\nu}\equiv t_{\lambda\nu}+\partial^\mu f_{\mu\lambda\nu} \tag{5}$$

such that

$$\theta_{\lambda\nu}=\theta_{\nu\lambda},\quad f_{\mu\lambda\nu}=-f_{\lambda\mu\nu}; \tag{6}$$

the last equality implies $\partial_\lambda\theta^{\lambda\nu}=0$.

Now

$$\partial^\mu j_{\mu\lambda\nu}=0=t_{\lambda\nu}-t_{\nu\lambda}+\partial^\mu s_{\mu\lambda\nu} \tag{7}$$

and

$$\theta_{\lambda\nu}-\theta_{\nu\lambda}=0=t_{\lambda\nu}-t_{\nu\lambda}+\partial^\mu(f_{\mu\lambda\nu}-f_{\mu\nu\lambda}), \tag{8}$$

so that we identify

$$s_{\mu\lambda\nu}=f_{\mu\lambda\nu}-f_{\mu\nu\lambda}. \tag{9}$$

Simple manipulation leads to

$$f_{\mu\lambda\nu}=\tfrac{1}{2}(s_{\mu\lambda\nu}-s_{\nu\mu\lambda}+s_{\lambda\nu\mu}). \tag{10}$$

We shall call $\theta_{\lambda\nu}$ the Noether SE tensor.

CANONICAL STRESS-ENERGY TENSOR

Given a Lagrangian field theory in curvilinear coordinates with Riemannian metric $g_{\mu\nu}$, symmetrix affine connection. According to Landau and Lifschitz, the quantity $T_{\lambda\nu}$ satisfies $D_\lambda T^{\lambda\nu}=0$, where $D_\lambda\equiv$ covariant derivative, and

$$T_{\lambda\nu}\equiv\frac{2}{\sqrt{-g}}\left(\partial_\mu\frac{\partial(\sqrt{-g}\mathcal{L})}{\partial(\partial_\mu g^{\lambda\nu})}-\frac{\partial\sqrt{-g}\mathcal{L}}{\partial g^{\lambda\nu}}\right), \tag{11}$$

with

$$g\equiv\det(g_{\alpha\beta}),\quad \mathcal{L}=\mathcal{L}(\phi_a,\ \partial_\mu\phi_a;\ g^{\alpha\beta},\ \partial_\mu g^{\alpha\beta}).$$

We shall let

$$g_{\lambda\nu}=\eta_{\lambda\nu}+h_{\lambda\nu},\quad g^{\lambda\nu}\approx\eta^{\lambda\nu}-h^{\lambda\nu}, \tag{12}$$

and keep terms to first order only in $h^{\lambda\nu}$, $\partial_\mu h^{\lambda\nu}$; the se-

cond equality above is valid to first order, where

$$h^{\lambda\nu} \equiv \eta^{\lambda\alpha}\eta^{\nu\beta}h_{\alpha\beta},\quad h_{\alpha\beta}=h_{\beta\alpha}. \tag{13}$$

Then we get

$$(T_{\lambda\nu})_{h=0}=\left\{\eta_{\lambda\nu}\mathcal{L}+2\frac{\partial\mathcal{L}}{\partial h^{\lambda\nu}}-2\partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu h^{\lambda\nu})}\right\}_{h=0}, \tag{14}$$

where the $\eta_{\lambda\nu}\mathcal{L}$ term comes from $\partial\sqrt{-g}/\partial g^{\lambda\nu}$. We shall call $T_{\lambda\nu}$ the canonical SE tensor.

SCALAR FIELD

Noether SF tensor

Let

$$\mathcal{L}_s=-\tfrac{1}{2}\partial_\mu\phi\ \partial^\mu\phi; \tag{15}$$

a mass term is irrelevant; $\phi(x)$ is the scalar field. Then the motion equations are

$$0=\frac{\delta\mathcal{L}_s}{\delta\phi}\equiv\partial_\mu\frac{\partial\mathcal{L}_s}{\partial(\partial_\mu\phi)}-\frac{\partial\mathcal{L}_s}{\partial\phi}\rightarrow(\partial_\mu\partial^\mu\phi)=0. \tag{16}$$

Under a Lorentz rotation, a scalar field has no components to be mixed (no spin part for the induced field transformation), so that $s^s_{\mu\lambda\nu}=0$. Using Eqs. (1), (16), we find

$$t^s_{\lambda\nu}=\theta^s_{\lambda\nu}=-\tfrac{1}{2}\eta_{\lambda\nu}\,\partial_\mu\phi\,\partial^\mu\phi+\partial_\nu\phi\,\partial_\lambda\phi. \tag{17}$$

Note that $\theta_{44}\geq 0$; when possible for Boson field we choose the over-all sign of the Lagrangian such that, with our definition of $\theta_{\mu\nu}$, we find $\theta_{44}\geq 0$.

Canonical SE tensor

$$\mathcal{L}_s(h)=-\tfrac{1}{2}\partial_\alpha\phi\,\partial_\beta\phi(\eta^{\alpha\beta}-h^{\alpha\beta}). \tag{18}$$

Using Eq. (14), we find

$$(T^s_{\lambda\nu})_{h=0}=-\frac{\eta_{\lambda\nu}}{2}\,\partial_\mu\phi\,\partial^\mu\phi+\partial_\lambda\phi\,\partial_\nu\phi=\theta^s_{\lambda\nu}, \tag{19}$$

where, of course, in the last expression $\partial^\mu\phi=\eta^{\mu\lambda}\partial_\lambda\phi$.

For conciseness in what follows, for each field theory considered we shall merely list the crucial quantities, and the results $\theta_{\lambda\nu}$, $(T_{\lambda\nu})_{h=0}$.

MAXWELL FIELD

$$\mathcal{L}_m=-(1/16\pi)f_{\alpha\beta}f^{\alpha\beta},\quad f_{\alpha\beta}\equiv\partial_\alpha A_\beta-\partial_\beta A_\alpha, \tag{20}$$

$$\frac{\delta\mathcal{L}_m}{\delta A_\mu}=0\rightarrow\partial_\nu f^{\nu\mu}=0,$$

$$(I_{\nu\lambda})_a{}^b=\eta_{\nu a}\delta_\lambda{}^b-\delta_\nu{}^b\eta_{\lambda a},$$

$$s^m_{\mu\lambda\nu}=(1/4\pi)(f_{\mu\lambda}A_\nu-f_{\mu\nu}A_\lambda),$$

$$f^m_{\mu\lambda\nu}=(1/4\pi)f_{\mu\lambda}A_\nu,\quad \partial^\mu f^m_{\mu\lambda\nu}=-(1/4\pi)f_{\lambda\nu}\partial^\mu A_\nu,$$

$$t^m_{\lambda\nu}=-(\eta_{\lambda\nu}/16\pi)f_{\alpha\beta}f^{\alpha\beta}+(1/4\pi)f_{\lambda\mu}\,\partial_\nu A^\mu,$$

$$\mathcal{L}_m(h)=-(1/16\pi)f_{\alpha\beta}f_{\gamma\delta}\,g^{\alpha\gamma}g^{\beta\delta}=\mathcal{L}_m+(1/16\pi)f_{\alpha\beta}f_{\gamma\delta}(\eta^{\alpha\gamma}h^{\beta\delta}+\eta^{\beta\delta}h^{\alpha\gamma}),$$

$$(T^m_{\lambda\nu})_{h=0}=\theta_{\lambda\nu}=-(\eta_{\lambda\nu}/16\pi)f_{\alpha\beta}f^{\alpha\beta}+(1/4\pi)f_{\lambda\mu}f_\nu{}^\mu\ . \tag{21}$$

(Note that the fully covariant $f_{\alpha\beta}$ still involves only ordinary derivatives of A_μ, because of its antisymmetry.) The equality $T=\theta$ is well known for the Maxwell theory.

Dirac spinor field

We choose a real-valued Lagrangian in metric $\eta_{\mu\nu}$:

$$\mathcal{L}_D=-\tfrac{1}{2}(\bar\psi\gamma^\mu\partial_\mu\psi-\partial_\mu\bar\psi\gamma^\mu\psi)-m\bar\psi\psi, \tag{22}$$

where the γ^μ are 4×4 constant matrices satisfying $[\gamma^\mu,\gamma^\nu]_+=2\eta^{\mu\nu}1$, $\gamma_i{}^\dagger=\gamma_i$, $\gamma_4^\dagger=-\gamma_4$, and for definiteness

$$\bar\psi\equiv\psi^\dagger i\gamma^4. \tag{23}$$

We regard $(\psi,\bar\psi)$ as independent (four-component) fields; then

$$\frac{\delta\mathcal{L}}{\delta\bar\psi}=0\rightarrow\gamma^\mu\partial_\mu\psi+m\psi=0,$$

$$\frac{\delta\mathcal{L}}{\delta\psi}=0\rightarrow\partial_\mu\bar\psi\gamma^\mu-m\bar\psi=0,$$

$$I_{\nu\lambda}=\tfrac{1}{2}\sigma_{\nu\lambda},\quad \sigma_{\nu\lambda}\equiv\tfrac{1}{2}[\gamma_\nu,\gamma_\lambda]_-,$$

$$s^D_{\mu\lambda\nu}=\frac{\partial\mathcal{L}}{\partial(\partial^\mu\psi)}I_{\nu\lambda}\psi-\bar\psi\,I_{\nu\lambda}\frac{\partial\mathcal{L}}{\partial(\partial^\mu\bar\psi)}=-\tfrac{1}{4}\bar\psi[\gamma_\mu,\sigma_{\nu\lambda}]_+\psi,$$

$$f^D_{\mu\lambda\nu}=-\tfrac{1}{8}\bar\psi(\gamma_\mu\gamma_\nu\gamma_\lambda-\gamma_\lambda\gamma_\nu\gamma_\mu)\psi,$$

$$\partial^\mu f^D_{\mu\lambda\nu}=-\tfrac{1}{4}(\bar\psi\gamma_\lambda\partial_\nu\psi-\bar\psi\gamma_\nu\partial_\lambda\psi-\partial_\nu\bar\psi\gamma_\lambda\psi+\partial_\lambda\bar\psi\gamma_\nu\psi), \tag{24}$$

$$t^D_{\lambda\nu}=\tfrac{1}{2}(\bar\psi\gamma_\lambda\partial_\nu\psi-\partial_\nu\bar\psi\gamma_\lambda\psi). \tag{25}$$

Note that $\mathcal{L}_D=0$ in virtue of the field equations.

The generally covariant Dirac equation was first presented long ago[5]; it may be written

$$\lambda^\mu(\partial_\mu-\Gamma_\mu)\psi+m\psi=0, \tag{26}$$

where λ^μ, Γ_μ are 4×4 coordinate-dependent matrices satisfying[6]

$$[\lambda^\mu,\lambda^\nu]_+=2g^{\mu\nu}1,\quad \Gamma_\mu=\tfrac{1}{4}[\lambda,\mu\nu]\Sigma^{\nu\lambda},$$

with $[\lambda,\mu\nu]=\frac{1}{2}(\partial_\mu g_{\lambda\nu}+\partial_\nu g_{\lambda\mu}-\partial_\lambda g_{\nu\mu})=$ Christoffel symbol of first kind, and $\Sigma^{\nu\lambda}\equiv\frac{1}{2}[\lambda^\nu,\lambda^\lambda]_-$. The field equations (26) are Euler–Lagrange equations of the Lagrangian $\sqrt{-g}\,\mathcal{L}_D(h)$, where

$$\mathcal{L}_D(h)=-\tfrac{1}{2}\{\bar\psi\lambda^\mu\partial_\mu\psi-\partial_\mu\bar\psi\lambda^\mu\psi-\bar\psi[\lambda^\mu,\Gamma_\mu]_+\psi\}-m\bar\psi\psi.$$

To first order in $h^{\alpha\beta}$, $\partial_\mu h^{\alpha\beta}$, we have

$$\lambda^\mu=\gamma^\mu-\tfrac{1}{2}h^{\mu\alpha}\gamma_\alpha,\quad \lambda_\mu=\gamma_\mu+\tfrac{1}{2}h_{\mu\beta}\gamma^\beta, \tag{27}$$

$$\Gamma_\mu=\tfrac{1}{4}(\partial_\nu h_{\lambda\mu})\sigma^{\nu\lambda},$$

$$[\gamma^\mu,\Gamma_\mu]_+=\tfrac{1}{8}(\partial_\nu h_{\lambda\mu})\{[\gamma^\mu,\sigma^{\nu\lambda}]_+-[\gamma^\lambda,\sigma^{\nu\mu}]_+\}=0, \tag{28}$$

$$\mathcal{L}_D(h)\approx\mathcal{L}_D+\tfrac{1}{8}h^{\mu\alpha}(\bar\psi\gamma_\alpha\partial_\mu\psi+\bar\psi\gamma_\mu\partial_\alpha\psi-\partial_\mu\bar\psi\gamma_\alpha\psi-\partial_\alpha\bar\psi\gamma_\mu\psi). \tag{29}$$

Whether we combine Eqs. (5), (24), (25) to get $\theta^D_{\lambda\nu}$, or Eqs. (14) and (29) to get $(T^D_{\lambda\nu})_{h=0}$, we find

$$(T^D_{\lambda\nu})_{h=0}=\theta^D_{\lambda\nu}=\tfrac{1}{4}(\bar\psi\gamma_\lambda\partial_\nu\psi+\bar\psi\gamma_\nu\partial_\lambda\psi-\partial_\nu\bar\psi\gamma_\lambda\psi-\partial_\lambda\bar\psi\gamma_\nu\psi).$$

COUPLED MAXWELL-DIRAC FIELDS

$$\mathcal{L}_{mD}=\mathcal{L}_m+\mathcal{L}_D+j_\mu A^\mu,\quad j^\mu\equiv ie\bar{\psi}\gamma^\mu\psi.$$

$\mathcal{L}_D+j_\mu A^\mu=0$ in virtue of field equations,

$$\frac{\delta\mathcal{L}}{\delta\bar{\psi}}=0\rightarrow\gamma^\mu(\partial_\mu-ieA_\mu)\psi+m\psi=0,$$

$$\frac{\delta\mathcal{L}}{\delta\psi}=0\rightarrow[(\partial_\mu+ieA_\mu)\bar{\psi}]\gamma^\mu-m\bar{\psi}=0,$$

$$\frac{\delta\mathcal{L}}{\delta A_\mu}=0\rightarrow\partial_\nu f^{\nu\mu}=-4\pi j^\mu,$$

$$t^{mD}_{\lambda\nu}=t^m_{\lambda\nu}+t^D_{\lambda\nu},\quad f^{mD}_{\mu\lambda\nu}=f^m_{\mu\lambda\nu}+f^D_{\mu\lambda\nu},$$

but $\partial^\mu f^{mD}_{\mu\lambda\nu}$ is different than in the free field case. After some manipulation, we get

$$\theta^{mD}_{\lambda\nu}=\theta^m_{\lambda\nu}+\theta^D_{\lambda\nu}-\tfrac{1}{2}(j_\lambda A_\nu+j_\nu A_\lambda). \tag{30}$$

We have for the curvilinear Lagrangian

$$\mathcal{L}_{mD}(h)=\mathcal{L}_m(h)+\mathcal{L}_D(h)+\tfrac{1}{2}(j_\alpha A_\beta+j_\beta A_\alpha)(\eta^{\alpha\beta}-h^{\alpha\beta});$$

inserting this in Eq. (14), we get

$$(T^{mD}_{\lambda\nu})_{h=0}=\theta^{mD}_{\lambda\nu}.$$

VECTOR FIELD

We use for the Lorentz metric Lagrangian density

$$\mathcal{L}_v=-\tfrac{1}{2}(\partial_\alpha\phi_\beta)(\partial^\alpha\phi^\beta).$$

This will provide an example in which the Christoffel symbols actually play a role:

$$\frac{\delta\mathcal{L}_v}{\delta\phi_\mu}=0\rightarrow\partial_\nu\partial^\nu\phi^\mu=0;$$

we take

$$\partial_\mu\phi^\mu=0,$$

$$(I_{\nu\lambda})_a{}^b=\eta_{\nu a}\delta_\lambda{}^b-\delta_\nu{}^b\eta_{\lambda a},$$

$$s^v_{\mu\lambda\nu}=(\partial_\mu\phi_\lambda)\phi_\nu-(\partial_\mu\phi_\nu)\phi_\lambda,$$

$$f^v_{\mu\lambda\nu}=\partial_{[\mu}\phi_{\lambda]}\phi_\nu+\partial_{(\lambda}\phi_{\nu)}\phi_\mu-\partial_{(\mu}\phi_{\nu)}\phi_\lambda,$$

where [ ] ≡ antisymmetric part, ( ) ≡ symmetric part. We have

$$t^v_{\lambda\nu}=-\tfrac{1}{2}\eta_{\lambda\nu}\,\partial_\alpha\phi_\beta\,\partial^\alpha\phi^\beta+\partial_\nu\phi^\mu\,\partial_\lambda\phi_\mu.$$

A simple calculation gives

$$\theta^v_{\lambda\nu}=\tfrac{1}{2}\eta_{\lambda\nu}\,\partial_\alpha\phi_\beta\,\partial^\alpha\phi^\beta+\partial_\nu\phi^\mu\partial_\lambda\phi_\mu-\tfrac{1}{2}(\partial_\lambda\phi_\mu)(\partial^\mu\phi_\nu)-\tfrac{1}{2}(\partial_\nu\phi_\mu)(\partial^\mu\phi_\lambda)+\phi_\mu\,\partial^\mu(\partial_{(\lambda}\phi_{\nu)}). \tag{31}$$

The curvilinear Lagrangian is

$$\mathcal{L}_v(h)=-\tfrac{1}{2}D_\alpha\phi_\beta\,D_\gamma\phi_\delta\,g^{\alpha\gamma}g^{\beta\delta}$$

with

$$D_\alpha\phi_\beta=\partial_\alpha\phi_\beta-\Gamma^\lambda_{\alpha\beta}\,\phi_\lambda,\quad \Gamma^\lambda_{\alpha\beta}\equiv g^{\lambda\delta}[\delta,\alpha\beta].$$

To first order in $h^{\alpha\beta}$, $\partial^\mu h^{\alpha\beta}$, this is

$$\mathcal{L}_v(h)=\mathcal{L}_v+\tfrac{1}{2}(\partial_\alpha\phi_\beta)(\partial_\gamma\phi_\delta)(\eta^{\alpha\gamma}h^{\beta\delta}+\eta^{\beta\delta}h^{\alpha\gamma})+\tfrac{1}{4}(\partial_{(\gamma}\phi_{\delta)}\,\phi_\sigma+\partial_{(\gamma}\phi_{\sigma)}\,\phi_\delta)\partial^\gamma h^{\sigma\delta}+\tfrac{1}{4}[\partial_{(\gamma}\phi_{\delta)}\,\phi_\sigma+\partial_{(\sigma}\phi_{\delta)}\,\phi_\gamma]\partial^\delta h^{\sigma\gamma}-\tfrac{1}{2}\partial_{(\gamma}\phi_{\delta)}\,\phi_\sigma\partial^\sigma h^{\gamma\delta}.$$

Using Eq. (14), we get

$$(T^v_{\lambda\nu})_{h=0}=\theta^v_{\lambda\nu}.$$

The Lagrangian $\mathcal{L}_v$ has been proposed for (massive) vector field theories; in addition to its lack of gauge invariance, its canonical SE tensor is quite clumsy; and θ^v_{44} is not manifestly positive semidefinite.

SUMMARY

Considering the above evidence that the canonical SE tensor is equal to the Noether SE tensor, it would seem that we really ought to be able to provide a general proof of their identity, perhaps by counting indices in some way. To our knowledge, a general proof has never been given, and we have not yet found one.

[1]L. Landau and L. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, MA, 1951).
[2]F.J. Belinfante, Physica **6**, 887 (1939).
[3]For a discussion of Noether's theorem, see, e.g., P. Roman, *Theory of Elementary Particles* (North-Holland, Amsterdam, 1961), 2nd ed., Chap. IV.
[5]V. Fock, Z. Physik **57**, 261 (1929); Schouten, J. Math. & Phys. **10**, 239 (1931); E. Schrödinger, Berl. Ber., 105 (1932); V. Bargmann, Berl. Ber., 346 (1932).
[6]See, e.g., R. Ingraham, Nuovo Cimento **10**, (1953), for a discussion of the quantity r_μ.

Higher-order diagrams in quantum gravity and the continuous dimension method*

D. M. Capper

International Centre for Theoretical Physics, Trieste, Italy

G. Leibbrandt

Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada
(Received 23 October 1973)

The following three third-order Feynman diagrams in quantum gravity are treated in the context of dimensional regularization: (a) the pure graviton triangle diagram and (b) two massless tadpole diagrams. It is shown that the graviton vertex integral is free of ultraviolet and infrared divergences as we approach Minkowski space ($\omega=2$). It is also shown that the ultraviolet divergences arising from two-loop massless tadpole integrals either vanish completely as $\omega\to 2$, or manifest themselves as poles of Weierstrass' gamma function.

1. INTRODUCTION

The continuous dimension method[1-3] has recently been employed in the calculation of the graviton self-energy and the associated fictitious particle loop.[4] That calculation, carried out to order K^2 in the gravitational coupling constant ($K^2=32\pi G$), demonstrated (i) that the basic graviton loop can indeed be regularized and (ii) that the procedure of dimensional regularization is gauge invariant in the sense that it respects the Slavnov–Ward identities.

The purpose of this article is to apply the continuous dimension method to the following higher-order diagrams arising in quantum gravity: (a) to the graviton triangle diagram and (b) to two graviton tadpoles, all diagrams being of order K^3 in the gravitational coupling constant.[5] Since for *massless* propagators the various integrals become ill defined when the regulating parameter $\omega\to 2^+$, the procedure of Refs. 1–3 is not applicable to the massless diagrams considered here. To treat the latter, one must introduce[6] an analytic function $f(\omega)$ which vanishes as we approach physical 4-space ($\omega=2$). [See Eq. (3) and also the Appendix]

2. PREVIOUS RESULTS

A detailed discussion of the technique of dimensional regularization as applied to *massless* fields can be found in Refs. 4 and 6. The basic idea is to define each momentum space integral over a complex 2ω-dimensional Euclidean space, to parametrize each propagator according to

$$1/q^2=\int_0^\infty dx\exp(-xq^2),\quad q^2>0, \tag{1}$$

and then to evaluate the various integrals by means of the generalized gaussian formula

$$\int d^{2\omega}q\exp(-xq^2+2b\cdot q)=(\pi/x)^\omega\exp[(b^2/x)-xf(\omega)],\quad x>0, \tag{2}$$

where the continuity function $f(\omega)$, such as[7]

$$f(\omega)=1-\cos(2\pi\cos(2\pi\cos(\cdots(\cos 2\pi\omega)\cdots))), \tag{3}$$

must satisfy specific conditions, in particular analyticity in the complex regulating parameter ω (cf. the Appendix). We also note that the structure of $f(\omega)$ in (3) in terms of *nested* cosine functions, where n is a finite integer, is superior to that previously chosen,[4,8] especially with regard to higher-order diagrams. The prescription (1)–(2) is not gauge invariant for $\omega\neq 2$. Since it is possible, however, with our particular $f(\omega)$, to make arbitrarily high—but not all—derivatives of $f(\omega)$ vanish, the technique discussed here can be made *gauge invariant to any finite order.*

The derivations in the next two sections depend extensively on results previously obtained with the aid of $f(\omega)$:

$$\text{(i)}\qquad \int d^{2\omega}q(q^2)^n=0,\quad n=0,\ 1,\ 2,\ \cdots, \tag{4a}$$

$$\text{(ii) (a)}\qquad \int d^{2\omega}q(q^2)^{-1}=\pi^\omega\Gamma(1-\omega)[f(\omega)]^{\omega-1}, \tag{4b}$$

$$\text{(b)}\qquad \lim_{\omega\to 2^+}\int d^{2\omega}q(q^2)^{-1}=0, \tag{4c}$$

$$\text{(iii)}\qquad \int d^{2\omega}q[q^2(q-p)^2]^{-1}=\pi^\omega\Gamma(2-\omega)\int_0^1 d\xi[p^2\xi(1-\xi)+f(\omega)]^{\omega-2}. \tag{4d}$$

Results (i), (ii), and (iii) are discussed in detail in Refs. 8, 6, and 4, respectively.

3. HIGHER-ORDER TADPOLES

We now examine the two third-order tadpoles shown in Figs. 1 and 2. These diagrams contain only massless graviton lines.

(a) The *basic* integral connected with Fig. 1 is given by

$$I_1=\int\int d^{2\omega}q_1d^{2\omega}q_2[q_1^2q_2^2(q_1-q_2)^2]^{-1}. \tag{5}$$

Applying formulas (1) and (2) or (4d), we can show that

$$I_1=\pi^\omega\Gamma(2-\omega)\int_0^1 d\xi$$

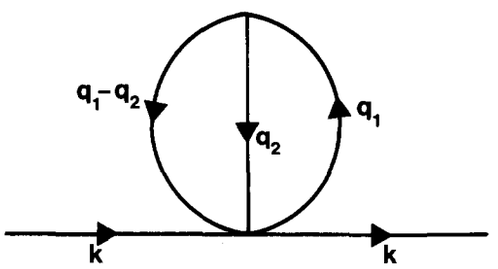


FIG. 1. Third-order massless tadpole.

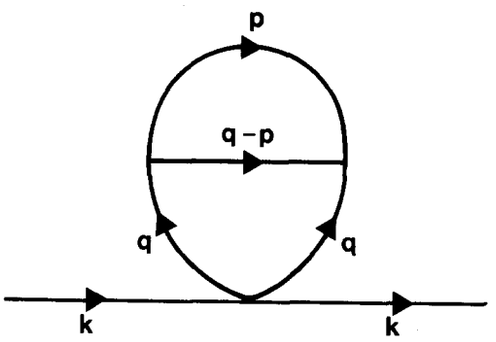


FIG. 2. Third-order massless tadpole.

$$\times \int d^{2\omega} q_1 (q_1^2)^{-1} [q_1^2 \xi(1-\xi) + f(\omega)]^{\omega-2}, \tag{6}$$

where $f(\omega)$ was introduced in Eqs. (2), (3). Expansion about $q_1^2 = 0$ yields

$$\begin{aligned} I_1 = {} & \pi^\omega \Gamma(2-\omega) \int_0^1 d\xi \ (\int d^{2\omega} q_1 (q_1^2)^{-1} \\ & \times [f^{\omega-2} + q_1^2 \xi(1-\xi)(\omega-2) f^{\omega-3} \\ & + \tfrac{1}{2}\xi^2 (1-\xi)^2 (\omega-2)(\omega-3) q_1^4 f^{\omega-4} + \cdots \\ & + (1/n!)(1-\xi)^n \xi^n (\omega-2) \\ & \times \cdots (\omega-1-n)(q_1^2)^n f^{\omega-2-n} + \cdots]). \end{aligned} \tag{7}$$

According to formula (4a), all terms in (7) vanish except the first one. Consequently

$$\begin{aligned} I_1 &= \pi^\omega \Gamma(2-\omega) f^{\omega-2} \int d^{2\omega} q_1 (q_1^2)^{-1} \\ &= \pi^{2\omega} \Gamma(1-\omega) \Gamma(2-\omega) f^{2\omega-3}, \end{aligned} \tag{8}$$

which is well behaved for noninteger ω. As we approach Minkowski space ($2\omega = 4$), we find that the basic integral associated with the massless tadpole in Fig. 1 vanishes:

$$\lim_{\omega \to 2} I_1 = 0. \tag{9}$$

(b) In this section we examine the two-loop tadpole integral

$$I_2 = \int \int d^{2\omega} q \, d^{2\omega} p [p^2 (q-p)^2 q^4]^{-1}, \tag{10}$$

associated with Fig. 2. Again all lines are massless graviton lines. Proceeding as in (a) and using result (4a), in conjunction with the formula

$$\int d^{2\omega} q \, q^{-4} = \pi^\omega \Gamma(2-\omega) f^{\omega-2}, \tag{11}$$

we obtain

$$I_2 = \pi^\omega \Gamma(2-\omega) \int d^{2\omega} q \, q^{-4} \int_0^1 d\xi [q^2 \xi(1-\xi) + f]^{\omega-2} \tag{12}$$

$$= \pi^{2\omega} \Gamma(2-\omega) \Gamma(2-\omega) f^{2\omega-4}. \tag{13}$$

I_2 is well defined for all values of the regulating parameter ω, except $\omega = 2, 3, \cdots$, but contrary to I_1, this tadpole integral does *not* vanish in (physical) 4-space. In fact, as $\omega \to 2^+$, both single and double poles emerge. Thus naively normal ordering massless theories (such as gravity) probably would not produce gauge invariant results in higher orders.

The question of course is how one should interpret the nonzero tadpole contribution I_2. There are two alternatives.

(a) It is possible that the finite I_2 value will be cancelled by other terms when all diagrams to order K^3 are added together. Unfortunately, without calculating all such diagrams, one cannot be sure that a cancellation does not take place. We are not certain at this stage of our programme what the physical implications of a non-vanishing tadpole diagram would be.

(b) If the I_2 contribution is *not* cancelled by similar third-order terms, it may still be possible to eliminate the associated single and double poles by suitable counterterms in the Lagrangian.

4. THE TRIANGLE DIAGRAM

We next apply the technique of dimensional regularization to the basic integral associated with the pure graviton triangle diagram shown in Fig. 3. The integral reads (in Euclidean space):

$$\begin{aligned} I_3 &= \int d^{2\omega} k [k^2 (k-p_2)^2 (k+p_3)^2]^{-1} \\ &= \int_0^\infty \int_0^\infty \int_0^\infty d\alpha d\beta d\gamma (\exp[-(\beta p_3^2 + \gamma p_2^2)]) \int d^{2\omega} k \\ &\quad \times \exp[-(\alpha+\beta+\gamma)k^2 + 2k \cdot (\gamma p_2 - \beta p_3)]. \end{aligned} \tag{14}$$

Integrating over k-space with the help of Eq. (2) and noting that $(p_2 + p_3)^2 = p_1^2$, we obtain

$$\begin{aligned} I_3 = {} & \pi^\omega \int_0^\infty d\alpha\beta d\gamma (\alpha+\beta+\gamma)^{-\omega} \\ & \times \exp[-(\alpha\beta p_3^2 + \alpha\gamma p_2^2 + \gamma\beta p_1^2)(\alpha+\beta+\gamma)^{-1} \\ & - (\alpha+\beta+\gamma) f(\omega)]. \end{aligned} \tag{15}$$

It is convenient now to introduce new variables ξ, τ, λ such that

$$\alpha = \xi\lambda, \quad \beta = \tau\lambda, \quad \gamma = \lambda(1-\xi-\tau), \tag{16}$$

the Jacobian of the transformation being $|J| = \lambda^2$. Integration over λ yields

$$I_3 = \pi^\omega (p_1^2)^{\omega-3} \Gamma(3-\omega) \int_0^1 d\xi \int_0^{1-\xi} d\tau (-\tau^2 + a\tau + b)^{\omega-3}, \tag{17}$$

where

$$a = (1-\xi) - \xi(p_2^2 - p_3^2)/p_1^2, \quad f_0(\omega) \equiv f(\omega)/p_1^2, \tag{18}$$

$$b = f_0(\omega) + \xi(1-\xi) p_2^2 / p_1^2.$$

The τ integral in (17) may be written compactly as the difference between two hypergeometric functions:

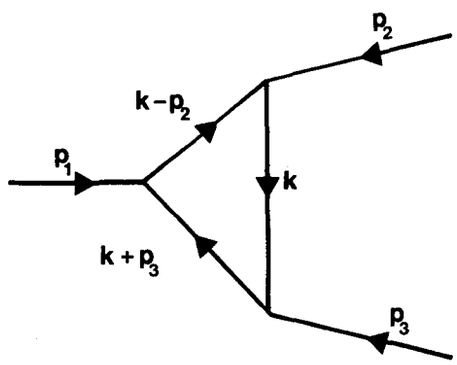


FIG. 3. Pure graviton triangle diagram.

$$I_3 = \tfrac{1}{2}\pi^{\omega}(p_1^2)^{\omega-3}[\Gamma(3-\omega)/(\omega-2)] \int_0^1 d\xi (c_0 + c_1\xi + c_2\xi^2)^{-1/2} \times [z_1^{\omega-2}\, {}_2F_1(\omega-2, \tfrac{1}{2}; \omega-1; z_1/R^2) - z_0^{\omega-2}\, {}_2F_1(\omega-2, \tfrac{1}{2}; \omega-1; z_0/R^2)], \quad (19)$$

where

$$R = (c_0 + c_1\xi + c_2\xi^2)^{1/2}, \quad c_0 = \tfrac{1}{4}(1+4f_0),$$
$$c_1 = -(p_2\cdot p_3)/p_1^2, \quad c_2 = [(p_2\cdot p_3)^2 - p_2^2 p_3^2]/p_1^4 \quad (20)$$
$$z_0 = f_0 + \xi(1-\xi)p_2^2/p_1^2, \quad z_1 = f_0 + \xi(1-\xi)p_3^2/p_1^2.$$

Again, as long as ω remains complex, the right-hand side of (19) is a perfectly well-behaved function of ω. The presence of $f(\omega)$ in (19) moreover ensures the absence of *real* poles in the ξ integrand. In particular the conventional end point singularities $\xi = 0,\ 1$ are absent now.

Since our primary concern is the behavior of I_3 near $\omega = 2$, we shall first write (19) in series form:

$$I_3 = \tfrac{1}{2}\pi^{\omega}(p_1^2)^{\omega-3}\,\Gamma(3-\omega)\int_0^1 \frac{d\xi}{(R^2)^{1/2}} \times\left(\frac{z_1^{\omega-2} - z_0^{\omega-2}}{\omega-2} + \frac{z_1^{\omega-1} - z_0^{\omega-1}}{2(\omega-1)R^2} + \frac{3(z_1^{\omega} - z_0^{\omega})}{8\omega R^4} + \cdots\right). \quad (21)$$

Since

$$\lim_{\omega\to 2^+}\left(\frac{z_1^{\omega-2} - z_0^{\omega-2}}{\omega-2}\right) = \log(p_3^2/p_2^2), \quad (22)$$

we conclude that the vertex integral (14) is indeed free of ultraviolet divergences near $\omega = 2$. This is in complete agreement with results obtained by more conventional techniques.

5. SUMMARY AND DISCUSSION

We have applied the continuous dimension method to several third-order Feynman diagrams in quantum gravity. It was found that the graviton vertex integral is indeed free of both ultraviolet and infrared divergences when the regulating parameter $\omega \to 2^+$. It was also shown that the ultraviolet divergences arising from two-loop massless tadpole integrals either vanish completely as $\omega \to 2^+$ or they manifest themselves as poles of Weierstrass' gamma function. The principal conclusion to be drawn here is that higher-order tadpoles do *not* vanish in general (contrary to lowest-order tadpoles). Their contributions must therefore be included in any satisfactory renormalization program.

Results (8), (13), and (19) may be utilized in the calculation of all graviton–graviton scattering diagrams to order K^3, including the appropriate fictitious particle loops. This complete program is rather lengthy, involving well over twenty thousand terms.

Finally it is essential to realize in connection with the two-loop tadpole integral (5) that the *same* ω was employed for both loops. (A similar statement holds for I_2.) The prescription differs from that given by Ashmore,[2] who suggests a *different* ω for each new loop.

APPENDIX

The continuity function

$$f(\omega) = 1 - \cos(2\pi\cos(2\pi\cos(\cdots(\cos 2\pi\omega)\cdots))) \quad (A1)$$

was introduced in the extended definition of the generalized Gaussian integral (2) and satisfies the following properties:

(i) $f(\omega)$ is a nonzero analytic function of the complex variable $\omega = \sigma + i\tau$;

(ii) $f(\omega) = 0$ for $\omega = \pm\lambda/2,\ \lambda = 0, 1, 2, \cdots$;

(iii) $f^{(l)}(\omega) = 0$ for $\omega = \pm\lambda/2,\ \lambda = 0, 1, 2, \cdots$, and $l + 1 < 2^n$, where n denotes the number of nested cosine functions in (A1); l is the number of derivatives;

(iv) $f^{(2^n)}(\omega) \neq 0$ for $\omega = \pm\lambda/2,\ \lambda = 0, 1, 2, \cdots$;

(v) $\mathrm{Re}[f(\omega)] > 0$ for *any* $\mathrm{Re}(\omega) \neq \pm\lambda/2,\ \lambda = 0, 1, 2, \cdots$, and *some* $\mathrm{Im}\omega$. (A2)

Although $f(\omega)$ is itself analytic, expressions like $f^{2\omega}$, $f^{2\omega-3}$ etc. will clearly not be analytic for general ω. It is therefore imperative—*before* using Taylor's series expansion and before invoking the principle of analytic continuation—to extract from $f^{2\omega}$, $f^{2\omega-3}$, etc. appropriate analytic branches. For example,

$$[f(\omega)]^{2\omega-3} = \exp[(2\omega-3)\mathrm{Log} f(\omega)] \quad (A3)$$

is such an analytic branch provided $|f(\omega)| > 0$ and $-\pi < \phi < \pi$, where $f(\omega) = |f(\omega)|\exp(i\phi)$. Having selected an analytic branch of $f^{2\omega-3}$, we can then expand $f^{2\omega-3}$ in a Taylor's series about the point $\omega = 2 + \epsilon\exp(i\alpha)$, $\epsilon > 0$ and $0 \leq \alpha \leq 2\pi$, allowing $\epsilon \to 0$ at the end of the calculation.

*Research supported in part by the National Research Council of Canada under Grant No. A8063.

[1] G. 't Hooft and M. Veltman, Nucl. Phys. B 44, 189 (1972).
[2] J.F. Ashmore, Lett. Nuovo Cimento 4, 289 (1972); Commun. Math. Phys. 29, 177 (1973).
[3] C.G. Bollini and J.J. Giambiagi, Nuovo Cimento B 12, 20 (1972).
[4] D.M. Capper, G. Leibbrandt, and M. Ramon Medrano, Phys. Rev. D 8, 4320 (1973).
[5] We employ natural units, $\hbar = c = 1$, throughout this paper, in which case $K^2 = 32\pi G \simeq 4\times10^{-44}(m_e)^{-2} \simeq 4\times10^{-38}(\mathrm{GeV})^{-2}$, where G is the newtonian constant and m_e the mass of the electron. All calculations are carried out in complex dimensional Euclidean space.
[6] D.M. Capper and G. Leibbrandt, J. Math. Phys. 15, 82 (1974).
[7] We are grateful to Dr. H.P. Heinig and the referee for their critical comments concerning this continuity function.
[8] D.M. Capper and G. Leibbrandt, J. Math. Phys. 15, 86 (1974).

A new method for the study of Feynman perturbation series

David S. Kershaw

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742
(Received 19 March 1973; final revised manuscript received 11 July 1973)

We have systematically studied a finite-dimensional version of Feynman perturbation series with arbitrary matrices for propagators and arbitrary tensors for interaction couplings. In this formalism we obtain many old results in an elegant new manner, and we obtain a completely new fundamental integral representation for the Green's functions of boson field theories which rigorously defines these functions in our finite-dimensional version and which we hope will enable us to calculate the Green's functions in the physically important infinite-dimensional case.

INTRODUCTION

From one point of view, the central mathematical problem of nonrelativistic quantum mechanics is to sum the infinite series:

$$\mathcal{R}(\lambda, M) = 1 + \lambda M + \lambda^2 M^2 + \lambda^3 M^3 + \cdots,$$

where M is a given infinite-dimensional matrix operating on a Hilbert space and λ is an arbitrary complex number. All physical consequences of the theory are easily extracted from $\mathcal{R}(\lambda, M)$ once it has been calculated. Many years before these problems were studied and solved, the corresponding problem for an arbitrary finite-dimensional matrix M was solved in closed form via the theory of determinants. Furthermore, the Fredholm theory, which is the basic theory for the infinite-dimensional case is based on this prior solution of the finite-dimensional case.

Now let us consider Feynman perturbation theory. We may regard the Feynman propagator as a given infinite-dimensional matrix and the interaction coupling as a given infinite dimensional tensor. The corresponding Hilbert space may be regarded as a space of functions of Minkowski space—time and a spinor representation of $SL(2, c)$. All physical consequences of the theory are easily extracted from the Green's functions which are the sums of infinite series of Feynman graphs. The central problem of Feynman perturbation theory, and perhaps of relativistic quantum mechanics, is to sum these infinite series. Historically, the corresponding problem for arbitrary finite-dimensional propagator matrices and arbitrary finite-dimensional coupling tensors has to my knowledge never been investigated, and perhaps this may prove to be as potent an approach in Feynman perturbation theory as it has been in the theory of linear integral equations.

Thus our new method for the study of Feynman perturbation series consists of writing down the perturbation series for the Green's functions of a given Lagrangian field theory and then replacing each propagator in the theory by an arbitrary finite-dimensional matrix and replacing each interaction coupling in the theory by an arbitrary finite-dimensional tensor. We then sum the infinite series in closed form, and we study the behavior of the Green's functions as functions of the arbitrary matrices and tensors.

We wish to emphasize two points. First, what we have done is to regard the Feynman perturbation series as a functional whose domain is a set of linear spaces of tensors (the propagators and interaction couplings of the theory) and whose range lies in a set of linear spaces of tensors (the Green's functions of the theory). We study the behavior of this functional on the whole domain, not just on those particular elements of the domain singled out by physical considerations (e.g., the Feyman propagators and the local interaction coupling). Second, many of the results which in the past have been obtained most elegantly by the introduction of arbitrary external sources and functional differentiation with respect to these sources are obtained more simply in our formulation by functional differentiation with respect to the arbitrary coupling tensors and propagator matrices without any necessity for the introduction of external sources.

Finally we wish to state our attitude toward renormalization theory. Throughout this paper we ignore the problems of renormalization. All expressions that appear are unrenormalized. Our treatment is entirely rigorous because for finite-dimensional matrices and tensors there are no divergences. Our general approach is to first sum the unrenormalized perturbation series functionals in closed form on domains where no divergences arise and then to try to extend these sums to larger domains which include the physical propagators and couplings. The divergences if any which arise in this extension of domain will, we hope, be more easily manageable and simpler in appearance than the usual graph by graph renormalization analysis of perturbation theory. A good example of this is the nonrelativistic Coulomb problem where the resolvent $\mathcal{R}(\lambda, M)$ is explicitly calculable[1] and involves no divergences while all terms of the perturbation expansion after the first are divergent, and the renormalization theory is relatively complicated.[2]

I. THE BOSON CASE

A. Formulation of the problem

We consider explicitly the case of scalar bonsons interacting through a local $g\phi^4(x)$ coupling. Our methods may be easily extended to other cases. The perturbation series for the n-point Green's function may then be succinctly expressed in the form

$$G_n(x_1, \ldots, x_n) = \langle 0 \,|\, a(x_1)a(x_2)\cdots a(x_n) \times \exp\{ig \int d^4x [a(x)]^4\} \times \exp\left[\tfrac{1}{2} \int d^4x_1 d^4x_2 a^*(x_1)a^*(x_2)\Delta_F(x_1 - x_2)\right] | \, 0\rangle, \tag{1}$$

where $a^*(x)$ and $a(x)$ are creation and destruction operators satisfying

$$[a(x_1),\ a(x_2)]=0,$$

$$[a(x_1),\ a^+(x_2)]=\delta^4(x_1-x_2),$$

the state $|0\rangle$ is defined by $a(x)|0\rangle=0$ for all x, and $\Delta_F(x)$ is the Feynman propagator. Equation (1) is derived in Appendix A.

In keeping with our general program we replace the continuous index x by a discrete finite index, $i=1, 2, \ldots, N$. We then replace

$$ig\int d^4x[a(x)]^4$$

by

$$\sum_{i_1=1}^{N}\cdots\sum_{i_4=1}^{N} M_{i_1i_2i_3i_4}\, a_{i_1}a_{i_2}a_{i_3}a_{i_4}=\mu,$$

and we replace

$$\tfrac{1}{2}\int d^4x_1 d^4x_2\, a^+(x_1)\,a^+(x_2)\Delta_F(x_1-x_2)$$

by

$$\tfrac{1}{2}\sum_{i,j=1}^{N} D_{ij}a_i^+a_j^+=\delta,$$

where $M_{i_1\cdots i_4}$ is an arbitrary 4-tensor and D_{ij} is an arbitrary matrix. The a_i and a_i^+ now satisfy

$$[a_i, a_j]=0$$

and

$$[a_i, a_j^+]=\delta_{ij},$$

and

$$a_i|0\rangle=0 \quad \text{for all } i=1,\ldots,N.$$

The object we wish to study is the n-point Green's function

$$G_n(i_1,\ldots,i_n)=\langle 0|\, a_{i_1}a_{i_2}\cdots a_{i_n}\, e^{\mu}e^{\delta}\,|0\rangle.$$

G_n is a function mapping an arbitrary 4-tensor and 2-tensor into a symmetric n-tensor.

B. Only the vacuum graphs need be considered

We first observe that we need only study the vacuum expectation value

$$G_0=G(M,D)=\langle 0|\, e^{\mu}\,e^{\delta}\,|0\rangle,$$

since G_n may always be obtained by taking appropriate partial derivatives of G with respect to M and D. If n is odd, then $G_n=0$. If $n=4s$ where s is an integer, then clearly

$$G_{(4s)}(i_1,\ldots,i_{(4s)})=\frac{\partial^s G(M,D)}{\partial M_{i_1i_2i_3i_4}\cdots\partial M_{i_{(4s-3)}i_{(4s-2)}i_{(4s-1)}i_{(4s)}}}.$$

If $n=4s+2$, let

$$\begin{aligned}\tilde{a}_i&=\exp(-\delta)\,a_i\exp(+\delta)\\&=a_i+[a_i,\delta]+(1/2!)[[a_i,\delta],\delta]\cdots\\&=a_i+\sum_j \mathcal{D}_{ij}a_j^+, \qquad \text{where } \mathcal{D}_{ij}=\tfrac{1/2}{}(D_{ij}+D_{ji}).\end{aligned}$$

Then we find that

$$\begin{aligned}G_n(i_1,\ldots,i_{(4s)},j_1,j_2)&=\langle 0|\, a_{i_1}\cdots a_{i_{(4s)}}a_{j_1}a_{j_2}\,e^{\mu}e^{\delta}\,|0\rangle\\&=\langle 0|\, a_{i_1}\cdots a_{i_{(4s)}}\,e^{\mu}e^{\delta}\,\tilde{a}_{j_1}\tilde{a}_{j_2}|0\rangle\\&=\langle 0|\, a_{i_1}\cdots a_{i_{(4s)}}\,e^{\mu}e^{\delta}\Big(\mathcal{D}_{j_1j_2}+\sum_{k_1k_2}\mathcal{D}_{j_1k_1}\mathcal{D}_{j_2k_2}a_{k_1}^+a_{k_2}^+\Big)|0\rangle\\&=\mathcal{D}_{j_1j_2}G_{(4s)}(i_1,\ldots,i_{(4s)})+2\sum_{k_1k_2}\mathcal{D}_{j_1k_1}\mathcal{D}_{j_2k_2}\frac{\partial G_{(4s)}(i_1,\ldots,i_{(4s)})}{\partial D_{k_1k_2}}.\end{aligned}$$

C. Reduction in the number of variables

Next assume that $\det(\mathcal{D})\neq 0$. Then as is shown in Appendix B, $(\mathcal{D}^{\pm 1/2})_{ij}$ always exists and is symmetric. Now let

$$a_i\to\sum_j(\mathcal{D}^{1/2})_{ij}a_j=\bar{a}_i,$$

$$a_i^+\to\sum_j(\mathcal{D}^{-1/2})_{ij}a_j^+=\overline{(a_i^+)}.$$

Since $\bar{a}_i$ and $\overline{(a_i^+)}$ satisfy the same commutation relations as a_i and a_i^+, $G(M,D)$ is invariant under this substitution. Therefore

$$\begin{aligned}G(M,D)&=G(\mathcal{M})\\&=\langle 0|\exp\Big(\sum_{i_1\cdots i_4}\mathcal{M}_{i_1\cdots i_4}\,a_{i_1}\cdots a_{i_4}\Big)\ \exp\Big(\tfrac{1}{2}\sum_i(a_i^+)^2\Big)\,|0\rangle,\end{aligned}\tag{2}$$

where

$$\mathcal{M}_{i_1\cdots i_4}=\sum_{j_1\cdots j_4}M_{j_1\cdots j_4}(\mathcal{D}^{1/2})_{j_1i_1}\cdots(\mathcal{D}^{1/2})_{j_4i_4}.$$

D. The differential equation for $G(\mathcal{M})$

Let

$$\alpha=\sum_{i_1\cdots i_4}\xi_{i_1i_2i_3i_4}\quad a_{i_1}a_{i_2}a_{i_3}a_{i_4}$$

and

$$\beta=\tfrac{1}{2}\sum_{ijk}\eta_{ij}\eta_{ik}a_j^+a_k^+,$$

where ξ and η are arbitrary tensors except that $\det(\eta)\neq 0$. Let

$$F(\xi,\eta)=\langle 0|\, e^{\alpha}e^{\beta}\,|0\rangle.$$

Make the substitution

$$a_i\to\sum_j a_j\eta_{ji},$$

$$a_i^+\to\sum_j(\eta^{-1})_{ij}a_j^+.$$

Then as in Eq. (2) we find that

$$F(\xi,\eta)=G(\mathcal{M}),$$

where now

$$\mathcal{M}_{i_1\cdots i_4}=\sum_{j_1\cdots j_4}\eta_{i_1j_1}\cdots\eta_{i_4j_4}\xi_{j_1\cdots j_4}$$

Differentiating, we find that

$$\begin{aligned}\frac{\partial F(\xi,\eta)}{\partial\xi_{i_1\cdots i_4}}&=\langle 0|\, e^{\alpha}a_{i_1}\cdots a_{i_4}e^{\beta}|0\rangle\\&=\langle 0|\, e^{\alpha}e^{\beta}\tilde{a}_{i_1}\cdots\tilde{a}_{i_4}|0\rangle,\end{aligned}$$

where now

$$\tilde{a}_i=e^{-\beta}a_ie^{+\beta}=a_i+\sum_{jk}\eta_{ji}\eta_{jk}a_k^+.$$

Commuting the a_i's through to the right, multiplying both sides by $\eta^{-1}_{i_1j_1}\cdots\eta^{-1}_{i_4j_4}$, summing over i_1 through i_4, and noting that since $F(\xi,\eta)=G(\mathcal{M})$, we have the relation

$$\frac{\partial F}{\partial \xi_{i_1\cdots i_4}} = \sum_{j_1\cdots j_4} \frac{\partial G}{\partial \mathcal{M}_{j_1\cdots j_4}} \eta_{j_1 i_1}\cdots\eta_{j_4 i_4},$$

we find that

$$\frac{\partial G}{\partial \mathcal{M}_{j_1\cdots j_4}} = \langle 0 | e^{\alpha} e^{\beta} S\{(\tfrac{1}{24})\rho_{j_1}\cdots\rho_{j_4} + (\tfrac{1}{4})\delta_{j_1 j_2}\rho_{j_3}\rho_{j_4} + (\tfrac{1}{8})\delta_{j_1 j_2}\delta_{j_3 j_4}\} | 0\rangle, \tag{3}$$

where $\rho_i = \sum_j \eta_{ij} a_j^*$, and S is the symmetrizing operator,

$$S\{f(i_1,\ldots,i_4)\} = \sum_{\sigma} f(i_{\sigma(1)}, i_{\sigma(2)}, i_{\sigma(3)}, i_{\sigma(4)}),$$

where $\sum_\sigma$ is the sum over all permutations on 4-elements. We see therefore that $\partial G/\partial \mathcal{M}_{i_1\cdots i_4}$ is a symmetric tensor. Now let $\mathcal{M}^s_{i_1\cdots i_4} = S\{\mathcal{M}_{i_1\cdots i_4}\}$. Using again the fact that $F(\xi,\eta) = G(\mathcal{M})$, we find that

$$\langle 0 | e^{\alpha} e^{\beta} \rho_i \rho_j | 0\rangle = \sum_k \eta_{ik} \frac{\partial F}{\partial \eta_{jk}} = (\tfrac{1}{6}) \sum_{i_1 i_2 i_3} \mathcal{M}^s_{i_1 i_2 i_3 i} \frac{\partial G}{\partial \mathcal{M}_{i_1 i_2 i_3 j}}$$

and

$$\langle 0 | e^{\alpha} e^{\beta} \rho_i \rho_j \rho_k \rho_l | 0\rangle = \sum_{pq} \eta_{ip}\eta_{kq} \frac{\partial^2 F}{\partial \eta_{jp}\partial \eta_{lq}} = (\tfrac{1}{36}) \sum_{p_1 p_2 p_3} \sum_{q_1 q_2 q_3} \mathcal{M}^s_{p_1 p_2 p_3 i}\, \mathcal{M}^s_{q_1 q_2 q_3 k} \times \frac{\partial^2 G}{\partial \mathcal{M}_{p_1 p_2 p_3 j}\, \mathcal{M}_{q_1 q_2 q_3 l}} + (\tfrac{1}{2}) \sum_{pq} \mathcal{M}^s_{pqik} \frac{\partial G}{\partial \mathcal{M}_{pqjl}}.$$

These two equations may be used instead of those derived in Subsection B to obtain the Green's functions from the vacuum expectation value G. They are more convenient to use when G is written as a function of $\mathcal{M}$ only, though, of course, they are completely equivalent.

Using these relations, we may rewrite Eq. (3) in the final form

$$\frac{\partial G}{\partial \mathcal{M}_{i_1\cdots i_4}} = L_{i_1\cdots i_4} G, \tag{4}$$

where

$$L_{i_1\cdots i_4} = S\{(\tfrac{1}{8})\delta_{i_1 i_2}\delta_{i_3 i_4} + (\tfrac{5}{144})\delta_{i_1 i_2} L_{i_3 i_4} + (\tfrac{1}{864}) L_{i_1 i_2} L_{i_3 i_4}\},$$

and

$$L_{ij} = \sum_{pqr} \mathcal{M}^s_{pqri} \frac{\partial}{\partial \mathcal{M}_{pqrj}}.$$

We see that $G(\mathcal{M})$ which is a single function of N^4 variables satisfies a set of N^4 second-order linear partial differential equations of the generalized hypergeometric type. Furthermore, since by the results of subsection B the first and second derivatives of G are just the four-point and eight-point Green's functions, Eq. (4) is just one integral equation satisfied by the Green's functions of ϕ^4 theory.

It is also easy to see that integral equations in the usual formulation go over into differential equations in our formulation.

E. The power series solution

The Feynman perturbation series for $G(M,D) = G(\mathcal{M})$ is just a power series expansion of G in powers of $\mathcal{M}_{i_1\cdots i_4}$, about the point $\mathcal{M}_{i_1\cdots i_4} = 0$. Instead of the usual Feynman graph form in which this power series is written, we wish to write it in the equivalent standard power series form,

$$G(\mathcal{M}_{i_1\cdots i_4}) = \sum_{n_{1111}=0}^{\infty} \cdots \sum_{n_{NNNN}=0}^{\infty} A(n_{1111},\ldots,n_{NNNN}) \times \prod_{i_1 i_2 i_3 i_4 = 1}^{N} (\mathcal{M}_{i_1 i_2 i_3 i_4})^{n_{i_1 i_2 i_3 i_4}}.$$

To determine the power series coefficients $A(n_{i_1\cdots i_4})$, we start from Eq. (2) with the exponentials expanded in power series. Let

$$\sum_{i_1\cdots i_4=1}^{N} n_{i_1\ldots i_4} = p$$

and let

$$\sum_{i_1\cdots i_4=1}^{N} (\delta_{i_1 j} + \delta_{i_2 j} + \delta_{i_3 j} + \delta_{i_4 j}) n_{i_1\cdots i_4} = p_j.$$

Clearly $A(n_{i_1\cdots i_4})$ is the coefficient of

$$\prod_{i_1\cdots i_4=1}^{N} (\mathcal{M}_{i_1\cdots i_4})^{n_{i_1\cdots i_4}}$$

in the expression

$$\frac{1}{r!\,s!} \langle 0 | \left(\sum_{i_1\cdots i_4} \mathcal{M}_{i_1\cdots i_4} a_{i_1}\cdots a_{i_4} \right)^r \left(\tfrac{1}{2} \sum_i (a_i^*)^2 \right)^s | 0\rangle,$$

where $r = p$ and $s = 2p$. Expanding by the multinomial theorem, we find that

$$A(n_{i_1\cdots i_4}) = \left(\prod_{i_1\cdots i_4=1}^{N} \frac{1}{(n_{i_1\cdots i_4})!} \right) \frac{1}{(2p)!} \times \langle 0 | \prod_{j=1}^{N} a_j^{p_j} (\tfrac{1}{2}\sum_i (a_i^*)^2)^{(2p)} | 0\rangle = \frac{1}{4^p} \left(\prod_{i_1\cdots i_4=1}^{N} \frac{1}{(n_{i_1\cdots i_4})!} \right) \left(\prod_{j=1}^{N} \frac{E(p_j)}{(p_j/2)!} \right) \times \langle 0 | \left(\prod_{j=1}^{N} a_j^{p_j} \right) \left(\prod_{j=1}^{N} (a_j^*)^{p_j} \right) | 0\rangle,$$

where

$$E(p_j) = \begin{cases} 0 & \text{if } p_j \text{ is odd,} \\ 1 & \text{if } p_j \text{ is even,} \end{cases}$$

$$= \frac{1}{4^p} \left(\prod_{i_1\cdots i_4=1}^{N} \frac{1}{(n_{i_1\cdots i_4})!} \right) \prod_{j=1}^{N} E(p_j) \frac{p_j!}{(p_j/2)!}.$$

Using the relation $\Gamma(2z) = 2^{2z-1}\pi^{-1/2}\Gamma(z)\Gamma(z+1/2)$ and the fact that $\sum_{j=1}^{N} p_j = 4p$, we find

$$A(n_{i_1\cdots i_4}) = \frac{4^p}{\pi^{(N/2)}} \left(\prod_{i_1\cdots i_4=1}^{N} \frac{1}{(n_{i_1\cdots i_4})!} \right) \times \prod_{j=1}^{N} E(p_j)\, \Gamma\left(\frac{1+p_j}{2}\right). \tag{5}$$

Let us now examine the convergence of this power series. The series diverges if there exists a ray $u_{i_1\cdots i_4}$, such that if we let $n_{i_1\cdots i_4} = \tau u_{i_1\cdots i_4}$, then $A(n_{i_1\cdots i_4}) \to \infty$ as $\tau \to \infty$. Set $u_{i_1\cdots i_4} = 1$ for all $i_1\cdots i_4$. Then $p = N^4\tau$, $p_j = 4N^3\tau$, and we have, as $\tau \to \infty$,

$$A(\tau) \sim (16N^6\tau/e)^{(N^4\tau)} \to \infty.$$

Therefore, the Feynman perturbation series diverges and is at best an asymptotic series. The interesting question is can we find a well-defined function whose asymptotic series is the Feynman perturbation series. This we shall do in the next subsection.

F. Integral representations

The conventional integral representation may be derived in our formalism as follows. We start from Eq. (2), and using the fact that

$$\exp\left(-\tfrac{1}{2}\sum_i (a_i^\dagger)^2\right) a_j \exp\left(+\tfrac{1}{2}\sum_i (a_i^\dagger)^2\right) = a_j + a_j^\dagger,$$

we find that

$$G(\mathcal{H}) = \langle 0 | \exp\left\{4 \sum_{i_1 \cdots i_4} \mathcal{H}_{i_1 \cdots i_4} q_{i_1} \cdots q_{i_4}\right\} | 0\rangle,$$

where $q_j = (a_j + a_j^\dagger)/\sqrt{2}$. In the orthonormal basis which diagonalizes the Hermitian operators q_j,

$$|0\rangle = \frac{1}{\pi^{N/4}} \prod_{j=1}^{N} \int_{-\infty}^{+\infty} dq_i \exp[-(q_i^2/2)] \, |q_1, \ldots, q_N\rangle,$$

so we find

$$G(\mathcal{H}) = \frac{1}{\pi^{N/2}} \int_{-\infty}^{+\infty} dq_i \cdots \int_{-\infty}^{+\infty} dq_N \times \exp\left\{-\sum_i q_i^2 + 4 \sum_{i_1 \cdots i_4} \mathcal{H}_{i_1 \cdots i_4} q_{i_1} \cdots q_{i_4}\right\}. \tag{6}$$

This is just the well-known functional integral for the Green's functions,[3] and we wish to point out that it is almost useless because over most of the domain of G, i.e., for most values of the arbitrary 4-tensor $\mathcal{H}_{i_1 \cdots i_4}$, the integrand diverges exponentially as the q_i go to infinity.

We shall now show that $G(\mathcal{H})$ is well defined everywhere on its domain by the integral representation

$$G(\mathcal{H}) = \frac{\Gamma(N/4)}{2^{[(N/2)+4]} \pi^{[(N+1)/2)]}} \int_{-\infty}^{+\infty} dp \exp[-(p/4)^2] \int d\Omega \times \left(\frac{p^2}{1 - p^2 \sum_{i_1 \cdots i_4} u_{i_1} u_{i_2} u_{i_3} u_{i_4} \mathcal{H}_{i_1 i_2 i_3 i_4}}\right)^{(N/4)}, \tag{7}$$

where $\int d\Omega$ is the integral over the surface of the unit sphere in N-dimensional Euclidean space, $\sum_{i=1}^{N} u_i^2 = 1$, and u_i is the corresponding real unit vector. This multiple integral is equal to $G(\mathcal{H})$ in the sense that it has an asymptotic expansion about the point $\mathcal{H}_{i_1 \cdots i_4} = 0$, which is just the power series we found in subsection E. In order to completely define $G(\mathcal{H})$, it is necessary to prescribe how the contour of integration is to be deformed in the neighborhood of points such that

$$p^2 \sum_{i_1 \cdots i_4} u_{i_1} \cdots u_{i_4} \mathcal{H}_{i_1 \cdots i_4} = 1,$$

so that the corresponding singularities of the integrand may be avoided. We leave this unspecified as all choices lead to the same asymptotic power series (i.e., to the same Feynman perturbation theory).

To investigate the behavior in the vicinity of the point $\mathcal{H}_{i_1 \cdots i_4} = 0$, we first interchange the order of integration and perform the p integral. Using the fact[4] that

$$\int_{-\infty}^{+\infty} dp \exp\left[-\left(\frac{p}{4}\right)^2 \left(\frac{p^2}{1 - p^2 z}\right)^{(N/4)}\right] = \frac{\Gamma(N/4 + 1/2)}{(-z)^{(N/4+1/2)}} \Psi\left(\left(\frac{N}{4} + \frac{1}{2}\right), \left(\frac{3}{2}\right); \left(\frac{-1}{16z}\right)\right),$$

where $\Psi(a, b; x)$ is the confluent hypergeometric function which goes as x^{-a} as $x \to \infty$, and both $(-z)^{(N/4+1/2)}$ and $\Psi((N/4 + 1/2), (3/2); (-1/16z))$ are taken to be always on their principal sheet in the complex z plane cut along the positive real axis, we find that

$$G(\mathcal{H}) = \int \frac{d\Omega}{\Omega} \left(\frac{-1}{16z}\right)^{(N/4+1/2)} \cdot \Psi\left[\left(\frac{N}{4} + \frac{1}{2}\right), \left(\frac{3}{2}\right); \left(\frac{-1}{16z}\right)\right], \tag{8}$$

where $\Omega = 2\pi^{(N/2)} / \Gamma(N/2) =$ the area of the unit sphere and where

$$z = \sum_{i_1 \cdots i_4} u_{i_1} \cdots u_{i_4} \mathcal{H}_{i_1 \cdots i_4}.$$

When $z \neq 0$, the integrand may be evaluated via the convergent power series,

$$\left(\frac{-1}{16z}\right)^{(N/4+1/2)} \Psi\left[\left(\frac{N}{4} + \frac{1}{2}, \left(\frac{3}{2}\right); \left(\frac{-1}{16z}\right)\right] = \frac{2^{(N/2-1)}\sqrt{\pi}}{\Gamma(N/2)} \times \left(\sum_{n=0}^{\infty} \frac{\Gamma(N/4 + n)}{\Gamma(1/2 + n)} \frac{(-1/16z)^{(N/4+n)}}{n!} - \sum_{n=0}^{\infty} \frac{\Gamma(N/4 + 1/2 + n)}{\Gamma(3/2 + n)} \frac{(-1/16z)^{(N/4+1/2+n)}}{n!}\right).$$

When $z = 0$, the integrand has an essential singularity. However, for all z on the principal sheet, the integrand has the well behaved asymptotic expansion,[5]

$$\left(\frac{-1}{16z}\right)^{(N/4+1/2)} \Psi\left[\left(\frac{N}{4} + \frac{1}{2}\right), \left(\frac{3}{2}\right); \left(\frac{-1}{16z}\right)\right] = \left(\frac{1}{\Gamma(N/2)}\right) \sum_{n=0}^{Q} \frac{\Gamma(N/2 + 2n)}{n!} (4z)^n + O(|z|^{(Q+1)}) \tag{9}$$

as $|z| \to 0$, and therefore the integral is well defined everywhere. Now let λ equal the maximum value of $|z|$ for all u_i on the unit sphere. Then for $\mathcal{H}_{i_1 \cdots i_4} \to 0$, we find that

$$G(\mathcal{H}) = \frac{1}{2\pi^{(N/2)}} \int d\Omega \left(\sum_{n=0}^{Q} \frac{\Gamma(N/2 + 2n)}{n!} (4z)^n + O(\lambda^{Q+1})\right) = \frac{1}{\pi^{(N/2)}} \sum_{n=0}^{Q} \frac{\Gamma(N/2 + 2n)}{n!} 4^n \int_{-\infty}^{+\infty} du_1 \cdots \times \int_{-\infty}^{+\infty} du_N \, \delta\left(1 - \sum_{i=1}^{N} u_i^2\right) z^n + O(\lambda^{Q+1}).$$

This is just an asymptotic power series in $\mathcal{H}_{i_1 \cdots i_4}$ whose coefficients are given by

$$A(n_{i_1 \cdots i_4}) = \frac{\Gamma(N/2 + 2p) 4^p}{\pi^{N/2}} \left(\prod_{i_1 \cdots i_4 = 1}^{N} \frac{1}{(n_{i_1 \cdots i_4})!}\right) \times \int_{-\infty}^{+\infty} du_1 \cdots \int_{-\infty}^{+\infty} du_N \delta\left(1 - \sum_i u_i^2\right) \prod_{j=1}^{N} u_j^{p_j},$$

where p and p_j are defined as in subsection E. By making the substitution $u_i \to -u_i$ for each i separately, we see that

$$A(n_{i_1\cdots i_4}) = \frac{\Gamma(N/2+2p)2^{(N+2p)}}{\pi^{(N/2)}} \left(\prod_{i_1\cdots i_4} \frac{1}{(n_{i_1\cdots i_4})!}\right) \times \int_0^\infty du_1 \cdots \int_0^\infty du_N \; \delta\left(1-\sum_i u_i^2\right) \prod_{j=1}^N E(p_j) u_j^{pj},$$

where $E(pj)$ is defined as in subsection E. Let $x_i = u_i^2$ and, using

$$\int_0^\infty dx_1 \cdots \int_0^\infty dx_n \, \delta\left(1-\sum_i x_i\right) \prod_{j=1}^N x_j^{(\alpha_j-1)} = \frac{\prod_{j=1}^N \Gamma(\alpha_j)}{\Gamma(\sum_{i=1}^N \alpha_j)},$$

we find

$$A(n_{i_1\cdots i_4}) = \frac{4^p}{\pi^{N/2}} \left(\prod_{i_1\cdots i_4=1}^N \frac{1}{(n_{i_1\cdots i_4})!}\right) \prod_{j=1}^N E(p_j)\, \Gamma\left(\frac{(1+p_j)}{2}\right),$$

which is just Eq. (5) of subsection E. Thus we have succeeded in constructing an explicit function $G(\mathcal{H})$, well defined everywhere on the domain of arbitrary 4-tensors $\mathcal{H}_{i_1\cdots i_4}$, whose asymptotic expansion about $\mathcal{H}_{i_1\cdots i_4}=0$ is just the Feynman perturbation series for the boson case.

We wish to study the analytic structure of $G(\mathcal{H})$ as a function of $\mathcal{H}$. Returning to Eq. (8), we see that the only singularities of the integrand in the region of integration occur at points u_i, such that

$$z = \sum_{i_1\cdots i_4} u_{i_1}\cdots u_{i_4} \mathcal{H}_{i_1\cdots i_4} = 0,$$

at which points the integrand has an essential singularity. By the well-known analysis of Eden, *et al.*[6] $G(\mathcal{H})$ will have a singularity at a point $\mathcal{H}$, only if there exists a unit vector u_i such that

$$z = \sum_{i_1\cdots i_4} u_{i_1}\cdots u_{i_4} \mathcal{H}_{i_1\cdots i_4} = 0$$

and

$$\frac{\partial z}{\partial u_j} = \left(\frac{1}{6}\right) \sum_{i_1 i_2 i_3} u_{i_1} u_{i_2} u_{i_3} \mathcal{H}^s_{i_1 i_2 i_3 j} = 0,$$

for all j. $\mathcal{H}^s$ is defined as in subsection D.

Since the "pinch" occurs at an essential singularity of the integrand, $G(\mathcal{H})$ has an essential singularity at such points. Thus $G(\mathcal{H})$ is an analytic function of $\mathcal{H}_{i_1\cdots i_4}$ except at points where the N equations

$$\sum_{i_1 i_2 i_3=1}^N u_{i_1} u_{i_2} u_{i_3} \mathcal{H}^s_{i_1 i_2 i_3 j} = 0, \tag{10}$$

$j=1,2,\ldots,N$, and u_i real, possess a solution. At such points $G(\mathcal{H})$ has an essential singularity. Finally we note that in the actual local $g\,\phi^4(x)$ field theory case,

$$\mathcal{H}^s(x_1,x_2,x_3,x_4) = ig \int d^4x \prod_{i=1}^4 K(x_i - x),$$

where

$$K(x) = \sqrt{i} \int \frac{d^4p}{(2\pi)^4} \frac{\exp(-ipx)}{(p^2-m^2+i\epsilon)^{1/2}}.$$

So, in this case Eq. (10) becomes

$$ig \int d^4x\, [v(x)]^3 K(y-x) = 0 \quad \text{for all } y, \tag{11}$$

where

$$v(x) = \int d^4y\, u(y) K(y-x),$$

$u(y)$ is a real function of y, and

$$\int d^4y\, [u(y)]^2 = 1.$$

Applying the operator $\sqrt{-i}\,(-\Box_y^2 - m^2)^{1/2}$ to both sides of Eq. (11), we obtain

$$ig[v(y)]^3 = 0 \quad \text{for all } y,$$

which implies (if $g \neq 0$)

$$v(y) = \int d^4x\, K(x-y)\, u(x) = 0 \quad \text{for all } y.$$

Applying $\sqrt{-i}\,(-\Box_y^2 - m^2)^{1/2}$ to both sides of this equation, we find

$$u(x) = 0 \quad \text{for all } x,$$

which contradicts the requirement

$$\int d^4x\, [u(x)]^2 = 1.$$

Therefore, when $g \neq 0$, local ϕ^4 field theory corresponds to a value of $\mathcal{H}_{i_1\cdots i_4}$, where $G(\mathcal{H})$ is analytic. As $g \to 0$, $\mathcal{H}$ moves onto an essential singularity of $G(\mathcal{H})$, for when $g=0$ Eq. (11) is trivially satisfied.

II. THE FERMION CASE

A. Formulation of the problem

Consider explicitly the case of quantum electrodynamics. The Green's function for n external electron lines, n external positron lines, and m external photon lines may be succinctly expressed in the form

$$\begin{aligned} G_{n,m}(i_1x_1, \cdots, i_nx_n; j_1y_1, \cdots; j_ny_n; \mu_1z_1, \cdots, \mu_mz_m) &= \langle 0 | b_{i_1}(x_1)\cdots b_{i_n}(x_n)\, c_{j_1}(y_1)\cdots c_{j_n}(y_n) \\ &\times a_{\mu_1}(z_1)\cdots a_{\mu_m}(z_m) \\ &\times \exp\left\{+ ie \sum_{ij\mu} \int d^4x\, c_i(x)(\gamma_\mu)_{ij} b_j(x)\, a_\mu(x)\right\} \\ &\times \exp\left\{\tfrac{1}{2} \sum_{\mu\nu} \int d^4x_1 d^4x_2\, a^+_\mu(x_1) D^F_{\mu\nu}(x_1-x_2) a^+_\nu(x_2)\right\} \\ &\times \exp\left\{- \sum_{ij} \int d^4x d^4y\, b^+_i(x) S^F_{ij}(x-y)\, c^+_j(y)\right\} | 0\rangle, \end{aligned} \tag{12}$$

where

$$[a_\mu(x_1),\ a_\nu(x_2)] = 0,$$

$$[a_\mu(x_1), a^+_\nu(x_2)] = \delta^4(x_1-x_2)\delta_{\mu\nu},$$

any a or $a^\dagger$ commutes with any b, b^+, c, or c^+,

$$[b_i(x_1), b_j(x_2)]_+ = [c_i(x_1), c_j(x_2)]_+ = [b_i(x_1), c_j(x_2)]_+ = [b_i(x_1),\ c^+_j(x_2)]_+ = 0,$$

$$[b_i(x_1), b^+_j(x_2)]_+ = [c_i(x_1),\ c^+_j(x_2)]_+ = \delta_{ij}\ \delta^4(x_1-x_2).$$

The vacuum state $|0\rangle$ is defined by

$$a_\mu(x)\,|\,0\rangle = b_i(x)\,|0\rangle = c_i(x)\,|0\rangle = 0 \quad \text{for all } i,\ \mu, \text{ and } x.$$

S^F is the electron propogator

$$S^F_{ij}(x-y)=(i\partial\!\!\!/_x-m)_{ij}\,\Delta^F(x-y).$$

Equation (12) is derived in Appendix C. In keeping with our general program we replace the photon indices (μ, z) by a discrete finite index $\alpha=1,2,\ldots,M$, and the electron indices (i,x) by a discrete finite index $i=1,2,\ldots,N$. We then replace

$$ie\sum_{ij\mu}\int d^4x\, c_i(x)(\gamma_\mu)_{ij}\,b_j(x)\,a_\mu(x)$$

by

$$\sum_{ij\alpha} c_i\Gamma^j_{i\alpha}b_j a_\alpha=\gamma$$

and

$$\tfrac{1}{2}\sum_{\mu\nu}\int d^4x_1 d^4x_2\, a^+_\mu(x_1)\,D^F_{\mu\nu}(x_1-x_2)a^+_\nu(x_2)$$

by

$$\tfrac{1}{2}\sum_{\alpha\beta} a^+_\alpha\, D_{\alpha\beta}\,a^+_\beta=\delta$$

and

$$\sum_{ij}\int d^4x\, d^4y\;\; b^+_i(x)S^F_{ij}(x-y)\,c^+_j(y)$$

by

$$\sum_{ij} b^+_i S_{ij}\,c^+_j=\sigma,$$

where Γ is an arbitrary 3-tensor and S and D are arbitrary matrices. Thus we really consider any fermion theory with an intermediate boson interaction. The commutation relations are the same except that we replace

$$\delta_{\mu\nu}\,\delta^4(x_1-x_2)\quad\text{by}\quad\delta_{\alpha\beta},$$

and

$$\delta_{ij}\,\delta^4(x_1-x_2)\text{ by }\delta_{ij}.$$

Then the object we wish to study is

$$G_{n,m}(i_1,\cdots,i_n;\,j_1,\cdots,j_n;\,\alpha_1,\cdots\alpha_m)$$
$$=\langle 0\,|\,b_{i_1}\cdots b_{i_n}c_{j_n}a_{\alpha_1}\cdots a_{\alpha_m}\,e^{\gamma}e^{\delta}e^{-\sigma}\,|\,0\rangle,\tag{13}$$

which is a function mapping an arbitrary 3-tensor and two arbitrary matrices into a $(2n+m)$-tensor which is antisymmetric in the first n and in the second n indices, and is symmetric in the last m indices.

B. Only the vacuum graphs need be considered

All Green's functions may be written as partial derivatives of the vacuum Green's function with respect to Γ, S, and D. The proof proceeds in the same manner as the proof in Sec. IB, only there are more cases to consider. We state the three particular results we shall use later:

$$G_{1,1}(i;j;\alpha)=\frac{-\partial G}{\partial\Gamma^i_{j\alpha}},\tag{14a}$$

$$G_{0,2}(\alpha,\beta)=\mathcal{D}_{\alpha\beta}G+2\sum_{\gamma\delta}\mathcal{D}_{\alpha\gamma}\,\mathcal{D}_{\beta\delta}\,\frac{\partial G}{\partial D_{\gamma\delta}},\tag{14b}$$

$$G_{1,0}(i;j)=S_{ij}\,G+\sum_{kl}S_{ik}\,\frac{\partial G}{\partial S_{lk}}\,S_{lj},\tag{14c}$$

where as in Sec. I, $\mathcal{D}_{\alpha\beta}=\frac{1}{2}(D_{\alpha\beta}+D_{\beta\alpha})$, and we denote $G_{0,0}$ simply by G.

C. Reduction in the number of variables

We start from Eq. (13),

$$G(\Gamma,S,D)=\langle 0\,|\,e^{\gamma}e^{\delta}e^{-\sigma}\,|\,0\rangle.$$

Inserting factors of $e^{+\delta}e^{-\delta}$ and using $\langle 0\,|\,a^+=0$, we find

$$G=\langle 0\,|\,e^{\tilde\gamma}e^{-\sigma}\,|\,0\rangle,$$

where

$$\tilde\gamma=e^{-\delta}\gamma e^{+\delta}=\gamma+[\gamma,\delta]+\frac{1}{2!}[[\gamma,\delta],\delta]+\cdots$$
$$=\gamma+2\sum_{ij\alpha\beta}c_i\Gamma^j_{i\alpha}b_j\mathcal{D}_{\alpha\beta}a^+_\beta.$$

Then using the relation

$$\exp(A+B)=\exp(A)\exp\{-[A,B]/2\}\exp(B),$$

where

$$[[A,B],A]=[[A,B],B]=0,$$
$$A=2\sum_{ij\alpha\beta}c_i\Gamma^j_{i\alpha}b_j\,\mathcal{D}_{\alpha\beta}\,a^+_\beta,\quad\text{and } B=\gamma,$$

we find

$$G=\langle 0\,|\,e^{\rho}e^{-\sigma}\,|\,0\rangle,\tag{15}$$

where

$$\rho=\sum_{i_1i_2j_1j_2}c_{i_1}c_{i_2}\rho^{j_1j_2}_{i_1i_2}b_{j_1}b_{j_2},$$

and

$$\rho^{j_1j_2}_{i_1i_2}=\sum_{\alpha\beta}\Gamma^{j_1}_{i_1\alpha}\mathcal{D}_{\alpha\beta}\,\Gamma^{j_2}_{i_2\beta}.$$

Equation (15) for arbitrary 4-tensors $\rho^{j_1j_2}_{i_1i_2}$ is just the expression we would have obtained at the outset if we had started with a pure Fermion theory with current–current interaction. Therefore, our considerations apply to this case also.

Now assume that $\det(S_{ij})\neq 0$, and let

$$b_i\to -S_{ij}b_j,\quad b^+_i\to -b^+_j(S^{-1})_{ji},$$
$$c_i\to c_i,\quad c^+_i\to c^+_i.$$

The commutation relations are invariant under this substitution, so that G is invariant under this substitution, and therefore

$$G(\Gamma,S,D)=G(\mathcal{M})=\langle 0\,|\exp\Big(\sum_{i_1i_2j_1j_2}c_{i_1}c_{i_2}\,\mathcal{M}^{j_1j_2}_{i_1i_2}\,b_{j_1}b_{j_2}\Big)$$
$$\times\exp\Big(\sum_i b^+_ic^+_i\Big)\Big|\,0\rangle,\tag{16}$$

where

$$\mathcal{M}^{j_1j_2}_{i_1i_2}=\sum_{k_1k_2}\rho^{k_1k_2}_{i_1i_2}\,S_{k_1j_1}S_{k_2j_2}$$
$$=\sum_{\alpha\beta k_1k_2}\Gamma^{k_1}_{i_1\alpha}\,\mathcal{D}_{\alpha\beta}\,\Gamma^{k_2}_{i_2\beta}S_{k_1j_1}S_{k_2j_2}.$$

Since G is a function only of $\mathcal{M}$, there exist two relations among the three partial derivatives of G with respect to Γ, S, and D. These are

$$\sum_{kl}S_{ik}\frac{\partial G}{\partial S_{lk}}S_{lj}=+\sum_{kl\alpha}\frac{\partial G}{\partial\Gamma^i_{l\alpha}}\Gamma^k_{l\alpha}S_{kj},$$

$$2\sum_{\gamma\delta}\mathcal{D}_{\alpha\gamma}\mathcal{D}_{\beta\delta}\frac{\partial G}{\partial D_{\gamma\delta}}=\sum_{ij\gamma}\frac{\partial G}{\partial\Gamma^i_{j\alpha}}\Gamma^i_{j\gamma}\mathcal{D}_{\gamma\beta}.$$

Using the results of subsection B, we find

$$\frac{G_{0,2}(\alpha,\beta)}{G} = D_{\alpha\beta} - \sum_{ij\gamma} \frac{G_{1,1}(i;j;\alpha)}{G} \Gamma^{i}_{j\gamma} D_{\gamma\beta},$$

and

$$\frac{G_{1,0}(i;j)}{G} = S_{ij} - \sum_{kl\alpha} \frac{G_{1,1}(i;l;\alpha)}{G} \Gamma^{k}_{l\alpha} S_{kj}.$$

These are just the first two Dyson equations[7] which we see are a simple consequence of the fact that $G(\Gamma, S, D) = G(\mathcal{M})$ is really only a function of $\mathcal{M}$.

D. The differential equation

We start from Eq. (15),

$$G(\rho, S) = \langle 0 | e^{\rho} e^{-\sigma} | 0 \rangle,$$

where

$$\rho = \sum_{i_1 i_2 j_1 j_2} c_{i_1} c_{i_2} \rho^{j_1 j_2}_{i_1 i_2} b_{j_1} b_{j_2},$$

and $\sigma = \sum_{ij} b^{\dagger}_i S_{ij} c^{\dagger}_j$. Differentiating, we find

$$\begin{aligned} \partial\rho^{j_1 j_2}_{i_1 i_2} &= \langle 0 | e^{\rho} c_{i_1} c_{i_2} b_{j_1} b_{j_2} e^{-\sigma} | 0 \rangle \\ &= \langle 0 | e^{\rho} e^{-\sigma} \tilde{c}_{i_1} \tilde{c}_{i_2} \tilde{b}_{j_1} \tilde{b}_{j_2} | 0 \rangle, \end{aligned} \tag{17}$$

where

$$\tilde{c}_i = e^{\sigma} c_i e^{-\sigma} = c_i + \sum_j b^{\dagger}_j S_{ji},$$

$$\tilde{b}_i = e^{\sigma} b_i e^{-\sigma} = b_i - \sum_j S_{ij} c^{\dagger}_j.$$

Note that $G(\rho, S) = G(\mathcal{M})$, where

$$\mathcal{M}^{j_1 j_2}_{i_1 i_2} = \sum_{k_1 k_2} \rho^{k_1 k_2}_{i_1 i_2} S_{k_1 j_1} S_{k_2 j_2},$$

and therefore

$$\frac{\partial G}{\partial \rho^{j_1 j_2}_{i_1 i_2}} = \sum_{k_1 k_2} \frac{\partial G}{\partial \mathcal{M}^{k_1 k_2}_{i_1 i_2}} S_{j_1 k_1} S_{j_2 k_2}.$$

Then, commuting the c's and b's through to the right in Eq. (17), multiplying both sides by $(S^{-1})_{l_1 j_1} (S^{-1})_{l_2 j_2}$, and summing over j_1 and j_2, we find

$$\begin{aligned} -\frac{\partial G}{\partial \mathcal{M}^{l_1 l_2}_{i_1 i_2}} &= \langle 0 | e^{\rho} e^{-\sigma} A\{(\tfrac{1}{2}) \delta_{i_1 l_1} \delta_{i_2 l_2} \\ &\quad + \delta_{i_1 l_1} \zeta_{i_2} c^{\dagger}_{l_2} + (\tfrac{1}{4}) \zeta_{i_1} c^{\dagger}_{l_1} \zeta_{i_2} c^{\dagger}_{l_2}\} | 0 \rangle, \end{aligned}$$

where $\zeta_i = \sum_j b^{\dagger}_j S_{ji}$, and A antisymmetrizes (i_1, i_2) and (l_1, l_2),

$$\begin{aligned} A\{f(i_1, i_2, l_1, l_2)\} &= f(i_1, i_2, l_1, l_2) + f(i_2, i_1, l_2, l_1) \\ &\quad - f(i_1, i_2, l_2, l_1) - f(i_2, i_1, l_1, l_2). \end{aligned}$$

We see that $\partial G / \mathcal{M}^{l_1 l_2}_{i_1 i_2}$ is antisymmetric in (i_1, i_2) and (l_1, l_2). Let ${}^{A}\mathcal{M}^{l_1 l_2}_{i_1 i_2} = A\{\mathcal{M}^{l_1 l_2}_{i_1 i_2}\}$. Using the fact that $G(\rho, S) = G(\mathcal{M})$, we find

$$\begin{aligned} \langle 0 | e^{\rho} e^{-\sigma} \xi_i c^{\dagger}_l | 0 \rangle &= -\sum_j \frac{\partial G}{\partial S_{jl}} S_{ji} \\ &= \tfrac{1}{2} \sum_{i_1 i_2 j} \frac{\partial G}{\partial \mathcal{M}^{l j}_{i_1 i_2}} {}^{A}\mathcal{M}^{ij}_{i_1 i_2} \end{aligned}$$

and

$$\begin{aligned} \langle 0 | e^{\rho} e^{-\sigma} \zeta_{i_1} c^{\dagger}_{l_1} \zeta_{i_2} c^{\dagger}_{l_2} | 0 \rangle &= \sum_{j_1 j_2} \frac{\partial^2 G}{\partial S_{j_1 i_1} \partial S_{j_2 i_2}} S_{j_1 i_1} S_{j_2 i_2} \\ &= \tfrac{1}{4} \sum_{\substack{j_1 k_1 k_2 \\ j_2 k_3 k_4}} \frac{\partial^2 G}{\partial \mathcal{M}^{l_1 j_1}_{k_1 k_2} \partial \mathcal{M}^{l_2 j_2}_{k_3 k_4}} \\ &\quad \times {}^{A}\mathcal{M}^{i_1 j_1}_{k_1 k_2} {}^{A}\mathcal{M}^{i_2 j_2}_{k_3 k_4} \\ &\quad + \tfrac{1}{2} \sum_{k_1 k_2} \frac{\partial G}{\partial \mathcal{M}^{l_1 l_2}_{k_1 k_2}} {}^{A}\mathcal{M}^{i_1 i_2}_{k_1 k_2}. \end{aligned}$$

Using these relations, we find that the differential equation for $G(\mathcal{M})$ is

$$-\frac{\partial G}{\partial \mathcal{M}^{j_1 j_2}_{i_1 i_2}} = L^{j_1 j_2}_{i_1 i_2} G, \tag{18}$$

where

$$L^{j_1 j_2}_{i_1 i_2} = A\{(\tfrac{1}{2}) \delta_{i_1 j_1} \delta_{i_2 j_2} + (\tfrac{3}{8}) \delta_{i_1 j_1} L^{j_2}_{i_2} + (\tfrac{1}{16}) L^{j_1}_{i_1} L^{j_2}_{i_2}\},$$

and

$$L^{j}_{i} = \sum_{klm} {}^{A}\mathcal{M}^{jk}_{lm} \frac{\partial}{\partial \mathcal{M}^{ik}_{lm}}.$$

The remarks at the end of Sec. ID apply here also.

E. The power series solution

We start from Eq. (16),

$$\begin{aligned} G(\mathcal{M}) = \langle 0 | \exp\Big(&\sum_{i_1 i_2 j_1 j_2} c_{i_1} c_{i_2} \mathcal{M}^{j_1 j_2}_{i_1 i_2} b_{j_1} b_{j_2} \Big) \\ &\times \exp\Big(\sum_i b^{\dagger}_i c^{\dagger}_i\Big) | 0 \rangle. \end{aligned}$$

Let $[N/2]$ be the greatest integer less than or equal to $(N/2)$. Then expanding the exponentials as power series we find,

$$G(\mathcal{M}) = \sum_{n=0}^{[N/2]} \frac{1}{n!(2n)!} \langle 0 | (\sum c_{i_1} c_{i_2} \mathcal{M}^{j_1 j_2}_{i_1 i_2} b_{j_1} b_{j_2})^n (\sum b^{\dagger}_i c^{\dagger}_i)^{2n} | 0 \rangle,$$

since the product of any n b_i's or n c_i's for $n > N$ is zero. Commuting the b_i's and c_i's through to the right, we find

$$\begin{aligned} G(\mathcal{M}) = \sum_{n=0}^{[N/2]} &\frac{(-1)^n}{n!(N-2n)!} \epsilon^{i_1 \cdots i_N} \epsilon_{j_1 \cdots j_N} \\ &\times \mathcal{M}^{j_1 j_2}_{i_1 i_2} \cdots \mathcal{M}^{j_{(2n-1)} j_{(2n)}}_{i_{(2n-1)} i_{(2n)}} \delta^{j_{(2n+1)}}_{i_{(2n+1)}} \cdots \delta^{j_N}_{i_N}, \end{aligned} \tag{19}$$

where $\epsilon^{i_1 \cdots i_N}$ is the fully antisymmetric tensor in N dimensions and repeated indices are summed over. Therefore, $G(\mathcal{M})$ is just a finite polynomial in $\mathcal{M}$ and is therefore convergent everywhere on its domain. Equation (19) seems to be a generalization of the determinant to the case of 4-tensors. To see the analogy more clearly, we note that if M^j_i is an $N \times N$ matrix, then

$$\begin{aligned} \det(I - M) = \sum_{n=0}^{N} &\frac{(-1)^n}{n!(N-n)!} \epsilon^{i_1 \cdots i_N} \epsilon_{j_1 \cdots j_N} \\ &\times M^{j_1}_{i_1} \cdots M^{j_n}_{i_n} \delta^{j_{(n+1)}}_{i_{(n+1)}} \cdots \delta^{j_N}_{i_N}. \end{aligned} \tag{20}$$

Also we might ask if expressions of the type appearing in Eq. (19) appear anywhere else in mathematics. The answer is yes! If we let $R^{j_1 j_2}_{i_1 i_2}$ be the Riemann curvature

tensor of a N-dimensional surface (where N is even), and let

$$I=\epsilon^{i_1\cdots i_N}\,\epsilon_{j_1\cdots j_N}\,R^{j_1 j_2}_{i_1 i_2}\cdots R^{j_{(N-1)} j_N}_{i_{(N-1)} i_N},$$

then I is just that invariant, which when integrated over the entire surface gives the Euler—Poincaré topological characteristic of the surface.[8]

Next we would like to point out that one usually calculates the physical Green's functions by first dividing out all the vacuum graphs and then summing the remaining graphs. If we let $G^P_{n,m}$ be the physical Green's function, then $G^P_{n,m}=G_{n,m}\times G^{-1}$, and the usual procedure corresponds to expanding G^{-1} in an infinite power series in $\mathcal{M}$ multiplying this by $G_{n,m}$ which is a finite polynomial in $\mathcal{M}$, and obtaining an infinite power series in $\mathcal{M}$ for $G^P_{n,m}$, which converges only if the power series of G^{-1} converges. Thus we see that, by the usual procedure of dividing out the vacuum graphs before summing the series, we may destroy the convergence of $G^P_{n,m}$ even though $G_{n,m}$ and G are finite polynomials in M. Finally we would like to point out that since $G(\mathcal{M})$ is a generalization of the determinant, we might expect that there would be a corresponding generalization of the Fredholm theory of linear integral equations for the infinite dimensional case. This theory has already been worked out by Caianiello[9] whose general formalism is similar to our Eq. (19). Unfortunately it applies only to sufficiently smooth and rapidly decreasing propagators and couplings, and therefore does not apply to the physical case.

F. Integral representations

Let $R^j_{i\alpha}=i\sum_{\beta k}\Gamma^k_{i\beta}S_{kj}(\sqrt{\Box})_{\beta\alpha}$, which implies that $\mathcal{M}^{j_1 j_2}_{i_1 i_2}=-\sum_\alpha R^{j_1}_{i_1\alpha}R^{j_2}_{i_2\alpha}$. Let $P^j_i=\sum_{\alpha=1}^M R^j_{i\alpha}u_\alpha$, where the u_α are M auxiliary variables. Then

$$G(\mathcal{M})=\frac{1}{(4\pi)^{(M/2)}}\int d^M u\exp(-|u|^2/4)\det(I+P),\quad(21)$$

where $\int d^M u=\int_{-\infty}^{+\infty}du_1\cdots\int_{-\infty}^{+\infty}du_M$, and $|u|^2=\sum_{\alpha=1}^M u_\alpha^2$. To prove this, we expand the determinant by Eq. (20), obtaining

$$\begin{aligned}G(\mathcal{M})&=\frac{1}{(4\pi)^{(M/2)}}\int d^M u\exp(-|u|^2/4)\\&\times\sum_{r=0}^{N}\frac{1}{r!(N-r)!}\epsilon^{i_1\cdots i_N}\epsilon_{j_1\cdots j_N}\\&\times P^{j_1}_{i_1}\cdots P^{j_r}_{i_r}\delta^{j_{(r+1)}}_{i_{(r+1)}}\cdots\delta^{j_N}_{i_N}\\&=\frac{1}{(4\pi)^{(M/2)}}\sum_{r=0}^{N}\frac{1}{r!(N-r)!}\epsilon^{i_1\cdots i_N}\epsilon_{j_1\cdots j_N}\\&\times\sum_{\alpha_1,\cdots,\alpha_r=1}^{M}R^{j_1}_{i_1\alpha_1}\cdots R^{j_r}_{i_r\alpha_r}\delta^{j_{(r+1)}}_{i_{(r+1)}}\cdots\delta^{j_N}_{i_N}\\&\times\int d^M u\exp(-|u|^2/4)\prod_{\beta=1}^{M}(u_\beta)^{r_\beta},\end{aligned}$$

where $r_\beta(\alpha_1,\ldots,\alpha_r)=\sum_{i=1}^r\delta_{\beta\alpha_i}$, and $\sum_{\beta=1}^M r_\beta=r$. Let $n_\beta=r_\beta/2$ and $\sum_{\beta=1}^m n_\beta=r/2=n$. Using

$$\int_{-\infty}^{+\infty}du_\beta\exp(-u_\beta^2/4)(u_\beta)^{r_\beta}=E(r_\beta)2^{(r_\beta+1)}\Gamma((r_\beta+1)/2),$$

$$\Gamma(n_\beta+\tfrac{1}{2})=\Gamma(r_\beta+1)\sqrt{\pi}/\Gamma(n_\beta+1)2^{r_\beta}$$

and rearranging terms, we find

$$\begin{aligned}G(\mathcal{M})&=\sum_{n=0}^{[N/2]}\frac{1}{n!(N-2n)}\epsilon^{i_1\cdots i_N}\epsilon_{j_1\cdots j_N}\\&\times\sum_{\alpha_1,\ldots,\alpha_n=1}^{M}R^{j_1}_{i_1\alpha_1}R^{j_2}_{i_2\alpha_1}\cdots R^{j_{(2n-1)}}_{i_{(2n-1)}\alpha_n}R^{j_{(2n)}}_{i_{(2n)}\alpha_n}\\&\times\delta^{j_{(2n+1)}}_{i_{(2n+1)}}\cdots\delta^{j_N}_{i_N}\\&=\sum_{n=0}^{[N/2]}\frac{(-1)^n}{n!(N-2n)!}\epsilon^{i_1\cdots i_N}\epsilon_{j_1\cdots j_N}\\&\times\mathcal{M}^{j_1 j_2}_{i_1 i_2}\cdots\mathcal{M}^{j_{(2n-1)}j_{(2n)}}_{i_{(2n-1)}i_{(2n)}}\delta^{j_{(2n+1)}}_{i_{(2n+1)}}\cdots\delta^{j_N}_{i_N},\end{aligned}$$

which is just Eq. (19). QED

This is just the well-known integral representation for the fermion case.[10]

CONCLUSION

We have completely solved the problem of determining the exact Green's functions in our finite-dimensional version of relativistic perturbation theory. The exact Green's functions are given by simple closed expressions which are finite polynomials in the fermion case and integrals over the unit sphere in the boson case.

The new integral representation [Eq. (8)] for the Green's functions of boson field theories seems very promising. It is an absolutely convergent integral over a compact space, with an integrand which is bounded everywhere in the region of integration. It explicitly exhibits the essential singularity at zero coupling constant. All this should make it far more useful for studies of the limit $N\to\infty$ than the usual integral representation [Eq. (6)] has been.

ACKNOWLEDGMENTS

I wish to thank O.W. Greenberg, D. Fivel, P.K. Mitter, P. Vinciarelli, and P. Weiss for helpful discussions.

APPENDIX A

We start from the standard expression[11]

$$\begin{aligned}G^g_n(x_1,\ldots,x_n)&=\langle\Phi\,|\,T(\phi(x_1)\cdots\phi(x_n)\\&\times\exp(ig\int d^4x[\phi(x)]^4))\,|\,\Phi\rangle,\end{aligned}$$

where $\phi(x)$ is the free field operator and $|\Phi\rangle$ is the free field vacuum state. We note that for $g=0$ we have for even n, $n=2m$ (for odd n, $G_n=0$),

$$\begin{aligned}G^0_n(x_1\ldots x_n)&=\langle\Phi\,|\,T(\phi(x_1)\ldots\phi(x_n))\,|\,\Phi\rangle\\&=\frac{1}{2^m m!}\sum_\sigma\Delta_F(x_{\sigma(1)}-x_{\sigma(2)})\cdots\\&\times\Delta_F(x_{\sigma(2m-1)}-x_{\sigma(2m)}),\end{aligned}$$

where $\sum_\sigma$ is the sum over all permutations $\sigma(i)$ of $i=1,2,\ldots,2m$,

$$\begin{aligned}&=\langle 0\,|\,a(x_1)\cdots a(x_n)\\&\times\frac{[\int d^4x_1 d^4x_2\,a^+(x_1)a^+(x_2)\Delta_F(x_1-x_2)]^m}{(2^m m!)}\,|\,0\rangle\end{aligned}$$

$$= \langle 0 | a(x_1) \cdots a(x_n) \times \exp\{\tfrac{1}{2} \int d^4x_1 d^4x_2 a^+(x_1) a^+(x_2) \Delta_F(x_1 - x_2)\}] | 0 \rangle.$$

Now for all g we have

$$\frac{d}{dg} G_n^g(x_1, \ldots, x_n) = i \int d^4x \langle \Phi | T(\phi(x_1) \cdots \phi(x_n) \phi^4(x) \times \exp[ig \int d^4x \, \phi^4(x)]) | \Phi \rangle = i \int d^4x \, G^g_{(n+4)}(x_1, \ldots, x_n, x, x, x, x).$$

Now consider

$$\tilde{G}_n(x_1, \ldots, x_n) = \langle 0 | a(x_1) \cdots a(x_n) \times \exp\{ig \int d^4x [a(x)]^4\} \times \exp[\tfrac{1}{2} \int d^4x_1 d^4x_2 a^+(x_1) a^+(x_2) \Delta_F(x_1 - x_2)]] | 0 \rangle.$$

Clearly $\tilde{G}_n$ satisfies the same differential equation as G_n^g for all g, and for $g = 0$, $\tilde{G}_n = G_n^0$. Therefore $G_n^g = \tilde{G}_n$. One may also, of course, simply expand the exponentials in the expression for $\tilde{G}_n$ and use the commutation relations and the fact that $a(x) | 0 \rangle = 0$, to see that one obtains the usual sum of Feynman diagrams.

APPENDIX B

Consider an arbitrary complex symmetric matrix $\mathcal{D}_{ij}$ with $\det(\mathcal{D}) \neq 0$. Then from the theory of the Jordan canonical form, there exists a similarity transformation S, such that

$$J = S \mathcal{D} S^{-1} = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_r \end{pmatrix}$$

where the J_i's are Jordan canonical blocks of the form

$$J_i = \begin{pmatrix} \lambda_i 1 & & & 0 \\ & \lambda_i 1 & & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}$$

and $\det(\mathcal{D}) \neq 0$ implies $\lambda_i \neq 0$ for all i. Now clearly $\mathcal{D}^{1/2} = S^{-1} J^{1/2} S$, where

$$J^{1/2} = \begin{pmatrix} J_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & J_r^{1/2} \end{pmatrix},$$

if J_i is an r-dimensional canonical block, then

$$J_i^{1/2} = \lambda_i^{1/2} \begin{pmatrix} C_0 & (C_1 \lambda_i^{-1}) & \cdots & (C_{(r-1)} \lambda_i^{-(r-1)}) \\ & C_0 & \cdots & (C_{(r-2)} \lambda_i^{-(r-2)}) \\ & & \ddots & \vdots \\ 0 & & & C_0 \; (C_1 \lambda_i^{-1}) \\ & & & C_0 \end{pmatrix}$$

where $C_0 = 1$ and for all $n \geq 0$

$$\sum_{i=0}^{n} C_i C_{(n-i)} = \begin{cases} 1, & n = 0, 1, \\ 0, & n = 2, 3, \cdots. \end{cases}$$

The unique solution of these equations is

$$C_n = \Gamma(\tfrac{3}{2}) / [\Gamma(n+1) \Gamma(\tfrac{3}{2} - n)].$$

Further we note that, for $\mathcal{D}$ sufficiently close to the identity,

$$\mathcal{D}^{1/2} = \exp\{\tfrac{1}{2} \log[I + (\mathcal{D} - I)]\},$$

is a convergent power series in $\mathcal{D}$ which explicitly satisfies $(\mathcal{D}^{1/2})_{ij} = (\mathcal{D}^{1/2})_{ji}$ and therefore by analytic continuation $\mathcal{D}^{1/2}$ is a symmetric matrix. Since the inverse of a symmetric matrix is symmetric, $\mathcal{D}^{-1/2}$ is also symmetric.

Finally we note that, for the case of boson field theory,

$$\mathcal{D} = \mathcal{D}(x_1, x_2) = \int \frac{d^4p}{(2\pi)^4} \exp[-ip(x_1 - x_2)] \frac{i}{p^2 - m^2 + i\epsilon},$$

and therefore

$$\mathcal{D}^{1/2}(x_1, x_2) = \sqrt{i} \int \frac{d^4p}{(2\pi)^4} \frac{\exp[-ip(x_1 - x_2)]}{\sqrt{p^2 - m^2 + i\epsilon}},$$

which is just the $O(4, 1)$ invariant solution of the Klein–Gordon equation in 4-space and 1 time dimensions, with the extra space coordinate set equal to zero. Explicitly,[12]

$$\mathcal{D}^{1/2}(x_1, x_2) = \frac{m^3}{(2\pi)^2} \frac{\sqrt{i}}{(ms)^2} \exp(-ims) \left(1 + \frac{1}{(ims)}\right),$$

where

$$s = +((x_1 - x_2)^2)^{1/2} \quad \text{if } (x_1 - x_2)^2 > 0,$$
$$s = -i(-(x_1 - x_2)^2)^{1/2} \quad \text{if } (x_1 - x_2)^2 < 0,$$
$$1/s^2 = 1/(x^2 - i\epsilon),$$

and

$$1/s^3 = 1/(x^2 - i\epsilon)^{3/2}.$$

$\mathcal{D}^{1/2}(x_1, x_2)$ is a well defined distribution since $(x^2 - i\epsilon)^\lambda$ is well defined everywhere except at the points $\lambda = -2 - k$, $k = 0, 1, 2, \cdots$.[13] Thus $\mathcal{D}^{1/2}(x_1, x_2)$ is actually a more elementary function than $\mathcal{D}(x_1, x_2)$.

APPENDIX C

We start from the standard expression[14]

$$G^e_{n,m}(i_1x_1, \ldots, i_nx_n; j_1y_1, \ldots, j_ny_n; \mu_1z_1, \ldots, \mu_mz_m) = \langle \Phi | T(\psi_{i_1}(x_1) \cdots \psi_{i_n}(x_n) \bar{\psi}_{j_1}(y_1) \cdots \bar{\psi}_{j_n}(y_n) \times A_{\mu_1}(z_1) \cdots A_{\mu_m}(z_m) \times \exp[ie \int d^4x \, \bar{\psi}(x) \not{A}(x) \psi(x)]) | \Phi \rangle,$$

where ψ and A are the free fields, $\bar{\psi} = \psi^+ \gamma_0$, and $| \Phi \rangle$ is the free particle vacuum state. For $e = 0$, we find (we set $m = 2l$, since, for $e = 0$, $G_{n,m} = 0$ if m is odd)

$$G^0_{n,m}(i_1x_1, \ldots, u_mz_m) = \frac{1}{2^l l!} \sum_{\sigma(m)} D^F_{\mu_{\sigma(1)} \mu_{\sigma(2)}}(x_{\sigma(1)} - x_{\sigma(2)}) \cdots \times D^F_{\mu_{\sigma(m-1)} \mu_{\sigma(m)}}(x_{\sigma(m-1)} - x_{\sigma(m)}) \times (-1)^{[n/2]} \sum_{\sigma(n)} t(\sigma) S^F_{i_1 j_{\sigma(1)}}(x_1 - y_{\sigma(1)}) \cdots$$

$$\times S^F_{i_n j_{\sigma(n)}}(x_n - y_{\sigma(n)})$$

where $[n/2]$ is the greatest integer less than $n/2$, $\sum_{\sigma(r)}$ is the sum over all permutations on r elements, and $t(\sigma) = \pm 1$, when σ is an even or odd permutation. It is then easy to show that

$$\begin{aligned} G^0_{n,m}(i_1x_1, \ldots, \mu_m z_m) &= \langle 0 | b_i(x_1) \cdots b_{i_n}(x_n) c_{j_1}(y_1) \cdots c_{j_n}(y_n) \\ &\times a_{\mu_1}(z_1) \cdots a_{\mu_m}(z_m) \\ &\times \exp\left(\tfrac{1}{2} \sum_{\mu\nu} \int d^4x_1 d^4x_2 \, a^+_\mu(x_1) a^+_\nu(x_2) D^F_{\mu\nu}(x_1 - x_2)\right) \\ &\times \exp\left(-\sum_{ij} \int d^4x d^4y \, b^+_i(x) S^F_{ij}(x-y) \, c^+_j(y)\right) | 0 \rangle. \end{aligned}$$

For all e we have

$$\begin{aligned} \frac{d}{de} G^e_{n,m}(i_1x_1, \ldots, \mu_m z_m) &= (-1)^{n+1} i \sum_{ij\mu} \int d^4x (\gamma_\mu)_{ji} \\ &\times G^e_{n+1,m+1}(i_1x_1, \ldots, i_nx_n, ix; j_1y_1 \cdots j_ny_n, jx; \\ &\qquad \mu_1 z_1 \cdots \mu_m z_m, \mu x). \end{aligned}$$

Consider

$$\begin{aligned} \tilde{G}_{n,m}(i_1x_1, \ldots, \mu_m z_m) &= \langle 0 | b_{i_1}(x_1) \cdots a_{\mu_m}(z_m) \\ &\times \exp\left\{ie \sum_{ij\mu} \int d^4x \, c_i(x) (\gamma_\mu)_{ij} b_j(x) a_\mu(x)\right\} \\ &\times \exp\left\{\tfrac{1}{2} \sum_{\mu\nu} \int d^4x_1 d^4x_2 \, a^+_\mu(x_1) a^+_\nu(x_2) D^F_{\mu\nu}(x_1 - x_2)\right\} \\ &\times \exp\left\{-\sum_{ij} \int d^4x d^4y \, b^+_i(x) S^F_{ij}(x-y) \, c^+_j(y)\right\} 0 \rangle. \end{aligned}$$

Clearly $\tilde{G}_{n,m}$ satisfies the same differential equation as $G^e_{n,m}$ for all e, and for $e = 0$, $G^e_{n,m} = \tilde{G}_{n,m}$. Therefore $G^e_{n,m} = \tilde{G}_{n,m}$.

*Supported in part by the National Science Foundation, Center for Theoretical Physics under Grant No. NSF GU 2061.

[1]E. Wichmann and C. H. Woo, J. Math. Phys. 2, 178 (1961).
[2]R. H. Dalitz, Proc. R. Soc. A 206, 509 (1951)
[3]See, for example, N. N. Bogoliubov and D. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1959), Chap. VII, Sec. 39.
[4]A. Erdelyi, *et al.*, *Higher Transcendental Functions, Vol. I* (McGraw-Hill, New York, 1953), Sec. 6.5., Eq. (3).
[5]Ref. 4, Sec. 6.13.1.
[6]R. J. Eden, *et al.*, *The Analytic S Matrix* (Cambridge U. P., Cambridge, 1966), see Sec. 2.1, Eq. 2.1.18.
[7]S. S. Schweber, *Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), see p. 611, Eqs. (34) and (36a), and p. 613, Eqs. (44) and (45).
[8]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Vol. II* (Interscience, New York, 1969). See Chap. XII, "Characteristic Classes," particularly Theorem 5.1.
[9]E. R. Caianiello, Nuovo Cimento, Suppl. 14, Ser. 10, 177 (1959).
[10]Bogoliubov, Ref. 3, Sec. 39.4. His function $S_0(u)$ is just our $\det(I+p)$.
[11]Schweber, Ref. 7, Sec. 17a, pp. 656—59.
[12]I. M. Gel'fand and G. E. Shilov, *Generalized Functions, Vol. I* (Academic, New York, 1964), p. 288, Eq. (5).
[13]Ref. 12, p. 275.
[14]Schweber, Ref. 7, Sec. 17a.

Equivalent Lagrangians and path integration for generalized mechanics*

Kuo-Shung Cheng

Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11790
(Received 12 November 1973; revised manuscript received 10 January 1974)

The equivalent Lagrangians for a generalized mechanical system are defined and the effects of using different members of the class of equivalent Lagrangians in classical mechanics and quantum mechanics are discussed. The path integration method is applied to quantize the generalized mechanical system.

I. INTRODUCTION

It is well known that for a mechanical system its equation of motion can be obtained from the principle of least action. The Lagrangian, which determines the action, is not uniquely defined. Thus, there is a class of equivalent Lagrangians all of which give the same equation of motion. Since the only thing which concerns classical mechanics is the equation of motion, choosing different members of the class of equivalent Lagrangians doesn't change anything. For quantum mechanics, different members of the class of equivalent Lagrangians mean different Hamiltonians and hence different Schrödinger equations. It is not obvious that all these equations represent the same system. After some investigations, we found that different members of the class of equivalent Lagrangians only change the wave functions up to a common phase factor. This common phase factor does not change any important properties of the system. That is, all the expectation values of observables are the same at all time, if they are the same at one time. We also investigated these properties for generalized mechanics (i.e., starting from a Lagrangian that depends on higher order time derivatives of the coordinates instead of just first order time derivatives) and found that the same properties exist, provided that we carefully defined the class of equivalent Lagrangians. It is precisely this point that has caused some troubles in the quantization of generalized mechanics.[1-5]

In order to prove the properties concerning quantum mechanics, we found that is is convenient to use path integration method to quantize the generalized mechanical system. The only thing that is not familiar is that the wavefunction now depends on $x, \dot{x}, \cdots$ up to $d^{n-1}x/dt^{n-1}$ if the Lagrangian contains d^nx/dt^n as the highest time derivative of the coordinate x. We will give some explanations to see why the wavefunction should depend on higher order time derivatives of the coordinates x.

II. DEFINITION OF THE CLASS OF EQUIVALENT LAGRANGIANS

In order to define the class of equivalent Lagrangians, we first define proper Lagrangian and improper Lagrangian. We feel that this step is necessary, although it is trivial, since it is precisely this point that causes troubles in the quantization of generalized mechanics. We define the proper Lagrangian as the *Lagrangian from which we can obtain the equation of motion via the variational principle*. If a Lagrangian is not a proper Lagrangian, we call it an improper Lagrangian.

Let us give an example. Consider the following Lagrangians

$$L_1 = -\tfrac{1}{2}x\ddot{x}, \quad L_2 = \tfrac{1}{2}\dot{x}^2 .$$

The actions are, respectively,

$$S_1 = \int_{t_1}^{t_2} (-\tfrac{1}{2}x\ddot{x})dt , \quad S_2 = \int_{t_1}^{t_2} (\tfrac{1}{2}\dot{x}^2)\,dt$$

and the variations of the actions are

$$\delta S_1 = \int_{t_1}^{t_2} -\ddot{x}\delta x dt - \tfrac{1}{2}x\delta\dot{x}\,\Big|_{t_1}^{t_2} + \tfrac{1}{2}\dot{x}\delta x\,\Big|_{t_1}^{t_2} , \tag{1}$$

$$\delta S_2 = \int_{t_1}^{t_2} -\ddot{x}\delta x dt + \tfrac{1}{2}\dot{x}\delta x\,\Big|_{t_1}^{t_2} . \tag{2}$$

From Eq. (1) we know that we have to fix $x(t_1)$, $x(t_2)$, $\dot{x}(t_1)$, and $\dot{x}(t_2)$ in order to get the equation of motion $\ddot{x}=0$. But in general a second order differential equation does not admit four initial conditions. From Eq. (2), we only have to fix $x(t_1)$ and $x(t_2)$ to obtain the equation of motion $\ddot{x}=0$. In general there is a solution. So, in these simple cases, we call L_1 an improper Lagrangian and L_2 a proper Lagrangian. We should exclude improper Lagrangians when we consider generalized mechanics, since they are not logically well defined. If we excluded the improper Lagrangian, we no longer have the troubles of Refs. 1—5.

According to the definition of proper Lagrangians, we have the following theorem:

For a proper Lagrangian if it involves d^nx/dt^n as the highest time derivative of the generalized coordinates x, then it must contain a term of the form[6]

$$A\left(\frac{d^{n-1}x}{dt^{n-1}}, \ldots, \frac{dx}{dt}, x, t\right)\left(\frac{d^nx}{dt^n}\right)^2$$

where

$$A\left(\frac{d^{n-1}x}{dt^{n-1}}, \ldots, \frac{dx}{dt}, x, t\right)$$

is not equal to zero at any time. We are not going to give the proof here, because it is obvious. Now we can define the class of equivalent Lagrangians.

Definition: Two proper Lagrangians are equivalent if they contain d^nx/dt^n as the common highest time derivative of the coordinates x and differ only by a term of the form

$$\frac{d}{dt} f\left(\frac{d^{n-1}x}{dt^{n-1}}, \ldots, \frac{dx}{dt}, x, t\right)$$

where f is an arbitrary function of its arguments.

III. DEFINITION OF PATH INTEGRATION FOR GENERALIZED MECHANICS

It was Feynman[7] who first introduced the path integration method into quantum mechanics. For practical purposes, it works as well as canonical quantization. But for a general Lagrangian of the form $L=\frac{1}{2}g_{ij}\dot{q}^i\dot{q}^j$, it seems that there is a difference between canonical quantization and path integration.[8] There are many reasons to believe the correctness of the path integration method.[9]

It is a *straightforward* generalization to define path integration for generalized mechanics. Let us consider a proper Lagrangian L:

$$L=L\left(\frac{d^nx}{dt^n},\ldots,\frac{dx}{dt},x,t\right). \tag{3}$$

L contains a term

$$A\left(\frac{d^{n-1}x}{dt^{n-1}},\ldots,\frac{dx}{dt},x,t\right)\left(\frac{d^nx}{dt^n}\right)^2.$$

We closely follow Feynman's original propositions. That is, we assume that for infinitesimal time intervals, the propagator of the wave function is proportional to the exponential of $i/\hbar$ times the classical action. Since the classical action is uniquely defined only when we fix $x(t),\dot{x}(t),\ldots,d^{n-1}x/dt^{n-1}$ at two end points, we conclude that the propagator depends on $x(t),\ldots,d^{n-1}x/dt^{n-1}$ at the two end points and hence the wavefunction depends on $x,\dot{x},\ldots,d^{n-1}x/dt^{n-1}$. This is not the common situation, but it is the most natural generalization from ordinary quantum mechanics to generalized quantum mechanics. For convenience we will define

$$Q_n=\frac{d^{n-1}x}{dt^{n-1}},\quad Q_{n-1}=\frac{d^{n-2}x}{dt^{n-2}}\ \ldots Q_1=x.$$

The wavefunction $\psi(Q(t_2),t_2)$ at time t_2 with coordinates $Q(t_2)$ is connected with the wavefunction $\psi(Q(t_1),t_1)$ at time t_1 as follows:

$$\psi(Q(t_2),t_2)=\int K(Q(t_2),t_2;Q(t_1),t_1)\psi(Q(t_1,t_1)g(Q(t_1))dQ(t_1), \tag{4}$$

where Q stands for $(Q_n,Q_{n-1},\ldots,Q_1)$ and

$$dQ(t_1)=dQ_n(t_1)\cdots dQ_1(t_1), \tag{5}$$

$$g(Q(t_1))=[A(Q_n,\ldots,Q_1)]^{n/2} \tag{6}$$

is the weighting factor to make sure that the measure gdQ is invariant under generalized coordinate transformations. Eq. (4) is the definition of the propagator $K(Q(t_2),t_2;Q(t_1),t_1)$. According to our assumption, we can write down the expression for $K(Q(t_2),t_2;Q(t_1),t_1)$, that is,

$$K(Q(t_2),t_2;Q(t_1),t_1)=\lim_{m\to\infty}\frac{1}{A_m}\int\exp\left(\frac{i}{\hbar}\sum_{k=0}^{m}S_{cl}(Q(t_1+k\epsilon),t_1+k\epsilon;Q(t_1+(k+1)\epsilon),t_1+(k+1)\epsilon)\right)\times\prod_{k=1}^{m}g(Q(t_1+k\epsilon))dQ(t_1+k\epsilon), \tag{7}$$

where $t_1+(m+1)\epsilon=t_2$ and A_m is a normalization constant which depends on m only. $S_{cl}(Q(t_k);Q(t_{k+1}))$ is the action along the classical path between $Q(t_k)$ and $Q(t_{k+1})$ in the time interval ϵ, that is

$$S_{cl}(Q(t_k);Q(t_{k+1}))=\text{stationary value of }\int_{t_1+k\epsilon}^{t_1+(k+1)\epsilon}L(t')dt'.$$

In order to get acquainted with the definition of the path integration method, we consider the simplest example of a generalized mechanical system,

$$L=\tfrac{1}{2}(\ddot{x})^2. \tag{8}$$

We shall first find the Schrödinger equation and then calculate the propagator for this Lagrangian. For infinitesimal time intervals, Eqs. (4) and (7) give

$$\psi(Q_2(t+\epsilon),Q_1(t+\epsilon),t+\epsilon)=\frac{1}{A}\int\exp\left(\frac{i}{\hbar}S_{cl}(Q_2(t+\epsilon),Q_1(t+\epsilon));Q_2(t),Q_1(t)\right)\times\psi(Q_2(t),Q_1(t),t)dQ_1(t)dQ_2(t) \tag{9}$$

where

$$S_{cl}=\int_t^{t+3}\tfrac{1}{2}(\dddot{x}_0)^2dt', \tag{10}$$

$x_0(t')$ satisfies the equation of motion

$$\ddddot{x}_0(t')=0 \tag{11}$$

and the boundary conditions

$$x_0(t')\big|_{t'=t}=Q_1(t),\quad x_0(t')\big|_{t'=t}=Q_2(t),$$
$$x_0(t')\big|_{t'=t+\epsilon}=Q_1(t+\epsilon),\quad x_0(t')\big|_{t'=t+\epsilon}=Q_2(t+\epsilon). \tag{12}$$

It is easy to obtain S_{cl}:

$$S_{cl}=\frac{2}{\epsilon}\left[3\left(\frac{\Delta}{\epsilon}-Q_2(t+\epsilon)\right)^2+3\left(\frac{\Delta}{\epsilon}-Q_2(t+\epsilon)\right)\dot{\Delta}+\dot{\Delta}^2\right] \tag{13}$$

where

$$\Delta=Q_1(t+\epsilon)-Q_1(t),$$
$$\dot{\Delta}=Q_2(t+\epsilon)-Q_2(t). \tag{14}$$

We can expand $\psi(Q(t),t)$ around $Q(t+\epsilon)$ and obtain

$$\psi(Q(t+\epsilon),t)+\epsilon\frac{\partial\psi(Q(t+\epsilon),t)}{\partial t}+\cdots=\frac{1}{A}\int\exp\left(\frac{i}{\hbar}S_{cl}\right)\times\left[\psi(Q(t+\epsilon),t)-\dot{\Delta}\frac{\partial\psi}{\partial Q_2}-\dot{\Delta}\frac{\partial\psi}{\partial Q_1}+\frac{1}{2!}\left(\dot{\Delta}^2\frac{\partial^2\psi}{\partial Q_2^2}+2\Delta\dot{\Delta}\frac{\partial^2\psi}{\partial Q_1\partial Q_2}+\Delta^2\frac{\partial^2\psi}{\partial Q_1^2}+\cdots\right)\right]\times dQ_1(t)dQ_2(t). \tag{15}$$

After a rather lengthy but *straightforward* calculation, we get

$$\psi+\epsilon\frac{\partial\psi}{\partial t}+\cdots=\frac{1}{A}\left[\frac{i\pi\hbar\epsilon^2}{\sqrt{3}}\psi+\epsilon\frac{i\pi\hbar\epsilon^2}{\sqrt{3}}\frac{i\hbar}{2}\left(\frac{\partial^2\psi}{\partial Q_2^2}-Q_2\frac{\partial\psi}{\partial Q_1}\right)+\cdots\right], \tag{16}$$

that is, we get

$$A=i\pi\hbar\epsilon^2/\sqrt{3} \tag{17}$$

and the equation for $\psi(Q(t),t)$

$$i\hbar\frac{\partial\psi}{\partial t}=-\frac{\hbar^2}{2}\frac{\partial^2\psi}{\partial Q_2^2}-i\hbar Q_2\frac{\partial\psi}{\partial Q_1}\,. \tag{18}$$

The propagator $K(Q_2(t_2),Q_1(t_2),t_2;Q_2(t_1),Q_1(t_1),t_1)$ can be obtained from the definition, that is, from Eq. (7). Since L is a quadratic in the variable x, we have other methods to calculate the propagator K. We symbolize Eq. (7) as

$$K(Q(t_2),t_2;Q(t_1),t_1)=\frac{1}{A}\int\exp\left(\frac{i}{\hbar}\int_{t_1,Q_1(t_1),Q_2(t_1)}^{t_2,Q_1(t_2),Q_2(t_2)}\tfrac{1}{2}(\ddot{x})^2dt\right)\prod_t dx(t) \tag{19}$$

where $\prod_t dx(t)$ means that we sum over all paths satisfying the conditions

$$x(t)\big|_{t=t_1}=Q_1(t_1),\quad x(t)\big|_{t=t_1}=Q_2(t_1),$$
$$x(t)\big|_{t=t_2}=Q_1(t_2),\quad x(t)\big|_{t=t_2}=Q_2(t_2).$$

Set

$$x(t)=x_0(t)+y(t) \tag{20}$$

where $x_0(t)$ satisfies the equation $\ddot{x}_0(t)=0$ and the boundary conditions

$$x_0(t_1)=Q_1(t_1),\quad x_0(t_2)=Q_1(t_2),$$
$$\dot{x}_0(t_1)=Q_2(t_1),\quad \dot{x}_0(t_2)=Q_2(t_2).$$

We substitute Eq. (20) into Eq. (19) and get

$$K(Q(t_2),t_2;Q(t_1),t_1)=\exp\left(\frac{i}{\hbar}\int_{t_1}^{t_2}\tfrac{1}{2}(\ddot{x}_0)^2dt\right)\frac{1}{A}\int\exp\left(\frac{i}{\hbar}\int_{t_1,0,0}^{t_2,0,0}\tfrac{1}{2}(\ddot{y})^2dt\right)\prod_t dy(t)\,. \tag{21}$$

Using the following condition

$$\lim_{t_2\to t_1}K(Q(t_2),t_2;Q(t_1),t_1)=\delta(Q(t_2)-Q(t_1)).$$

We get

$$K(Q(t_2),t_2;Q(t_1),t_1)=\frac{\sqrt{3}}{i\pi\hbar(t_2-t_1)^2}\exp\left\{\frac{2i}{\hbar(t_2-t_1)}\left[3\left(\frac{Q_1(t_2)-Q_1(t_1)}{t_2-t_1}-Q_2(t_2)\right)^2+3\left(\frac{Q_1(t_2)-Q_1(t_1)}{t_2-t_1}-Q_2(t_2)\right)[Q_2(t_2)-Q_2(t_1)]+[Q_2(t_2)-Q_2(t_1)]^2\right]\right\}. \tag{22}$$

IV. EFFECTS OF USING EQUIVALENT LAGRANGIANS TO GENERALIZED CLASSICAL MECHANICS AND QUANTUM MECHANICS

According to the definition of equivalent Lagrangians, one can generate the whole class of equivalent Lagrangians by adding a term of the form

$$\frac{d}{dt}f\left(\frac{d^{n-1}x}{dt^{n-1}},\ldots,x,t\right)$$

to a particular Lagrangian

$$L_0=L_0\left(\frac{d^nx}{dt^n},\ldots,x,t\right).$$

The action becomes

$$S=\int_{t_1}^{t_2}\left(L_0+\frac{df}{dt}\right)dt=S_0+f(t_2)-f(t_1).$$

Classical mechanics is concerned with the problem of finding the path $x(t)$ which makes S stationary with $x(t)$, $\dot{x}(t),\ldots$ and $d^{n-1}x/dt^{n-1}$ at time $t=t_1$ and $t=t_2$ fixed. Since $f(t_1)$ and $f(t_2)$ depend on fixed quantities only, we have

$$\delta f(t_1)=\delta f(t_2)=0\quad\text{and}\quad\delta S=\delta S_0\,.$$

Thus the equation of motion is the same for all members of the class of equivalent Lagrangians.

For quantum mechanics, let us set ψ_0 and ψ to represent the wave functions corresponding to the Lagrangians L_0 and L_0+df/dt. If ψ_0 and ψ have the same expectation values for all observables $h(\partial/\partial Q_i,\ Q_i,t)$ at a time t_1, that is

$$\int\psi_0^*(Q(t_1),t_1)h\left(\frac{\hbar\partial}{i\partial Q_i},Q_i,t\right)\psi_0(Q(t_1),t_1)\,g(Q(t_1))dQ(t_1)$$
$$=\int\psi^*(Q(t_1),t_1)h\left(\frac{\hbar\partial}{i\partial Q_i}-\frac{\partial f}{\partial Q_i},Q_i,t\right)\psi(Q(t_1),t_1)\times g(Q(t_1))dQ(t_1)\,. \tag{23}$$

Then we are going to prove that they have the same expectation value at all time. The two systems are equivalent. We write down ψ_0 and ψ at time t as

$$\psi_0(Q(t),t)=\int K_0(Q(t),t;Q(t_1),t_1)\psi_0(Q(t_1),t_1)g(Q(t_1))dQ(t_1)\,, \tag{24}$$

$$\psi(Q(t),t)=\int K(Q(t),t;Q(t_1),t_1)\psi(Q(t_1)t_1)g(Q(t_1))dQ(t_1)\,, \tag{25}$$

where

$$K(Q(t),t;Q(t_1),t_1)=K_0((Q(t),t;Q(t_1),t_1)\exp\left(\frac{i}{\hbar}[f(Q(t))-f(Q(t_1))]\right) \tag{26}$$

according to Eq. (7).

From Eq. (23), we are forced to conclude that

$$\psi(Q(t_1),t_1)=\psi_0(Q(t_1),t_1)\exp\left(+\frac{i}{\hbar}f(Q(t_1),t_1)\right). \tag{27}$$

Using Eqs. (24), (25), (26), and (27), we get

$$\psi(Q(t),t)=\psi_0(Q(t),t)\exp\left(\frac{i}{\hbar}f(Q(t),t)\right) \tag{28}$$

thus

$$\int\psi_0^*(Q(t),t)h\left(\frac{\hbar}{i}\frac{\partial}{\partial Q_i},Q_i,t\right)\psi_0(Q(t),t)g(Q(t))dQ(t)$$
$$=\int\psi^*(Q(t),t)h\left(\frac{\hbar}{i}\frac{\partial}{\partial Q_i}-\frac{\partial f}{\partial Q_i},Q_i,t\right)\psi(Q(t),t)g(Q(t))dQ(t)\,. \tag{29}$$

Note that the changes from $(\hbar/i)(\partial/\partial Q_i)$ in the lhs of Eqs. (23) and (29) to

$$\left(\frac{\hbar}{i}\frac{\partial}{\partial Q_i}-\frac{\partial f}{\partial Q_i}\right)$$

in the rhs of Eqs. (23) and (29) are due to the changes of canonical conjugate variables. For example, P_n, the canonical conjugate variable corresponding to Q_n are, respectively,

$$P_n = \frac{\partial L_0}{\partial Q_n} = A \frac{d^n x}{dt^n} \quad \text{in } L_0$$

and

$$P_n = \frac{\partial L}{\partial Q_n} = A \frac{d^n x}{dt^n} + \frac{\partial f}{\partial Q_n} \quad \text{in } L .$$

From Eq. (29) we conclude that the two systems L_0 and $L_0 + df/dt$ are equivalent.

ACKNOWLEDGMENT

The author would like to thank Professor C.N. Yang for his constant encouragement and delightful discussions.

*Work supported in part by National Science Foundation Grant No. GP-32998X.

[1]C.F. Hayes and J.M. Jankowski, Nuovo Cimento B 58, 494 (1968).
[2]C.F. Hayes, J. Math. Phys. 10, 1555 (1969).
[3]C. Ryan, J. Math. Phys. 13, 283 (1972).
[4]H. Tesser, J. Math. Phys. 13, 796 (1972).
[5]D. Anderson, J. Math. Phys. 14, 934 (1973).
[6]We restrict to the form $A(d^n x/dt^n)^2$ because it is simple. We can have $A(d^n x/dt^n)^m$ with $m \neq 1$.
[7]R.P. Feynman, Rev. Mod. Phys. 20, 327 (1948).
[8]K.S. Cheng, J. Math. Phys. 13, 1733 (1972).
[9]K.S. Cheng, J. Math. Phys. 14, 980 (1973).

Hamiltonian cosmology: Death and transfiguration*

Michael P. Ryan Jr.

Center for Relativity Theory, The University of Texas at Austin, Austin, Texas 78712
(Received 26 April 1973)

Following work of MacCallum and Taub, we show that for Bianchi-type universes of Ellis–MacCallum class B (Bianchi types II, III, IV, V, $VI_{h\neq -1}$, $VII_{h=0}$) Hamiltonian cosmology breaks down. We construct a quasi-Hamiltonian theory for these models.

I. INTRODUCTION

A. Death

Hamiltonian cosmology is dead. The application of Hamiltonian techniques to the study of homogeneous cosmologies has been found to be invalid. Ehlers[1] questioned whether simply inserting a homogeneous metric into an action functional for general relativity led to an action for the Einstein equations of that metric. For the most part Hamiltonian techniques were applied to Bianchi-type universes[2] and it is now certain that for some of these Bianchi types Ehlers' worry was justified. Ellis and MacCallum[3] have divided the nine Bianchi-type models into two groups, Class A (types I, II, VI_{-1}, VII_0, VIII, IX), and Class B (types II, III, IV, V, $VI_{h\neq -1}$, $VII_{h\neq 0}$), and Hamiltonian cosmology fails for universes of Class B. It is relieving to note that the models most closely studied in Hamiltonian cosmology (types I and IX) are in Class A. It is also important to know that type V models face problems, because if the present value for the cosmic luminous-matter density is any reasonable fraction of the total density, the real universe is type V.

The failure of Hamiltonian cosmology was first noted by MacCallum and Taub[4] in a Hamiltonian formulation of their own. Most of the work on Hamiltonian cosmology has been based on the techniques of Arnowitt, Deser, and Misner[5] (ADM), so we must investigate ADM Hamiltonian cosmology for the various Bianchi types. In this paper we show that the ADM method applied in a straightforward fashion is valid for Class A vacuum cosmologies but, like the MacCallum–Taub Hamiltonian method, fails for Class B models. We also show that this failure may be a symptom of a very general problem with the Einstein[6] variational principle in noncoordinate frames.

B. Transfiguration

It is possible to retrieve something of the ADM approach in Class B universes. In fact, the evolution equations of these models can be reduced to an almost-Hamiltonian form, and this is done below. A by-product of this reduction is that the "potential" terms in the Hamiltonians of Class A and the "quasi-Hamiltonians" of Class B can be written out explicitly for all Bianchi types. This is done in Appendix A.

Hamiltonian cosmology has often led to quantum cosmology, quantized models based on the Hamiltonians which appear. The fact that the evolution equations of Class B models are not Hamiltonian leads to difficulties in quantizing these models. This is discussed briefly in the final section of the paper.

II. THE APPLICATION OF THE ADM METHOD TO HOMOGENEOUS METRICS

A. The ADM method and homogenization

ADM[5] write the action for general relativity in the absence of matter as $(i,j=1,2,3)$

$$I=\int \mathcal{L}d^4x=\int [-g_{ij}\partial_\tau\pi^{ij}+Ng^{1/2}\{{}^3R+g^{-1}(\tfrac{1}{2}\pi^2-\pi^{ij}\pi_{ij})\} +2N_i\pi^{ij}{}_{|j}-2(\pi^{ij}N_j-\tfrac{1}{2}\pi N^i+N^{|i}\sqrt{g})_{,i}]d\tau d^3x. \tag{2.1}$$

We use the superscript[4] to label four-dimensional quantities and[3] to label quantities on $\tau=$ constant hypersurfaces. We define

$$g_{ij}\equiv {}^4g_{ij},\quad N\equiv(-{}^4g^{00})^{-1/2},$$
$$N_i\equiv {}^4g_{0i},\quad \pi^{ij}\equiv(-{}^4g)^{1/2}({}^4\Gamma^0_{pq}-g_{pq}{}^4\Gamma^0_{rs}g^{rs})\,g^{ip}g^{jq}.$$

The matrix g^{ij} is the reciprocal of g_{ij} and henceforth all i,j indices will be raised and lowered by means of g_{ij} and g^{ij} unless we specifically say otherwise. In the action (2.1) τ is whatever coordinate we take to be "time", $|$ means covariant differentiation on a $\tau=$ constant surface, and ${}^3R$ is the curvature scalar formed from the spatial metric g_{ij}. Finally, $\pi\equiv\pi^k{}_k$. The variational equations for this action (varying with respect to π^{ij}, g_{ij}, N, and N_i) are

$$\partial_\tau g_{ij}=2Ng^{-1/2}(\pi_{ij}-\tfrac{1}{2}g_{ij}\pi)+N_{i|j}+N_{j|i}, \tag{2.2a}$$

$$\partial_\tau\pi^{ij}=-N\sqrt{g}({}^3R^{ij}-\tfrac{1}{2}g^{ij}{}^3R)+Ng^{-1/2}g^{ij}(\pi^{mn}\pi_{mn}-\tfrac{1}{2}\pi^2) -2Ng^{-1/2}(\pi^{im}\pi^j_m-\pi\pi^{ij})+\sqrt{g}(N^{|ij}-g^{ij}N^{|m}{}_{|m}) +(\pi^{ij}N^m)_{|m}-N^i{}_{|m}\pi^{mj}-N^j{}_{|m}\pi^{mi}, \tag{2.2b}$$

$${}^3R+g^{-1}(\pi^2-\pi^{ij}\pi_{ij})=0, \tag{2.2c}$$

$$\pi^{ij}{}_{|j}=0. \tag{2.2d}$$

ADM[5] show that these are equivalent to Einstein's equations in vacuum.

In the next section we attempt to apply the action (2.1) and Eqs. (2.2) to homogeneous, Bianchi-type cosmological metrics. In anticipation of this we study the "homogenization" of (2.1) and (2.2) and do this in a noncoordinate frame. Again anticipating we shall take the frame to be $(d\tau,\sigma^i)$, where the $\sigma^i=\xi^i_j(x^k)dx^j$, $i,j,k=1,2,3$. By "homogenization" we mean that for the components of any geometric quantity A (metric components, symbols, Riemann tensor components) in the frame $(d\tau,\sigma^i)$, $A_{,i}\equiv X_iA=0$, where the X_i are vectors dual to the σ^i. Finally we shall let $N_i=0$, a coordinate condition which is usual in cosmology. Because π^{ij} is a density and not a tensor, we must be careful in transforming from one frame to another, but one can see that Eqs. (2.2) with $N_i=0$ transform easily because only time derivatives of π^{ij} appear and the frame we are

transforming to has $\sigma^0 = d\tau$. Equation (2.2d) is also easily transformed because the covariant derivative of a density is well-defined.

The homogenized action and the homogenization of (2.2) are

$$I = \int [-g_{ij}\partial_\tau \pi^{ij} + Ng^{1/2}\{{}^3R + g^{-1}(\tfrac{1}{2}\pi^2 - \pi^{ij}\pi_{ij}\} + 2N_i\pi^{ij}{}_{|j}]\, d\tau \wedge \sigma^1 \wedge \sigma^2 \wedge \sigma^3, \tag{2.3a}$$

$$\partial_\tau g_{ij} = 2Ng^{-1/2}(\pi_{ij} - \tfrac{1}{2}g_{ij}\pi), \tag{2.3b}$$

$$\partial_\tau \pi^{ij} = -N\sqrt{g}({}^3R^{ij} - \tfrac{1}{2}g^{ij}{}^3R) + \tfrac{1}{2}Ng^{-1/2}g^{ij}(\pi^{mn}\pi_{mn} - \tfrac{1}{2}\pi^2) - 2Ng^{-1/2}(\pi^{im}\pi^j_m - \tfrac{1}{2}\pi\pi^{ij}), \tag{2.3c}$$

$${}^3R + g^{-1}(\tfrac{1}{2}\pi^2 - \pi^{ij}\pi_{ij}) = 0, \tag{2.3d}$$

$$\pi^{ij}{}_{|j} = 0. \tag{2.3e}$$

Notice that the choice $N_i = 0$ has enabled us unambiguously to discard the total-divergence term in (2.1). We must now ask if Eqs. (2.3b–e) result from varying (2.3a). Because the variations for Eqs. (2.3b, d) involve no space derivatives [and because $N_{i|j} = 0$, also (2.3e)] these equations are obtained from (2.3a). The same is true for the term on the left-hand side of (2.3c) and the last two terms on the right-hand side. It is in the first term on the right-hand side of (2.3c) that the variational principle breaks down. This may seem strange as it comes from the variation of the term $Ng^{1/2}{}^3R$ in (2.3a) and is a reflection of the three-dimensional analog of the Einstein variational principle for general relativity.

This variational equation is not necessarily straightforward in noncoordinate frames, as we shall see in the next subsection.

What happens in the specific homogeneous case we are considering? For a homogeneous three-space we have[7]

$${}^3R = -C^k_{il}C^l_{kj}g^{ij} - \tfrac{1}{4}C^k_{il}C^m_{sj}g^{ls}g_{km}g^{ij} + C^k_{km}g^{ml}C^s_{ls} \tag{2.4a}$$

$${}^3R^{ij} = -\tfrac{1}{2}C^k_{ls}C^l_{kt}g^{si}g^{tj} - \tfrac{1}{2}C^a_{ls}C^b_{kt}g^{lk}g_{ab}g^{si}g^{tj} + \tfrac{1}{4}g^{st}g^{pr}C^a_{sp}C^b_{tr}g_{am}g_{bn}g^{mi}g^{nj} + \tfrac{1}{2}C^s_{sr}g^{rt}(C^p_{tm}g_{pn} + C^p_{tn}g_{pm})g^{mi}g^{nj}, \tag{2.4b}$$

where C^i_{jk} is, as usual $\Gamma^i_{kj} - \Gamma^i_{jk}$. From these two expressions it is easy to see that

$$\delta(\sqrt{g}{}^3R) = (-{}^3R^{ij} + \tfrac{1}{2}g^{ij}{}^3R)\sqrt{g}\delta g_{ij} + (\tfrac{1}{2}C^s_{sr}g^{rt}(C^j_{tm}g^{mi} + C^i_{tn}g^{nj}) - C^s_{sr}C^p_{tp}g^{ri}g^{tj})\sqrt{g}\delta g_{ij}. \tag{2.5}$$

We shall see in the next section how this affects the Hamiltonian formulation of Bianchi-type models.

B. The Einstein variational principle in non-coordinate frames

The usual Einstein variational principle[6] for a Riemannian or a pseudo-Riemannian space is

$$\delta \int R\sqrt{\pm g}\, dV = \int (-R^{ij} + \tfrac{1}{2}g^{ij}R)\sqrt{\pm g}\delta g_{ij}\, dV. \tag{2.6}$$

The usual proof of this relation begins by writing

$$\delta \int g^{ij}R_{ij}\sqrt{\pm g}\, dV = \int -R^{ij}\sqrt{\pm g}\delta g_{ij} + \tfrac{1}{2}g^{ij}R\sqrt{\pm g}\delta g_{ij}\, dV + \int g^{ij}\delta R_{ij}\sqrt{\pm g}\, dV \tag{2.7}$$

and then showing[8] that $g^{ij}\delta R_{ij} = w^l_{;l}$, where w^l is the vector $g^{ij}\delta\Gamma^l_{ij} - g^{il}\delta\Gamma^j_{ij}$. Rewriting $w^l_{;l}$ as

$$\frac{1}{\sqrt{\pm g}}\frac{\partial}{\partial x^l}(\sqrt{\pm g}w^l),$$

and using Gauss' theorem, the integral of $g^{ij}\delta R_{ij}\sqrt{\pm g}$ over volume can be transformed into an integral of w^l over a hypersurface surrounding the volume. Since the variations of the field are zero at the integration limits, $w^l = 0$ there and the term in δR_{ij} drops out.

Let us examine this proof in a noncoordinate frame. One might already be suspicious, as the vector w^l involves the Γ^i_{jk} which are not frame-independent quantities. Let us write $\tilde{g}_{ij}$, $\tilde{R}_{ij}$, $\tilde{\Gamma}^i_{jk}$ for the metric, Ricci tensor, and Christoffel symbols in a noncoordinate frame $\{\sigma^1, \sigma^2, \ldots, \sigma^n\}$. As before we have (where $dV \equiv \sigma^1 \wedge \sigma^2 \wedge \ldots \wedge \sigma^n$):

$$\delta \int \tilde{R}_{ij}\tilde{g}^{ij}\sqrt{\pm\tilde{g}}\, dV = -\int \tilde{R}^{ij}\delta g_{ij}\sqrt{\pm g}\, dV + \tfrac{1}{2}\int \tilde{g}^{ij}\tilde{R}\sqrt{\pm\tilde{g}}\delta\tilde{g}_{ij}\, dV + \int \tilde{g}^{ij}\delta\tilde{R}_{ij}\sqrt{\pm\tilde{g}}\, dV. \tag{2.8}$$

In a general frame $\tilde{R}_{ij} = \tilde{\Gamma}^k_{ij,k} - \tilde{\Gamma}^k_{ik,j} + \tilde{\Gamma}^l_{ij}\tilde{\Gamma}^k_{lk} - \tilde{\Gamma}^l_{ik}\tilde{\Gamma}^k_{lj} - C^l_{kj}\tilde{\Gamma}^k_{il}$. Once one notices that $\delta\tilde{\Gamma}^i_{jk}$ is a tensor as it is in a coordinate frame it is not difficult to show that

$$\tilde{g}^{ij}\delta\tilde{R}_{ij} = \tilde{w}^l_{;l} - C^k_{km}\,\delta\tilde{\Gamma}^m_{ji}\tilde{g}^{ij}, \tag{2.9}$$

where w^l is a vector whose components are the analogues of those in the coordinate frame with each quantity in the expression replaced by the same quantity with a tilde. By means of the generalized Stokes' theorem[9] $\int_M d\omega = \int_{\partial M}\omega$ we can show that $\int \tilde{w}^l_{;l}\sqrt{\pm\tilde{g}}\, dV = 0$ as in a coordinate frame. The second term on the right-hand side of (2.9) is not the divergence of a vector.

This result means that one must be careful in applying the Einstein variational principle in noncoordinate frames. The exception to this caveat (to borrow from Ellis and MacCallum's terminology describing Bianchi-type models) are *spaces of class A* for which $C^k_{km} = 0$ and frames in which C^k_{km} can be made to vanish.

Gowdy[10] has constructed a correct variational principle for the density $\sqrt{g}R$ in which the objects which are varied are not the metric components in the frame. Even this variational principle is not free from problems of boundary conditions which arise because $\delta\Gamma^m_{ij}$ which appears in w^l contains derivatives of the metric, so that we may say roughly that not only the δg_{ij}, but $(\delta g_{ij})_{,h}$ must be held zero on the boundary of the integration region. MacCallum and Taub[4] feel that these boundary problems preclude finding any valid variational principle for Class B models, but whether the illness that they found is related to the problem we have discussed in the ADM case or is actually of this more basic type needs further study.[11]

In the previous subsection we saw that choosing one of the forms defining the frame we used to be $d\tau$ made the only modification of the four-dimensional variational principle appear as a modification of the variation of $\sqrt{g}{}^3R$.

III. THE ADM METHOD AND BIANCHI-TYPE UNIVERSES

A. The Bianchi-type models and the Ellis-MacCallum classification scheme

We write the metric of any Bianchi-type model as

$$ds^2 = -dt^2 + e^{-2\Omega(t)} e^{2\beta(t)}_{ij} \sigma^i \sigma^j, \tag{3.1}$$

where $\Omega(t)$ is a scalar and $\beta^{(t)}_{ij}$ is a 3×3 traceless matrix. Wherever $\Omega(t)$ is a monotonic function of t it makes sense to make the coordinate transformation $t \to \Omega(t)$ and the metric becomes (with the definitions of Sec. II)

$$ds^2 = (-N^2 + N^i N_i) d\Omega^2 + 2N_i d\Omega \sigma^i + e^{-2\Omega} e^{2\beta(\Omega)}_{ij} \sigma^i \sigma^j. \tag{3.2}$$

Because $N_i = 0$ in (3.1) and $\Omega = \Omega(t)$, $N_i = 0$ in (3.2), and we will discard them from here on.

It is useful to parametrize β_{ij} by means of a parametrization due to Ryan[2] and Misner[12] where we write

$$\beta_{ij} = R^{-1}_{ik} (\beta_d)_{kl} R_{lj},$$

where R_{ij} is an Euler rotation matrix and β_d is a diagonal traceless matrix parametrized by $\beta_d = \mathrm{diag}(\beta_+ + \sqrt{3}\beta_-, \beta_+ - \sqrt{3}\beta_-, -2\beta_+)$. We parametrize a generic rotation matrix by $R = \exp(\psi\kappa_3)\exp(\theta\kappa_2)\exp(\varphi\kappa_3)$, where

$$\kappa_3 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \kappa_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

For a general Bianchi-type model a solution of the Einstein equations gives the five quantities $\beta_+, \beta_-, \varphi, \psi, \theta$ as functions of Ω. The procedure of Misner then gives these quantities as functions of cosmic time t.

For any Bianchi-type model the σ^i obey

$$d\sigma^i = C^i_{jk} \sigma^j \wedge \sigma^k, \tag{3.3}$$

where the C^i_{jk} are constant. Ellis and MacCallum[3] have used the symmetries of the C^i_{jk} to write

$$C^i_{jk} = \epsilon_{jkl} m^{li} + \delta^i_k a_j - \delta^i_j a_k, \tag{3.4}$$

where m^{ij} is a 3×3 symmetric matrix of constants and a_i is a triple of constants. Various choices of m and a give the various Bianchi types. For example $m = 0$, $a = 0$ (Type I); $a = 0$, $m^{ij} = \delta^i_j$ (Type IX); $m = 0$, $a_i = -\delta^i_3$ (Type V); $a_i = -\frac{1}{2}\delta^i_3$, $m = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ (Type IV).

B. The application of the ADM method to Bianchi-type universes

Since the Bianchi-type universes are homogeneous in the sense of Sec. II we need only insert $g_{ij} = e^{-2\Omega} e^{2\beta}_{ij}$ into (2.3) and replace τ by Ω. We can achieve an explicit Hamiltonian form for (2.3) by parametrizing π^{ij} by

$$\pi^{ij} = e^{-\beta}[(p_{ij}/2\pi) + \tfrac{1}{3}\delta_{ij}\pi] e^{-\beta}, \tag{3.5}$$

letting $\pi \equiv H/2\pi$ and parametrizing p_{ij} by

$$6p_{ij} = R^{-1} \Bigg\{ \alpha_1 p_+ + \alpha_2 p_- + \alpha_3 \frac{3p_\psi}{\sinh(2\sqrt{3}\beta)} + \alpha_4 \frac{3(p_\varphi \sin\psi - p_\psi \cos\theta \sin\psi + p_\theta \cos\psi \sin\theta}{\sin\theta \sinh(3\beta_+ + \sqrt{3}\beta_-)} + \alpha_5 \frac{3(p_\theta \sin\psi \sin\theta - p_\varphi \cos\psi + p_\psi \cos\psi \cos\theta}{\sin\theta \sinh(3\beta_+ - \sqrt{3}\beta_-)} \Bigg\} R, \tag{3.6}$$

where

$$\alpha_1 = \mathrm{diag}(1, 1, -2), \quad \alpha_2 = \mathrm{diag}(\sqrt{3}, -\sqrt{3}, 0),$$

$$\alpha_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \alpha_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \alpha_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

With these definitions (2.3a) becomes

$$I = \int [p_+ d\beta_+ + p_- d\beta_- + p_\psi d\psi + p_\varphi d\varphi + p_\theta d\theta - H d\Omega] \tag{3.7}$$

(assuming the spatial constraint equations are solved). Solving the constraint (2.3d) for H we find

$$H^2 = 6\mathrm{Tr}[(p)^2] - 24\pi^2 g^3 R. \tag{3.8}$$

The space constraints may be easily computed (see Appendix B). The potential term g^3R is given for all Bianchi types in Appendix A. The combination of (3.7) and (3.8) make up a well-defined Hamiltonian system, but because of the failure of the variational principle we must examine Equations (2.3b) and (2.3c) carefully. Inserting our parametrization for π^{ij} and β_{ij} into (2.3b) we quickly find that

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \tag{3.9}$$

where a dot means $d/d\Omega$ and we have let $(\beta_+ \to q_1, \ \beta_- \to q_2, \ \psi \to q_3, \ \theta \to q_4, \ \varphi \to q_5)$ and $p_+ \to p_1$, etc.

Inserting our parametrizations into (2.3c) we find, after an unconscionable amount of algebra (in which the beauty of mathematics is shown by the most likely collections of terms combining to give exactly what is needed) we find two sets of equations

$$\dot{H} = \frac{\partial H}{\partial \Omega} \tag{3.10}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} - e^{-4\Omega} H^{-1} \mathcal{V}_i(q_k). \tag{3.11}$$

The $\mathcal{V}_i$ are

$$\mathcal{V}_1 = 144\pi^2 \{ -A_2 M^{13} e^{-2\beta_+ - 2\sqrt{3}\beta_-} + A_1 M^{23} e^{-2\beta_+ - 2\sqrt{3}\beta_-} - (A_1)^2 e^{-2\beta_+ - 2\sqrt{3}\beta_-} - (A_2)^2 e^{-2\beta_+ + 2\sqrt{3}\beta_-} + 2(A_3)^2 e^{4\beta_+} \}, \tag{3.12a}$$

$$\mathcal{V}_2 = 144\pi^2 \{ \tfrac{1}{3} A_1 M^{23} e^{-2\beta_+ - 2\sqrt{3}\beta_-} - \tfrac{1}{3} A_2 M^{13} e^{-2\beta_+ + 2\sqrt{3}\beta_-} - \sqrt{3}(A_1)^2 e^{-2\beta_+ - 2\sqrt{3}\beta_-} - \sqrt{3}(A_2)^2 e^{-2\beta_+ + 2\sqrt{3}\beta_-} \}, \tag{3.12b}$$

$$\mathcal{V}_3 = 144\pi^2 \{ \tfrac{1}{3} A_3 M^{22} e^{4\beta_+ - 2\sqrt{3}\beta_-} - \tfrac{1}{3} A_1 M^{23} e^{-2\beta_+} + \tfrac{1}{3} A_1 M^{13} e^{-2\beta_+} + \tfrac{1}{3} A_3 M^{11} e^{4\beta_+ + \sqrt{3}\beta_-} - A_1 A_2 e^{-2\beta_+ + 2\sqrt{3}\beta_-} - A_1 A_2 e^{-2\beta_+ - 2\sqrt{3}\beta_-} \} \sinh(2\sqrt{3}\beta_-),$$

$$\mathcal{V}_4 = 144\pi^2 \{ \sin\psi \sin\theta \sinh(3\beta_+ + \sqrt{3}\beta_-) [\tfrac{1}{3} A_3 M^{32} e^{2\beta_+ - 2\sqrt{3}\beta_-} - \tfrac{1}{3} A_2 M^{33} e^{-5\beta_+ + \sqrt{3}\beta_-} - \tfrac{1}{3} A_1 M^{12} e^{\beta_+ - \sqrt{3}\beta_-} + \tfrac{1}{3} A_2 M^{11} e^{\beta_+ + 3\sqrt{3}\beta_-} - A_1 A_3 e^{4\beta_+} - A_1 A_3 e^{-2\beta_+ - 2\sqrt{3}\beta_-}]$$
$$- \cos\psi \sin\theta \sinh(3\beta_+ - \sqrt{3}\beta) [\tfrac{1}{3} A_1 M^{33} e^{-5\beta_+ - \sqrt{3}\beta_-} - \tfrac{1}{3} A_3 M^{13} e^{\beta_+ + \sqrt{3}\beta_-} + \tfrac{1}{3} A_2 M^{12} e^{\beta_+ + \sqrt{3}\beta_-} - \tfrac{1}{3} A_1 M^{22} e^{-3\beta_+ - 3\sqrt{3}\beta_-} - A_2 A_3 e^{4\beta_+} - A_2 A_3 e^{-2\beta_+ + 2\sqrt{3}\beta_-}] \},$$

$$\mathcal{V}_5 = 144\pi^2\{\cos\psi \sinh(3\beta_+ + \sqrt{3}\beta_-)[\tfrac{1}{3}A_3M^{32}e^{2\beta_+ - 2\sqrt{3}\beta_-} - \tfrac{1}{3}A_3M^{33}e^{-5\beta_+ + \sqrt{3}\beta_-} - \tfrac{1}{3}A_1M^{12}e^{\beta_+ - \sqrt{3}\beta_-} - \tfrac{1}{3}A_2M^{11}e^{\beta_+ + \sqrt{3}\beta_-} - A_1A_2e^{4\beta_+} - A_1A_2e^{-2\beta_+ - 2\sqrt{3}\beta_-}] + \sin\psi \sinh(3\beta_+ - \sqrt{3}\beta_-)[\tfrac{1}{3}A_1M^{33}e^{-5\beta_+ - \sqrt{3}\beta_-} - \tfrac{1}{3}A_3M^{13}e^{\beta_+ + \sqrt{3}\beta_-} + \tfrac{1}{3}A_2M^{12}e^{\beta_+ + \sqrt{3}\beta_-} - \tfrac{1}{3}A_1M^{22}e^{-3\beta_+ - 3\sqrt{3}\beta_-} - A_2A_3e^{4\beta_+} - A_2A_3e^{-2\beta_+ + 2\sqrt{3}\beta_-}]\},$$

where $M^{ij} = R_{ik}m^{kl}R^{-1}_{lj}$ and $A_i = R_{ij}a_j$. Notice that all the terms in each of the $\mathcal{V}_i$ are proportional to one of the A_i so that the $\mathcal{V}_i$ are all zero if the $C^k_{ki} \propto R^{-1}_{ij}A_j$ are zero.

Classically, the fact that Equation (3.11) are not exactly Hamiltonian is not too important, since they are close enough for calculations to be carried out. In fact, the terms in $\mathcal{V}_i$ will be negligible for large Ω except in certain small portions of $\beta_+, \beta, \varphi, \psi, \theta$ space. In the "wall" picture of Misner[12] these terms would add "quasiwalls" where the universe point would change direction, and the non-Hamiltonian form would only make the computation of bounce laws from these quasiwalls more difficult. When we try to study the quantization of Class B universes, we find ourselves in more difficulty. This will be discussed below.

We have neglected the space constraints (2.3e). These are given in Appendix B for all Bianchi-type models. In diagonal type V models a helpful cancellation occurs. $\mathcal{V}_2$ becomes zero and the space constraint (see Appendix B) implies $\beta_+ = \text{const}$. This allows one to write the evolution equations in Hamiltonian form. For other Class B models the space constraints are not helpful.

While reducing the Einstein equations to the forms (3.9), (3.10), (3.11) is helpful, it is necessary to attempt to find a true Hamiltonian formulation based on a correct variational principle.

IV. QUANTIZATION

There is very little that can be said about quantizing Class B models in the context of the ADM method as it was applied above. It is impossible to reconcile the commutation relations $[p_i\ q_i] = i\hbar$ with the equation of motion (3.11). Notice that it is possible to quantize the equations far from the quasiwalls where the non-Hamiltonian terms become important, so we could construct wavepackets in these regions, but we could not connect these solutions with any kind of bounce law. In order to quantize these models properly, we would need a theory based on a correct variational principle freed from problems of boundary conditions.

APPENDIX A: MISCELLANY

For Class B universes we have

Bianchi Type	m^{ij}	a_i
III	$\frac{1}{2}\alpha_3$	$-\frac{1}{2}\delta^i_3$
IV	diag(1, 0, 0)	$-\delta^i_3$
V	0	$-\delta^i_3$
VI_h	$-\frac{1}{2}(1-h)\alpha_3$	$-(1+h)\delta^i_3$
$VIII_h$	diag(0, −1, 0) + α_3	$-h\delta^i_3$

Here α_3 is the same matrix as in the expression (3.6).

The potential terms in the "Hamiltonian" in (3.8) are $24\pi^2g^3R$. In terms of β_ϕ, R, M^{ij}, and A_i we find that (for any Bianchi type)

$$24\pi^2g^3R = e^{-4\Omega}\{e^{-2\beta\delta}_{ij}(\text{adj}M^{ij} - M^{ij}M^{kl}e^{2\beta\delta}_{il}e^{2\beta\delta}_{jk} - 6A_iA_je^{-2\beta\delta}_{ij}\},$$

where adj$M$ is the classical adjoint of M.

The triplet A_i is not difficult to write out for any Class B model, in fact,

$$A_i = E(\sin\psi \sin\theta,\ \cos\psi \sin\theta,\ \cos\theta),$$

where E is the factor miltiplying δ^i_3 in the table above. The matrix M is quite complicated even in the simplest cases. The diligent can convince themselves of this by computing M for type IV.

APPENDIX B: THE SPACE CONSTRAINTS

If we let $\alpha_{ij} \equiv 6R^{-1}_{ik}p_{kl}R_{lj}$, then the space constraints (2.3e) become

$$e^{-\beta\delta}_{if}\epsilon_{gfr}e^{\beta\delta}_{rw}\alpha_{gr}e^{\beta\delta}_{ad}M^{dw} + 3e^{-\beta\delta}_{ib}A_ie^{-\beta\delta}_{iq}\alpha_{qb} = 0.$$

For a type V model as an example these three equations read

$$A_1(p_+ + \sqrt{3}p_-)e^{-2(\beta_+ + \sqrt{3}\beta_-)} + 3A_2e^{-2\beta_+}\frac{p_\psi}{\sinh(2\sqrt{3}\beta_-)} + 3A_3e^{\beta_+ - \sqrt{3}\beta_-}\frac{(p_\varphi \sin\psi - p_\psi\cos\theta\sin\psi + p_\theta\cos\psi\sin\theta)}{\sin\theta\sinh(3\beta_+ + \sqrt{3}\beta_-)} = 0, \tag{B1}$$

$$3A_1e^{-2\beta_+}\frac{p_\psi}{\sinh(2\sqrt{3}\beta_-)} + A_2(p_+ - \sqrt{3}p_-)e^{-2(\beta_+ - \sqrt{3}\beta_-)} + 3A_3e^{\beta_+ + \sqrt{3}\beta_-} \times \frac{(p_\theta\sin\psi\sin\theta - p_\varphi\cos\psi + p_\psi\cos\psi\cos\theta)}{\sin\theta\sinh(3\beta_+ - \sqrt{3}\beta_-)} = 0 \tag{B2}$$

$$3A_1e^{\beta_+ - \sqrt{3}\beta_-}\frac{(p_\varphi\sin\psi - p_\psi\cos\theta\sin\psi + p_\theta\cos\psi\sin\theta)}{\sin\theta\sinh(3\beta_+ + \sqrt{3}\beta_-)} + 3A_2e^{\beta_+ + \sqrt{3}\beta_-} \times \frac{(p_\theta\sin\psi\sin\theta - p_\varphi\cos\psi + p_\psi\cos\psi\cos\theta)}{\sin\theta\sinh(3\beta_+ - \sqrt{3}\beta_-)} - 2A_3e^{4\beta_+}p_+ = 0. \tag{B3}$$

Notice that for diagonal type-V models these equations reduce to $p_+ = 0$.

*Supported in part by National Science Foundation Grant GP 32039.

[1]J. Ehlers (private communication).
[2]See M. Ryan, *Hamiltonian Cosmology* (Springer, Heidelberg, 1972).
[3]G. Ellis and M. MacCallum, Comm. Math. Phys. **12**, 108 (1969).
[4]M. MacCallum and A. Taub, Comm. Math. Phys. **25**, 173 (1972).
[5]See A. Arnowitt, S. Deser, and C. Misner in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).
[6]A. Einstein, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl., 1111 (1916).
[7]M. Ryan and L. Shepley, *Homogeneous Relativistic Cosmologies* (Princeton U.P., Princeton, N.J., to be published).
[8]See L. Landau and E. Lifshitz, *Classical Theory of Fields* (Addison—Wesley, Reading, Mass., 1965).
[9]See M. Spivak, *Calculus on Manifolds* (Benjamin, New York, 1965).
[10]R. Gowdy, Ph.D. thesis, Yale University (unpublished).
[11]While this paper was in press, C. Misner pointed out the possibility of an error in (2.9). If the last term is (2.9) is an error, then the failure of the variational principle is due entirely to boundary terms.
[12]C. Misner, Phys. Rev. **186**, 1328 (1969).

Spherically symmetric static space-times which admit stationary Killing tensors of rank two

I. Hauser and R. J. Malhiot

Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616
(Received 7 July 1972)

All spherically symmetric static space–times for which a nonredundant stationary Killing tensor of rank two exists are found and discussed. The results include explicit forms for the line elements, the Killing tensors, and the corresponding quadratic constants of geodesic motion. Included as special cases are relativistic analogs of the Lenz vector and of the quadratic constant of the motion peculiar to the spatial harmonic oscillator.

1. INTRODUCTION

Any symmetric tensor $K_{\alpha\beta}$ which satisfies the condition

$$K_{(\alpha\beta;\gamma)}=0 \tag{1}$$

is called a *Killing tensor[1] of the second rank*. $K_{\alpha\beta}$ will be called *redundant* if it is equal to some linear combination with constant coefficients of the metric tensor $g_{\alpha\beta}$ and of terms of the form $A_{(\alpha}B_{\beta)}$ where A_α and B_β are Killing vectors. Our objective is to find all spherically symmetric static space–times for which a nonredundant stationary[2] Killing tensor of rank two exists.

This objective is a limited part of a more general program of determining space–times which admit nonredundant Killing tensors in addition to some given group of isometries. The program is still in its infancy with regard to techniques of calculation. There is the well-known technique[3] of handling a symmetric tensor by choosing an orthonormal tetrad relative to which the tensor has a Jordan canonical form. However, as discussed in Sec. 5, this method is not appropriate when there is a given group of isometries. A subsidiary purpose of this paper is to explore methods of handling Killing tensors when working in a coordinate system adapted to the given isometries.

With few exceptions,[4,5] previous work[6–9] on space–times or other Riemannian spaces which admit Killing tensors did not impose specific isometry constraints such as ours. Nor was the problem of redundancy considered in any detail.

Even apart from problems of redundancy, little has been done on Killing tensors in general relativity. A noteworthy exception is the work of V. I. Golikov[7] and of C. F. Martin[8] who independently found all space–times which admit geodesic correspondences. Each geodesic correspondence uniquely determines a Killing tensor which typifies the correspondence. Further details pertinent to our own work are given in the discussion of results in Sec. 5.

It is worthwhile reviewing the principal motivations for studying such hidden symmetry conditions as the existence of nonredundant Killing tensors. We recall that Eq. (1) is equivalent to the statement that, for any geodesic motion of a test particle with a world velocity p^μ, the scalar $K_{\mu\nu}p^\mu p^\nu$ is a constant of the motion. Therefore, our particular problem is one of finding all spherically symmetric static space–times which admit stationary quadratic first integrals of geodesic motion subject to the proviso that these first integrals be functionally independent of $g_{\mu\nu}p^\mu p^\nu$ and of all linear first integrals of geodesic motion. This is an extension of the classical problem of determining those central force fields in which a nonrelativistic particle has a constant of the motion which is a second degree polynomial in the momentum components and which is functionally independent of the Hamiltonian and the angular momentum. There are exactly two central potentials, viz. r^{-1} and r^2, which satisfy these criteria.

The above classical example illustrates how hidden symmetry criteria can select dynamically interesting fields. This example and its analogues in particle physics are the principal motivations for searching for space–times which admit nonredundant Killing tensors.

The Kerr metric furnishes additional support for such a program in general relativity. Here, we have in mind the quadratic first integral of geodesic motion which Carter[5] has found as a byproduct of his work demonstrating that the Hamilton–Jacobi and Klein–Gordon equations are separable for the Kerr metric.

Carter[10] has also derived a broad class of line elements which include the Kerr metric as a special case and which satisfy certain separability criteria in addition to having the two Killing vectors characteristic of axially symmetric stationary space–times. The general relation[11] between separability criteria and hidden symmetry criteria is still an open question and deserves further investigation. However, Carter's separability conditions have no direct bearing on the present paper, because all spherically symmetric static space–times satisfy these conditions.

The present paper contains perhaps the first calculation of Killing tensors in which the nonredundant ones are completely sifted out from the others. We are able to do this easily, because we are working with a class of spaces whose Killing vectors are all explicitly known.[4] These Killing vectors and the corresponding first integrals of geodesic motion are given in Sec. 2.

The key steps in our calculation of the Killing tensors and the corresponding line elements are outlined in Sec. 3. Almost all calculational details are omitted, since the calculations are long, though straightforward in our opinion. Schwarzschild coordinates θ, ϕ, r, t were found convenient for the calculations of Sec. 3, but the key results are simplest when expressed in terms of coordinates θ, ϕ, s, t for which the line element has the form

$$\Phi=r^2(d\theta^2+\sin^2\theta d\phi^2)+(r/s)^4ds^2-e^\nu dt^2, \tag{2}$$

TABLE I. Constants of geodesic motion V_i and W_i peculiar to the Minkowski, Einstein static model, and de Sitter spaces. $i=1,2,3$. The line elements are all of the form $\Phi = r^2(d\theta^2 + \sin^2\theta d\phi^2) + (\exp\lambda)dr^2 - (\exp\nu)dt^2$.

Name of Space	$\exp\lambda$	$\exp\nu$	V_i	W_i
Minkowski	1	1	$\dot{x}_i$	$t\dot{x}_i + p_4 x_i$
Einstein static model	$\left[1-\left(\frac{r}{R}\right)^2\right]^{-1}$	1	u_i	0
de Sitter	$\left[1-\left(\frac{r}{R}\right)^2\right]^{-1}$	$1-\left(\frac{r}{R}\right)^2$	v_i	w_i

where r and ν depend only on s. The results for the metric tensor, the constants of the motion, and the Cartan components of the matter tensor are all given in terms of these coordinates in Sec. 4.

The results and perspectives for further work are discussed in the fifth section. The results do not include the Schwarzschild line element. Nor do they include the Einstein and de Sitter static models of the universe, but several of our line elements have the Einstein and de Sitter line elements as limiting cases, and the corresponding Killing tensors have redundant limits corresponding to the special isometries of these models.

The fifth section also includes a discussion of the singularities of the matter tensor and the sign of the energy density. Only the limiting cases of the Einstein and de Sitter static models have perfect fluid matter tensors. The new line elements obtained by us cannot be attributed to any obviously interesting single source.

We now review the well-known[3] Killing vectors of spherically symmetric static space–times. In addition to the generators of infinitesimal rotations and time translations, there are the Killing vectors which are peculiar to the Einstein and de Sitter static models of the universe. These Killing vectors will be needed in later calculations for the purpose of identifying redundant Killing tensors.

2. THE KILLING VECTORS

Greek small letter scripts will have the values 1, 2, 3, 4. Latin small scripts will have the values 1, 2, 3. Schwarzschild coordinates,

$$\xi^1=\theta, \quad \xi^2=\phi, \quad \xi^3=r, \quad \xi^4=t$$

are used. Our line element is

$$\Phi = g_{\alpha\beta}d\xi^\alpha d\xi^\beta = r^2(d\theta^2+\sin^2\theta d\phi^2) + (\exp\lambda)\,dr^2 - (\exp\nu)\,dt^2, \tag{3}$$

where λ and ν depend on r alone. The summation convention will apply to all small letter scripts.

The metric components from Eq. (3) are inserted into the defining equation, $K_{(\alpha;\beta)}=0$, for a Killing vector K_α. Ten distinct differential equations for the Killing vector components are thereby obtained. The solution of these equations is given below in a form which covers all possible choices of $\lambda(r)$ and $\nu(r)$:

$$K_\alpha = \epsilon g_{\alpha 4} + \omega_{ij}x_i\frac{\partial x_j}{\partial\xi^\alpha} + \rho^\beta_\alpha\frac{\partial}{\partial\xi^\beta}(Ca_jx_j + Sb_jx_j). \tag{4}$$

ϵ, $\omega_{ij}=-\omega_{ji}$, a_j, and b_j are arbitrary real parameters, and

$$x_1 = r\sin\theta\cos\phi, \quad x_2 = r\sin\theta\sin\phi, \quad x_3 = r\cos\theta; \tag{5}$$

$$\rho^1_1=\rho^2_2=\exp(-\tfrac{1}{2}\lambda), \quad \rho^3_3=\exp(\tfrac{1}{2}\lambda),$$
$$\rho^4_4=-\exp(\tfrac{1}{2}\nu); \quad \rho^\beta_\alpha=0, \quad \alpha\neq\beta. \tag{6}$$

C and S depend at most on t and are both zero except for the Minkowski space, the Einstein static model, and the de Sitter model. For the Minkowski space,

$$C=1, \quad S=t; \tag{7a}$$

for the Einstein static model,

$$C=1, \quad S=0; \tag{7b}$$

for the de Sitter model,

$$C=\cosh(t/R), \quad S=\sinh(t/R), \tag{7c}$$

where R is a constant. $\exp\lambda$ and $\exp\nu$ for these spaces are given in Table I.

The Killing vector of Eq. (4) yields the following constant of geodesic motion:

$$K_\alpha\dot{\xi}^\alpha = \epsilon p_4 + \tfrac{1}{2}\omega_{ij}l_{ij} + a_iV_i + b_iW_i, \tag{8}$$

where $\dot{\xi}^\alpha = d\xi^\alpha/d\tau$, and τ is an arbitrarily chosen affine parameter. The constants of geodesic motion p_4 and l_{ij} are given by the equations

$$p_4 = -(\exp\nu)\dot{t}, \tag{9}$$
$$l_{ij} = x_i\dot{x}_j - x_j\dot{x}_i.$$

The constants of geodesic motion V_i and W_i are both zero except for the Minkowski space, the Einstein static model, and the de Sitter model. For these spaces, they are given in Table I in terms of dynamical variables which have broader application than the present context and which will now be dfined.

The dynamical variable u_i is defined for all spherically symmetric static space-times by the equation

$$u_i = [(\exp\tfrac{1}{2}\lambda)\dot{r}\delta_{ij} - s^{-1}l_{ij}]\,(x_j/r). \tag{10}$$

The variable s in Eq. (10) is the same one used in Eq. (2) and is defined apart from a sign and a constant of integration by the equation

$$ds^2 = (s/r)^4\,(\exp\lambda)dr^2. \tag{11}$$

s will be chosen positive for all $\lambda(r)$. In the special cases of the Minkowski, Einstein static model, and de Sitter spaces, the constant of integration is chosen so that integration of Eq. (11) yields

$$s = (\exp\tfrac{1}{2}\lambda)r. \tag{12}$$

Note that $u_i=\dot{x}_i$ for the Minkowski space.

The dynamical variables v_i and w_i which appear in Table I are defined for all spherically symmetric static space–times for which

$$\exp(-\nu) = a_0 + b_0s^2, \tag{13}$$

where a_0 and b_0 are constants. They are defined as follows:

$$v_i = u_i\cosh(\sqrt{b_0}\,t) + s_ip_4\sqrt{b_0}\,\sinh(\sqrt{b_0}\,t),$$
$$w_i = u_i\sinh(\sqrt{b_0}\,t) + s_ip_4\sqrt{b_0}\cosh(\sqrt{b_0}\,t), \tag{14}$$

where

$$s_i = (s/r)x_i. \tag{15}$$

In the special case of the de Sitter model, $\exp\nu = \exp(-\lambda)$, and $\sqrt{b_0} = 1/R$.

The variables p_4, l_{ij}, u_i, v_i, and w_i as defined above by Eqs. (9), (10), and (14) will also appear later in our Killing tensor calculations. The procedure for computing the nonredundant Killing tensors is considered next.

3. THE KILLING TENSORS

Corresponding to all possible values of the scripts $(\alpha\beta\gamma)$ in Eq. (1), there are twenty distinct differential equations which are to be used to determine the Killing tensor components. We work with the contravariant components, since the equations are then a little simpler in our particular problem. The form of the metric given in Eq. (3) will be used from the outset of the calculations; also, only those solutions of Eq. (1) which are independent of t are considered.[2]

Those seven of the differential equations which correspond to the script values $(ij4)$ and (444) involve only the components K^{i4} and are identical with the differential equations which define any time independent Killing vector K^α whose fourth component is zero. Since none of the other equations contain K^{i4}, the solution of Eq. (1) is completely reducible to the sum,

$$K^{\alpha\beta} = A^{(\alpha}B^{\beta)} + C^{\alpha\beta},$$

where

$$K^{i4} = A^i B^4, \quad K^{ij} = C^{ij}, \quad K^{44} = C^{44},$$

A^α is the Killing vector given by Eq. (4) with $\epsilon = 0$ and $dC/dt = dS/dt = 0$, and B^α is the Killing vector given by Eq. (4) with $\omega_{ij} = 0$ and $C = S = 0$.

Since K^{i4} is redundant, only the thirteen equations involving K^{ij} and K^{44} are of further interest to us. The following radial functions play a key role in the analysis of these equations:

$$R_1 = \int dr(e^{\lambda/2}/r^2), \quad R_2 = \int dr(R_1 e^{\lambda/2}/r^2),$$
$$R_3 = e^{-\lambda/2}/r, \quad R_4 = r^{-2} - R_1 R_3, \quad R_5 = 1. \tag{16}$$

Out of the thirteen equations, the seven corresponding to the script values (333), (443), (332), (331), (123), (113), and (223) can be integrated in succession to furnish us with the radial dependences of the Killing tensor components. The integrations are straightforward and yield

$$K^{33} = e^{-\lambda} f^{33}, \quad K^{44} = f^{44} - e^{-\nu} f^{33},$$

$$K^{23} = e^{\lambda/2}\left(f^{23} - \frac{R_1}{2\sin^2\theta}\frac{\partial f^{33}}{\partial \phi}\right),$$

$$K^{31} = e^{-\lambda/2}\left(f^{31} - \frac{R_1}{2}\frac{\partial f^{33}}{\partial \theta}\right),$$

$$K^{11} = f^{11} + \frac{1}{r^2} f^{23} - 2R_1\frac{\partial f^{31}}{\partial \theta} + R_2\frac{\partial^2 f^{33}}{\partial \theta^2},$$

$$K^{12} = f^{12} - R_1\left(\frac{\partial f^{23}}{\partial \theta} + \frac{1}{\sin^2\theta}\frac{\partial f^{31}}{\partial \phi}\right)$$

$$+ \frac{R_2}{\sin^2\theta}\left(\frac{\partial^2 f^{33}}{\partial \phi \partial \theta} - \cot\theta\frac{\partial f^{33}}{\partial \phi}\right),$$

$$K^{22} = f^{22} + \frac{1}{r^2\sin^2\theta} f^{33} - \frac{2R_1}{\sin^2\theta}\left(\frac{\partial f^{23}}{\partial \phi} + \cot\theta\, f^{31}\right)$$

$$+ \frac{R_2}{\sin^2\theta}\left(\frac{1}{\sin^2\theta}\frac{\partial^2 f^{33}}{\partial \phi^2} + \cot\theta\frac{\partial f^{33}}{\partial \theta}\right). \tag{17}$$

Above, f^{ij} and f^{44} depend only on θ and ϕ and are still to be determined. (The dependence on ϕ will clearly involve only the terms 1, $\cos\phi$, $\sin\phi$, $\cos 2\phi$, and $\sin 2\phi$).

We have exhausted all but six of the equations (1). Those four of the remaining equations which correspond to the script values (112), (221), (111), and (222) are now used to determine f^{ij}. The four equations all have the following structure after substituting the expressions (17) into them and grouping terms:

$$\sum_{A=1}^{5} R_A F_{AB}(\theta, \phi) = 0; \quad B = 1, 2, 3, 4. \tag{18}$$

The radial functions R_A are defined by Eqs. (16). The spherical functions F_{AB} are linear homogeneous expressions in f^{ij}. We do not list these expressions since they are rather long and do not illuminate the key steps in the calculations.

Before solving Eqs. (18) for the spherical functions f^{ij}, it is expedient to classify the various cases of linear interdependence of the radial functions R_A. Let the number of functions in any maximal linearly independent subset of the set $\{R_1, R_2, R_3, R_4, R_5\}$ be called its *rank*. Since R_1, R_2, and R_5 are linearly independent for any $\lambda(r)$, the rank is always 3, 4, or 5. The cases of rank 3 and 4 have, in turn, a variety of subcases depending on the particular linear relations which hold between these radial functions.

The number of subcases is significantly lessened by appropriate choices of the constants of integration in the definitions (16) of R_1 and R_2. By such choices, the linear structure of the functions R_A corresponding to any given $\lambda(r)$ can be reduced to one of the following types:

I. $R_3 = -\beta R_1$, $R_4 = 4\beta R_2$, rank = 3;

II. $R_3 = \alpha$, $R_4 = -3\alpha R_1$, rank = 3;

III. $R_3 = \alpha R_2$, rank = 4;

IV. $R_4 = \gamma R_1 + \delta$, rank = 4;

V. $R_4 = 4\beta R_2 + \gamma R_1$, rank = 4;

VI. rank = 5.

Above, α and β are nonzero constants, whereas γ and δ are constants which may be zero.

For each of the above $\lambda(r)$ types I to VI, specific linear relations are obtained between the expressions F_{AB} in Eqs. (18). These relations are then solved for f^{ij}. For all $\lambda(r)$ except those corresponding to type I with $\beta = 1$, $\frac{1}{4}$, or 4 and type V with $\gamma = 0$ and $\beta = 1$ or $\frac{1}{4}$, the solution is as follows:

$$f^{11} = a^{11}\cos 2\phi + b^{11}\sin 2\phi + c^{11}, \tag{19a}$$

$$f^{12} = \alpha^{12}\cos\phi + \beta^{12}\sin\phi + \cot\theta(-a^{11}\sin 2\phi + b^{11}\cos 2\phi), \tag{19b}$$

$$f^{22}=2\cot\theta(-\alpha^{12}\sin\phi+\beta^{12}\cos\phi)$$
$$-\cot^2\theta(a^{11}\cos2\phi+b^{11}\sin2\phi)+c^{11}\csc^2\theta+c^{22}, \tag{19c}$$

$$f^{33}=c^{33}, \tag{19d}$$

$$f^{31}=f^{23}=0. \tag{19e}$$

Above and in the sequel, a^{ij}, b^{ij}, c^{ij}, α^{ij}, β^{ij}, and $\gamma^{\alpha\beta}$ are arbitrary constants. Eqs. (19d) or (19e) are the only ones which are changed in the exceptional cases.

In the exceptional cases, whether type I or type V, when $\lambda(r)$ is such that

$$R_4=4n^2R_2, \quad n=1,\tfrac{1}{2}, \tag{20}$$

Eq. (19d) is replaced by

$$f^{33}=[(2n-1)(a^{33}\cos2\phi+b^{33}\sin2\phi)+c^{33}](\cos2n\theta-1)$$
$$+(\alpha^{33}\cos\phi+\beta^{33}\sin\phi)\sin2n\theta$$
$$+\gamma^{33}, \quad (n=1,\tfrac{1}{2}). \tag{21}$$

In the exceptional case, type I, where $\lambda(r)$ is such that both

$$R_4=4n^2R_2, \text{ and } R_3=-n^2R_1, \quad n=1,2, \tag{22}$$

Eqs. (19e) are replaced by

$$f^{31}=(a^{31}\cos2\phi+b^{31}\sin2\phi+c^{31})\sin n\theta$$
$$+(\alpha^{31}\cos\phi+\beta^{31}\sin\phi)\cos n\theta,$$
$$f^{23}=[n(a^{31}\sin2\phi-b^{31}\cos2\phi)+(2-n)c^{23}]\cos n\theta$$
$$+[(n-1)(-\alpha^{31}\sin\phi+\beta^{31}\cos\phi)$$
$$+(2-n)(\alpha^{23}\cos\phi+\beta^{23}\sin\phi)]\sin n\theta \tag{23}$$
$$+\cot\theta[2(-a^{31}\sin2\phi+b^{31}\cos2\phi)\sin n\theta$$
$$+(-\alpha^{31}\sin\phi+\beta^{31}\cos\phi)\cos n\theta], \quad (n=1,2).$$

Equations (20) to (23) cover all exceptions to Eqs. (19). We stress that Eq. (21) holds if Eq. (20) holds regardless of whether the rank is three or four. However, Eqs. (23) hold only in that rank three case specified by Eqs. (22).

There remain only the following two differential equations corresponding to the script values (441) and (442) in Eq. (1):

$$\frac{\partial}{\partial\theta}\left[f^{44}+\left(\frac{s}{2}\frac{de^{-\nu}}{ds}-e^{-\nu}\right)f^{33}\right]+s^2\frac{de^{-\nu}}{ds}f^{31}=0,$$
$$\frac{\partial}{\partial\phi}\left[f^{44}+\left(\frac{s}{2}\frac{de^{-\nu}}{ds}-e^{-\nu}\right)f^{33}\right]+s^2\frac{de^{-\nu}}{ds}(\sin^2\theta)f^{23}=0. \tag{24}$$

The radial coordinate s was defined by Eq. (11). Note, from Eqs. (16), that $R_1=-1/s$.

Upon applying the condition of integratability to the above equations (24), we obtain

$$\frac{de^{-\nu}}{ds}\left[\frac{\partial f^{31}}{\partial\phi}-\frac{\partial}{\partial\theta}(\sin^2\theta f^{23})\right]=0, \tag{25}$$

which provides a possible constraint on f^{31} and f^{23} in the event that $\nu(s)$ is not a constant. By examining Eqs. (19e) and (23), we see that this integrability condition is automatically satisfied except when

$$R_4=4R_2, \quad R_3=-R_1, \quad d\nu(s)/ds\neq0. \tag{26}$$

In this case, Eq. (25) imposes constraints on Eqs. (23) such that

$$f^{31}=(\alpha^{31}\cos\phi+\beta^{31}\sin\phi)\cos\theta+c^{31}\sin\theta,$$
$$f^{23}=(-\alpha^{31}\sin\phi+\beta^{31}\cos\phi)\csc\theta. \tag{27}$$

There are no further changes in the preceding expressions for f^{ij}.

We now use Eqs. (24) to find f^{44}. We examine these equations for each of the possible f^{ij} given by Eqs. (19) to (23) and by Eq. (27).

First, suppose Eqs. (19d) and (19e) hold. Then Eqs. (24) have the following solution for arbitrary $\nu(s)$:

$$f^{44}=\gamma^{44}. \tag{28}$$

We may now show that the Killing tensor obtained by substituting Eqs. (19) and (28) into Eqs. (17) has the form

$$K^{\alpha\beta}=\gamma^{33}g^{\alpha\beta}+\tfrac{1}{2}(A^\alpha B^\beta+A^\beta B^\alpha),$$

where A^α is the Killing vector (4) with $C=S=0$, and B^α is the same except for different real parameters. Therefore, Eqs. (19) yield a redundant Killing tensor, viz. the one which is characteristic of all spherically symmetric static space–times.

Next, suppose that $f^{31}=f^{23}=0$, but that f^{33} is not a constant and is given by Eq. (21). Then Eqs. (24) are satisfied if and only if

$$\exp(-\nu)=a_0+b_0s^2, \tag{29}$$

$$f^{44}=a_0(f^{33}-\gamma^{33})+\gamma^{44}, \tag{30}$$

where a_0 and b_0 are constants. Such is necessarily the case if the rank of the radial functions R_A is four and Eq. (20) holds. Also, such can be the case if Eqs. (22) hold, with $n=1$.

Next, suppose that f^{33} is constant and that f^{31} and f^{23} are not both zero and are given by Eqs. (23). Then

$$\exp(-\nu)=a_0+2ms^{-1}, \tag{31}$$

$$\frac{\partial f^{44}}{\partial\theta}=2mf^{31}, \quad \frac{\partial f^{44}}{\partial\phi}=2m(\sin^2\theta)f^{23}, \tag{32}$$

where a_0 and m are constants. Such is necessarily the case if Eqs. (22) hold with $n=2$. Also, such can be the case if Eqs. (22) hold with $n=1$.

Next, suppose that f^{33} is not a constant and, also, f^{31} and f^{23} are not both equal to zero. Such can be the case only if Eqs. (22) hold with $n=1$. From the linear independence of f^{33} and f^{31} (or f^{23}), we can show that this occurs if and only if (with the usual choice of the scale of the coordinate t)

$$\exp\nu=1, \quad f^{44}=f^{33}-\gamma^{33}+\gamma^{44}. \tag{33}$$

This case turns out to be the Einstein static model and yields only a redundant Killing tensor, which will be given in the next section.

We have now exhausted all cases. There remain the computation of $\exp(-\lambda)$ from Eqs. (20) and (22) and of

the relation between s and r from Eq. (11). We skip further calculational details and will now summarize the results for the line elements and for the corresponding constants of geodesic motion $K^{\alpha\beta}p_\alpha p_\beta$.

4. RESULTS

For all line elements which have the forms specified by Eqs. (2) or (3) and which yield nonredundant stationary Killing tensors, $\exp\nu$ is given by Eqs. (29) or (31), and

$$\exp(-\lambda)=(n^2-\kappa^2 s^4)^2\,(r/s)^2, \tag{34}$$

$$(s/r)^2=n^2+bs^2+\kappa^2 s^4;\quad n=1,\tfrac{1}{2},2. \tag{35}$$

Above, b and κ are constants; κ is arbitrary if Eq. (29) holds and is zero if Eq. (31) holds. We choose $\kappa\geq 0$ without loss of generality. The arbitrary sign in s has been chosen so that $s>0$.

Equation (35) can be solved for r as a function of s, or vice versa. When $\kappa=0$ and $b>0$, note that r is a monotonic increasing function of s defined over the intervals,

$$0<r<1/\sqrt{b},\quad 0<s<\infty.$$

When $\kappa=0$ and $b<0$, again r is a monotonic increasing function of s, but this time over the intervals,

$$0<r<\infty,\quad 0<s<n/\sqrt{-b}.$$

When $\kappa>0$, the coordinates θ, ϕ, s cover a larger portion of space–time than the coordinates θ, ϕ, r. For example, if $\kappa>0$ and $b>-2n\kappa$ than the metric tensor components r^2, $r^2\sin^2\theta$, and $(r/s)^4$ in the line element (2) are rational functions of s^2 over the domain $0<s^2<\infty$. On the other hand, $\exp(-\lambda)$ as a function of r has two distinct branches obtained by replacing s^2 in Eq. (34) by the two roots $s^2=(2\kappa^2)^{-1}\{r^{-2}-b\pm[(r^{-2}-b)^2-4n^2\kappa^2]^{1/2}\}$, derived from Eq. (35). In one branch, r is a monotonic increasing function of s^2 over the intervals,

$$0<s^2<n/\kappa,\quad 0<r<(b+2n\kappa)^{-1/2}.$$

In the other branch, r decreases monotonically with s over the intervals

$$n/\kappa<s^2<\infty,\quad (b+2n\kappa)^{-1/2}>r>0.$$

All line elements which yield nonredundant Killing tensors are given in Table II, together with the corresponding quadratic constants of geodesic motion. The dynamical variables p_4, l_{ij}, u_i, v_i, and w_i which appear in the table are defined by Eqs. (9), (10), and (14) respectively; l^2 is the square of the angular momentum:

$$l^2=\tfrac{1}{2}l_{ij}l_{ij}. \tag{36}$$

The constants of geodesic motion H_{ij} and H_i are derived from Eqs. (17), (21), and (30); A_i and A_{ij} are derived from Eqs. (17) and (23). (Actually, we obtained the constants of geodesic motion in the forms $\epsilon_{ij}H_{ij}$, ϵ_iH_i. ϵ_iA_i, and $\epsilon_{ij}A_{ij}$ where $\epsilon_{ij}=\epsilon_{ji}$ and ϵ_i are arbitrary real parameters; these parameters are omitted in Table II). It is interesting to note that the traces of H_{ij} and A_{ij} are redundant as can easily be verified.

Table II covers all spherically symmetric static space–times which yield nonredundant stationary Killing tensors. There are two and only two line elements which are special cases of those listed in Table II and which yield only redundant Killing tensors. One of these is the Einstein static model line element which corresponds to the following values of the parameters used in the table:

$$n=1,\quad \kappa=b_0=0,\quad a_0=1.$$

For these values, the two $n=1$ line elements of Table II become identical in form, and the corresponding constants of geodesic motion reduce to

$$H_{ij}=u_iu_j,\quad A_i=l_{ij}u_j,$$

which are redundant as can be seen by referring to Table I.

The other redundant Killing tensor corresponds to the values $n=1$, $\kappa=0$, and

$$\exp(-\nu)=\exp\lambda=1+bs^2.$$

This is the de Sitter line element. From Table II we obtain

$$H_{ij}=v_iv_j-w_iw_j, \tag{37}$$

whose redundancy can be seen from Table I. It is interesting to note here how a stationary Killing tensor is constructed from the nonstationary Killing vectors peculiar to the de Sitter line element.

We next consider the Cartan components of the Einstein tensor relative to the orthonormal tetrad forms

$$e_\theta=r\,d\theta,\quad e_\phi=r\sin\theta\,d\phi,$$
$$e_s=(r/s)^2ds,\quad e_t=-(\exp\tfrac{1}{2}\nu)dt. \tag{38}$$

The Einstein tensor is related to the matter tensor and to a cosmological constant Λ by the equation

$$G_{\alpha\beta}=-kT_{\alpha\beta}+\Lambda g_{\alpha\beta}. \tag{39}$$

Those Cartan components of $kT_{\alpha\beta}-\Lambda g_{\alpha\beta}$ which are not identically zero are given below in forms which are readily applicable to any of our line elements of Table II:

$$kT_{\theta\theta}-\Lambda=kT_{\phi\phi}-\Lambda=-2\kappa^4s^6-3b\kappa^2s^4-6n^2\kappa^2s^2-n^2b-(n^2+bs^2+\kappa^2s^4)(n^2+2bs^2+3\kappa^2s^4)\times\left(\frac{e^\nu}{2s}\frac{de^{-\nu}}{ds}\right),$$

TABLE II. Non-redundant quadratic constants of geodesic motion $K^{\alpha\beta}p_\alpha p_\beta=K_{\alpha\beta}\dot{\xi}^\alpha\dot{\xi}^\beta$. The line elements are all of the form $ds^2=r^2(d\theta^2+\sin^2\theta d\phi^2)+(r/s)^4ds^2-(\exp\nu)dt^2$ where $(s/r)^2=\kappa^2s^4+bs^2+n^2$.

$\exp(-\nu)$	n	κ	$K^{\alpha\beta}p_\alpha p_\beta$
$a_0+b_0s^2$	1	arbitrary constant	$H_{ij}=u_iu_j+(\kappa^2l^2-b_0p_4^2)s_is_j$ $=v_iv_j-w_iw_j+\kappa^2l^2s_is_j$
$a_0+b_0s^2$	$\frac{1}{2}$	arbitrary constant	$H_i=\left(H_{ij}-\frac{l^2}{4s^2}\delta_{ij}\right)\left(\frac{x_j}{r}\right)$
a_0+2ms^{-1}	1	0	$A_i=l_{ij}u_j-mp_4^2\,(x_i/r)$
a_0+2ms^{-1}	2	0	$A_{ij}=(u_il_{jk}+u_jl_{ik})(x_k/r)$ $+\left(\frac{2l^2}{s}-mp_4^2\right)\left(\frac{x_ix_j}{r^2}\right)$

$$-(n^2+bs^2+\kappa^2s^4)^2\left[\frac{e^\nu}{2}\frac{d^2e^{-\nu}}{ds^2}-\tfrac{3}{4}\left(e^\nu\frac{de^{-\nu}}{ds}\right)^2\right], \tag{40a}$$

$$kT_{rr}-\Lambda=\kappa^4s^6-(1+2n^2)\kappa^2s^2-b-n^2(1-n^2)s^{-2} -(n^2-\kappa^2s^4)\,(n^2+bs^2+\kappa^2s^4)\left(\frac{e^\nu}{s}\frac{de^{-\nu}}{ds}\right), \tag{40b}$$

$$kT_{tt}+\Lambda=3\kappa^4s^6+6b\kappa^2s^4+(1+14n^2)\kappa^2s^2+(1+2n^2)b +n^2(1-n^2)s^{-2}. \tag{40c}$$

The Weyl conform tensor was also computed and we found that conformal flatness is obtained for the following three sets of parameter values corresponding to $n=\frac{1}{2}, 1$, and 2 respectively: $n=\frac{1}{2}, a_0=0$; $n=1, b_0^2-a_0b_0b+a_0^2\kappa^2=0$, $a_0\neq0$, $b^2>4\kappa^2$; $n=2$, $a_0^2+m^2b=0$. The $n=1$ set of conformally flat space–times specializes to the de Sitter or Einstein static model space–times when $\kappa=0$.

We now initiate our discussion of results with a summary of the main features of the matter tensor given by Eqs. (40).

5. DISCUSSION

Some features of the above matter tensor are evident by inspection. For example, each Cartan component is a rational function of s. Also, there is no difficulty about satisfying the condition $T_{tt}\geq0$ if $n=1$ or $\frac{1}{2}$. If $n=2$, however, then the condition is violated in some neighborhood of zero s.

As regards singularities, each Cartan component has poles at $s=\infty$ if and only if $\kappa\neq0$. When $\kappa\neq0$ and we transform to Schwarzschild coordinates in a neighborhood of infinite s, then $s=\infty$ maps into $r=0$, and the dominant term near this origin is proportional to r^{-6}.

Except when $n=1$, $\exp(-\nu)=a_0=b_0s^2$, and $a_0\neq0$, there are singularities at $s=0$ in one or more of the Cartan components of the matter tensor. If $n\neq1$, there are poles of order two in the energy density and in the radial stress. If $\exp(-\nu)=a_0+2ms^{-1}$ $(m\neq0)$ or $\exp(-\nu)=b_0s^2$, there are poles of order two in both stress components. All of these second order poles are aphysical in the sense that their strengths are independent of the nonzero values assigned to the parameters in the metric tensor.

We next consider the various types of matter which might conceivably fit Eqs. (40). If the matter tensor is due only to a perfect fluid or to a vacuum, then $T_{rr}-T_{\theta\theta}$ must vanish identically. As expected, this perfect fluid condition is satisfied if and only if $n=1$, $\kappa=0$, and either

$$\exp(-\nu)=a_0$$

or

$$\exp(-\nu)=a_0(1+bs^2),\quad b\neq0.$$

The first of the above alternatives is the Minkowski space or the Einstein static model depending on whether b is zero or not zero. The second alternative is the de Sitter space. As we have already noted in the preceding section, the Killing tensors corresponding to these models are redundant.

Our nonredundant Killing tensors do not correspond to any simple physical source as far as we can tell. We are referring here to a perfect fluid taken alone or to any classical vector or scalar field taken one at a time.

We have also tried some combinations of pairs of sources without great success. For some of our line elements, the matter tensor is consistent over a limited range of s with an admixture of a perfect fluid with a single classical field such as an electrostatic or a scalar field. By "consistency", we mean that the conditions of positive definiteness for the pressure p, the proper mass density ρ, and even ρc^2-3p can be satisfied. However, it appears to us that such combinations divorced as they are from any likely physical model are too eclectic to deserve more than passing mention.

It seems worthwhile to describe some features of the matter tensor of those special metrics for which $n=1$, $\kappa=0$, $a_0>0$, and b_0, m, and b are nonnegative. With the usual choice of the time scale, $a_0=1$, and

$$\exp(-\nu)=1+b_0s^2,\quad r^2=s^2(1+bs^2)^{-1}, \tag{41}$$

or

$$\exp(-\nu)=1+2ms^{-1},\quad r^2=s^2(1+bs^2)^{-1}. \tag{42}$$

These selected metrics include the Einstein and de Sitter models as special cases. They each correspond to a uniform energy density $T_{tt}=k^{-1}(3b-\Lambda)$. There are no singularities except for the second order pole at the origin corresponding to the metric in Eq. (42). As regards the perfect fluid condition, $T_{rr}-T_{\theta\theta}=0$, the metric (41) satisfies it at $s=0$, and the metric (42) satisfies it on the sphere $s=(R/2m)\,\{R^2+[R^2+(2m)^2]\}^{1/2}$, where $R=b^{-1/2}$. If any of our line elements are candidates for some application, we would bet on Eqs. (41) and (42). However, we have found nothing as yet.

It appears as if the physical pertinence of the new line elements which we have obtained does not derive from any coincidence between them and some solution corresponding to an obvious interesting physical model. However, like many of the central potentials which were studied in the heyday of Newtonian mechanics, our line elements may serve didactic purposes. For example, the geodesic orbits have interesting properties analogous to Kepler's laws. This is illustrated by the orbital equation for the metric given by Eq. (42). To obtain this orbital equation, choose the angle θ so that the orbit lies in the plane $\theta=\frac{1}{2}\pi$; then compute x_iA_i/r where A_i is the generalized Lenz vector which is defined in Table II. Another illustration, this time reminiscent of harmonic oscillator orbits, is derived from the quadratic constant of the motion H_{ij}. Note the absence of precession for both illustrations.

Apart from such didactic purposes, there are the techniques which we have used and which can serve as a guide for further investigations, especially in connec-

tion with axially symmetric problems. Useful results may be found as suggested by the Carter[5] work on the Kerr metric.

With respect to this last remark, it is pertinent to note the way some of the Killing tensors which we found become redundant when we specialize the line elements to those of the Einstein and de Sitter models. Analogously, we expect some of the axially symmetric stationary space–times to have Killing tensors which reduce to the redundant cases of Eqs. (19) and (28) in the special case of spherical symmetry. The quadratic constant of the motion[5] peculiar to the Kerr metric illustrates this kind of reduction to redundancy.

The methods which we have used in this paper differ markedly from those used in most previous investigations[3,8,9] of quadratic first integrals of geodesic motion. The starting point of these previous calculations has been to choose an orthonormal ennuple relative to which the matrix $\{K^{\beta}_{\alpha}\}$ has a Jordan canonical form. However, the coordinate systems appropriate to these ennuples may well differ from those which display some given group of isometries to advantage. So, if we are interested in space–times which are spherically symmetric or which are axially symmetric and stationary, it seems more appropriate to us to choose a coordinate system in which the given isometries are manifest. This should facilitate both the calculations and the later checks for redundancy.

In this paper, we have drawn a distinction only between those Killing tensors which are redundant and those which are not redundant. There are other ways of characterizing Killing tensors. For example, we may consider those symmetric tensors $K_{\alpha\beta}$ which satisfy conditions of the form

$$K_{\alpha\beta;\gamma}=A_{\alpha\gamma}B_{\beta}+A_{\beta\gamma}B_{\alpha}-2A_{\alpha\beta}B_{\gamma}. \tag{43}$$

It is easy to show that Eq. (43) implies that $K_{\alpha\beta}$ is a Killing tensor. In particular, if $A_{\alpha\beta}=K_{\alpha\beta}$ and if $K_{\alpha\beta}$ is nonsingular and has the same signature as $g_{\alpha\beta}$, then Eq. (43) is equivalent to the statement that the space–time admits a geodesic correspondence[6,8] onto another space–time whose metric is

$$\bar{g}_{\alpha\beta}=(g/K)^{2/3}K_{\alpha\beta}, \tag{44}$$

where g and K are the determinants of $g_{\alpha\beta}$ and $K_{\alpha\beta}$ respectively. On the other hand, if $A_{\alpha\beta}=g_{\alpha\beta}$, and $K_{\alpha\beta}$ has the structure

$$K_{\alpha\beta}=K_{\alpha;\beta}+K_{\beta;\alpha}-\tfrac{4}{5}g_{\alpha\beta}K^{\mu}{}_{;\mu}, \tag{45}$$

then Eq. (43) is equivalent to the statement that the infinitesimal mapping $x^{\alpha}\to x^{\alpha}+K^{\alpha}\,\delta\epsilon$ is a projective collineation.[6,8]

As regards the vector field B_{γ} in Eq. (43), if $A_{\alpha\beta}$ is nonsingular, then B_{γ} is uniquely determined by $K_{\alpha\beta}$ and $A_{\alpha\beta}$. For a geodesic correspondence,

$$B_{\gamma}=-(1/6K)K_{,\gamma}; \tag{46}$$

for a projective collineation,

$$B_{\gamma}=\tfrac{1}{5}(K^{\alpha}{}_{;\alpha})_{,\gamma}. \tag{47}$$

Note, therefore, that the defining equations for a geodesic correspondence and a projective collineation are inherently nonlinear.

If the Killing tensors of a given space–time have all been worked out as is true in our case, then a residual problem is to discover which of these Killing tensors are characteristic of geodesic correspondences or projective collineations. Preliminary calculations have turned up negative results so far for all except redundant Killing tensors, but the work remains to be completed before definitive statments can be made. The test for a geodesic correspondence or a projective collineation cannot be done on one Killing tensor at a time. Instead, it is necessary to construct a linear combination of *all* Killing tensors of the given line element as follows:

$$c_1K^{(1)}_{\alpha\beta}+\cdots+c_NK^{(N)}_{\alpha\beta}.$$

Then this linear combination is to be inserted into Eq. (43) and the values of the coefficients $c_1, \ldots, c_N$ are to be determined. The final result must be checked to see if it is nonsingular and has the same signature as $g_{\alpha\beta}$.

As we mentioned in the introduction, V.I. Golikov[7] and C. F. Martin[8] have found all space–times which admit geodesic correspondences.[12] Martin[8] has also specialized some of these space–times to ones admitting projective collineations. Their results deserve further investigation. In particular, some of them may contain interesting axially symmetric stationary line elements as special cases.

The authors are currently working on the problem of axially symmetric stationary line elements which admit nontrivial quadratic first integrals of geodesic motion. This Killing tensor criterion is an alternative to that of separation of variables.[10] It is still not clear whether it can function as an independent and useful constraint on solutions corresponding to interesting matter tensors. In fact, the whole problem of hidden symmetries in general relativity can still bear a lot of investigation.

ACKNOWLEDGMENTS

We thank Dr. Frederick J. Ernst and Dr. Richard A. Isaacson for their encouragement and many helpful discussions.

[1]This definition differs from that given by J. L. Synge, in *Relativity: The General Theory* (North-Holland, Amsterdam, 1960). He uses the term Killing tensor for $\xi_{\alpha;\beta}$ where ξ_{α} is any Killing vector.

[2]There is no reason, other than expediency, for this restriction. We hope to return to the problem of solving the much more complicated equations for time-dependent Killing tensors.

[3]See, for example, L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N. J., 1926, 1960 printing), p. 107 et seq.

[4]Some aspects of geodesic correspondences for spherically symmetric space–times have been analyzed by H. Takeno, *The Theory of Spherically Symmetric Space–Times*, Revised edition, Sci. Rep. Res. Inst. Theor. Phys. Hiroshima University, No. 5 (1966). This report includes references to other work by this author on spherically symmetric space–times.

[5]A special Killing tensor peculiar to the Kerr metric has been found by B. Carter, Phys. Rev. **174,** 1559 (1968).
[6]A detailed account, with references, of work on Killing tensors and geodesic correspondences, prior to 1926 is given by L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N. J., 1926, 1960 printing), pp. 128–141. Also, Appendix 13 of this reference deals with the separable systems of Stäckel and includes some later original contributions of Eisenhart.
[7]V. I. Golikov, Uch. Zap., Kazan State Univ. 123, book 12, 26, 59 (1963). Also, see Ref. 9, p. 278.
[8]C. F. Martin, Ph.D. Thesis, North Carolina State College, 1962 (unpublished). This reference also has some new results on space–times admitting projective collineations.
[9]A detailed account, with references of work on geodesic correspondences and projective collineations prior to 1962, is given by A. Z. Petrov, *Einstein Spaces,* English Edition (Pergamon, Oxford, 1969).
[10]B. Carter, Commun. Math. Phys. **10,** 280 (1968).
[11]An interesting example of a particular relation is given in Appendix 13 of Ref. 6.
[12]A family (not the most general) of spherically symmetric space–times which admit geodesic correspondences is given by H. Takeno, *The Theory of Spherically Symmetric Space–Times*, Sci. Rep. Res. Inst. Theoret. Phys., Hiroshima University, No. 5 (1966), p. 273. We find that the static space–times in this family all correspond to redundant Killing tensors.

Causality and nonlocalizable fields

F. Constantinescu
Department of Applied Mathematics, University of Frankfurt, Frankfurt, Germany
J. G. Taylor
Department of Mathematics, King's College, London, England

(Received 20 September 1972)

We give a model-independent and Lorentz invariant prescription for the manner in which a nonlocalizable quantum field extends outside the light cone, in particular specifying the order of this extension. We show how our definition applies to several examples, including the nonlocalizable free field and certain functions of the massive free field; nonlocalizable functions of the massless free field presents a difficulty which has not yet been resolved.

1. INTRODUCTION

There has been much recent discussion of nonpolynomial functions of the free field, both with respect to mathematical problems arising in setting up such a functional calculus[1] and also with respect to the application of such a calculus to nonpolynomial chiral and gravitational theories.[2] One of the most important questions involved in such problems is that of the properties that such nonpolynomial functions possess, especially causality and positive-definiteness of the metric in the state space.[3] In this paper we attempt an analysis of causality, in particular investigating the manner in which it is broken by nonlocalizable functions of the free field.

These nonlocalizable functions have physical interest since they can arise when particular field coordinates are chosen in either the chiral or gravitational interactions. A class of these functions has been investigated elsewhere and shown[4] to possess various useful properties, such as the existence of a PCT operator and of a scattering theory, as the vestiges of local commutativity. However, such nonlocalizable fields cannot satisfy strict causality, and a description of the extension of the commutator bracket outside the light cone has been given for the zero mass case.[5] We wish here to give a general discussion of this extension of the commutator bracket of a nonlocalizable field outside the light cone, both from a mathematical and a physical point of view. This necessitates the introduction of a new class of test functions which can suitably probe the behavior of the commutator bracket both outside and inside the light cone. This allows us to indicate, in a Lorentz invariant fashion, both the rate of decrease of the commutator outside the light cone as well as the range of the extension. In other words, we specify the degree of noncausality in an invariant fashion, and expect the range defined in this way to have physical interest and so determine the energy at which violations of causality can appear.

The detailed plan of the paper is as follows. In the next section we give a general discussion of the way the extension of the commutator bracket can be described in a Lorentz invariant and mathematically precise fashion. We give a definition which classifies the degree of nonlocalizability of nonlocalizable fields. In Sec. 3 we show how functions of the zero mass free field nearly satisfy conditions for this definition to apply, and consider a special class of nonzero mass fields in the subsequent section for which our definition can be used. A more complete discussion of the nonzero mass case is given in Sec. 5. The physical significance of the discussion is considered in the final section.

2. NONLOCALIZABLE FIELDS

In this section we wish to formulate the properties of nonlocalizable fields in a manner independent of the fact that the most interesting applications are to functions of a free field. We will consider only one neutral scalar field $\phi(x)$, the extension to complex spin fields presenting no conceptual difficulty. We suppose that the field $\phi(x)$ is an operator-valued generalized function over a suitable test function space C of functions $f(x)$, so that $\phi(f)=\int\phi(x)f(x)\,dx$ is an operator for which the usual Wightman axioms,[6] except that of local commutativity, apply. Thus there are a sequence of Wightman functions $W_n(x_1,\ldots,x_n)$, defined by

$$W_n(x_1,\ldots,x_n)=\langle 0\,|\,\phi(x_1)\cdots(x_n)\,|\,0\rangle$$

from which the fields may be reconstructed.

The usual tempered field theory results if C is the space S of indefinitely differentiable functions decreasing at infinity faster than any polynomial; the localizable case occurs if C contains a dense subset of functions of compact support. The nonlocalizable situation in which we are interested in here corresponds to C being comprised of analytic functions due to their rapid decrease in momentum space. This fall-off is required in order that the rapid high energy increase of the Fourier transforms $\overline{W}_n$ of the Wightman functions can be satisfactorily taken account of by the Fourier transforms of the functions of C, which we denote by $\mathcal{M}$. If $\overline{W}_n$ increases like $\exp[|p|^{\alpha}]$, for $\alpha<1$, a satisfactory choice for $\mathcal{M}$ is any space S^{β}, for $\beta<\alpha$ of Gelfand and Shilov,[7] being the set of indefinitely differentiable functions, which together with all derivations, are bounded at infinity by $\exp[-b|p|^{1/\beta}]$ for some positive b (which depends on the function considered). A more precise specification[8] is given by the indicatrix function $g(p^2)$ which is an entire function satisfying

$$\int_0^{\infty}[\log g(t^2)/t^2]\,dt<\infty. \tag{1}$$

The related function space, which we denote C^g in momentum space or C_g in coordinate space, is composed of momentum space functions $\phi(p)$ which, together with all derivatives, are bounded by $g^{-1}(Ap^2)$ for some constant A. If (1) is not satisfied the space C_g is composed

of analytic functions, so again corresponds to the nonlocalizable case.

We suppose, then, that each Wightman function W_n belongs to the dual of one of the above spaces $C(\mathbf{R}^{4(n-1)})$ in $4(n-1)$ variables. As usual we denote this by

$$W_n \in \cdot C'(\mathbf{R}^{4(n-1)}). \tag{2}$$

It is not possible, in the nonlocalizable case, even to formulate local commutativity. Analysis of special cases[5] has shown that there may still be a trace of local commutativity in that commutator brackets of the field operators at two points, whilst not zero for spacelike separations, can decrease fast as the spacelike separation increases. In fact, we will be led to consider an exponential fall-off. If we denote by $\phi(x)$ the nonlocalizable field of interest, then we expect[5] for large spacelike values of $(x-y)$, that

$$|<0|[\phi(x),\phi(y)]_-|0>| \leq \exp[-a|(x-y)^2|^{\gamma/2} \tag{3}$$

for some positive constant a and γ, wtih $\gamma<1$. We note that (3) does not contradict the result of Pohlmeyer and Borchers[9] that such a fast fall-off implies strict local commutativity, since this result was obtained on the basis of analyticity of Wightman functions in space–time; such properties no longer persist in the nonlocalizable case.

We will now discuss how we may formulate (3) in a general fashion. The idea behind our approach is to determine how far outside C we may extend the commutator bracket (3). If it is to test functions which actually increase for large spacelike values like the inverse of the rhs of (3) then we can conclude that the decrease of the commutator is roughly given by the rhs of (3). This extension has to be investigated in detail for each particular field being considered, but we will attempt in this section to give a general formulation of it which is model independent. To do that let us take the case when C is required to be an S^α space, with $\alpha<1$. We consider the commutator brackets

$$C_{n,j}(x_1\cdots x_n)=W(x_1\cdots x_j,x_{j+1}\cdots x_n)-W(x_1\cdots x_{j+1}x_j\cdots x_n) \tag{4}$$

If ϕ were local then causality would indicate that each C_j vanishes for $\xi_j=x_j-x_{j+1}$ spacelike, but would be nonzero for timelike separations. Thus we have to choose test functions which are in S^α for timelike values; we denote this space of test functions by S_x^α. Since the set $X=\{\xi;\xi^2\geq 0\}$ is a closed set we have to define S_X^α by a suitable limiting procedure. It may be possible to use the closed set X directly, without using the following construction, but there are various points which need to be resolved before that can be done. We will not consider that further here but use the better known inductive limit approach.

Let $S^{\alpha B}(\Omega_\mu)$ be the set of functions ϕ of the 4-vector x which are defined on the set $\Omega_\mu=\{x;x^2>-\mu^2\}$, and satisfy there

$$|\mathbf{x}^k\phi^{(q)}(x)| \leq C_k B^{|q|}q^{q\alpha}, \tag{5}$$

where

$$\mathbf{x}^k=\prod_{i=1}^4 x_i^{k_i},\quad |q|=q_1+q_2+q_3+q_4,\quad q^{q\alpha}=\prod_{i=1}^4 q_i^{q_i\alpha},$$

and B, C_k are given constants. In order to take account of the constant a on the rhs of (3), we should take B to be independent of ϕ in (5). However, that does not give a Lorentz covariant space, but initially we allow only the constants C_k to vary with ϕ. We now define

$$S_x^\alpha=\bigcup_{\substack{\mu>0\\ B=1,2,\ldots}} S^{\alpha B}(\Omega_\mu). \tag{6}$$

Each space $S^{\alpha B}(\Omega_\mu)$ is a complete, countably normed perfect space with the norms

$$\|\phi\|_{k,\rho,\mu}\doteq \sup_{\substack{x\in\Omega_\mu\\ q}} |\mathbf{x}^k\phi^{|q|}(x)|/[(B+\rho)^{|q|}q^{q\alpha}],$$

for $k=0,1,2,\ldots,\rho=1,\frac{1}{2},\ldots$, being a standard $S^{\alpha B}$ space but on the open set Ω_μ.[10] Since $S^{\alpha B_1}(\Omega_\mu)\subseteq S^{\alpha B_2}(\Omega_\nu)$ for $\mu\geq\nu$ and $B_2\geq B_1$ with topological inclusion, then we may take the topology on S_x^α as the inductive limit as μ tends to zero and B to infinity of these topologies on the $S^{\alpha B}(\Omega_\mu)$, so that a sequence $\{\phi_n\}$ converges to zero in S_x^α if all the functions ϕ_n belong to some space $S^{\alpha B}(\Omega_\mu)$ for some μ and B, and converge to zero in its topology. We note that if $\alpha>1$ the space $S^{\alpha B}(\Omega_\mu)$ is not complete, and the quantity $\|\phi\|_{k,\rho,\mu}$ is not a true norm but only a prenorm. However, we do not wish to apply our construction of $S^{\alpha B}(\Omega_\mu)$ and S_x^α to the localizable case.

We construct the space of test functions appropriate to the commutator $C_{n,j}$ as

$$T_{n,j}(\alpha)=S^\alpha(\mathbf{R}^{4(n-2)})\otimes S_x^\alpha, \tag{7}$$

where the functions $\phi(\xi_1,\ldots,\xi_{n-1})$ in $T_{n,j}(\alpha)$ are in $S^\alpha(\mathbf{R}^{4(n-2)})$ with respect to the variables $\xi_i(=x_i-x_{i+1})$ for $i=1,2,\ldots,j-1,j+1,\ldots,n-1$ and are in S_x^α in ξ_j. At this stage it is not evident how this space $T_{n,j}(\alpha)$ is relevant to the exponential decrease given by the rhs of (3). This will become apparent from the following lemma.

Lemma 1: Any function $\phi\in S_x^\alpha$ has an analytic continuation to the whole of $\mathbf{R}^4+i\mathbf{R}^4=\mathbf{C}^4$, and increases at infinity at most like $\exp 4(B\|x\|)^{1/(1+\alpha)}$ where $\|x\|=\sup_{1\leq i\leq 4}|x_i|$, for some constant B.

Proof: To prove this we note that $\phi\in S^{\alpha B}(\Omega_\mu)$, for some B and μ. Thus by (5)

$$|\phi^{(q)}(0)| \leq C_0 B^{|q|}q^{q\alpha}. \tag{8}$$

Thus the series

$$\psi(x)=\sum_q \mathbf{x}^q\phi^{(q)}(0)/q_{q!},$$

where

$$\mathbf{q}!=\prod_{i=1}^4 q_i!,\ \mathbf{x}^q=\prod_{i=1}^4 x_i^{q_i},\quad \phi^{(q)}(0)=\prod_{i=1}^4\frac{\partial^{q_i}}{\partial x_i^{q_i}}\ \phi\Big|_{x=0}$$

has infinite radius of convergence and is bounded at infinity, by (8), by

$$C_0\sum_q B^{|q|}q^{q\alpha}|\mathbf{x}|^q/_{q!} \leq C_0\exp[4B\|x\|^{1/(1-\alpha)}], \tag{9}$$

where

$$|\mathbf{x}|^q=\prod_{i=1}^4 |x_i|^{q_i}.$$

Since ψ coincides with the analytic function ϕ in Ω_μ, then the lemma is proved. We see that if we can extend $C_{n,j}$ from $S^\alpha(\mathbf{R}^{4(n-1)})$ to $T_{n,j}(\beta)$ then we obtain the decrease

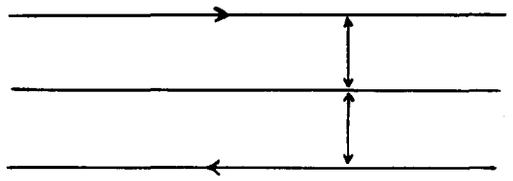


FIG. 1. The integration contour C in the complex x_0 plane over which the massless commutator $\langle C_2, \phi\rangle$ of Eq. (16) is evaluated.

like the rhs of (3) with $\gamma = 1-\beta$ and $a = 4(B)^{1/(1-\beta)}$. Thus we define the degree of extension of the commutator bracket outside the light cone as follows.

Definition 1: If for each n and each j we may extend $C_{n,j}$ from $S^\alpha(\mathbf{R}^{4(n-1)})$ to $T_{n,j}(\beta)$, where β may be chosen independently of n and j, then the largest such values of $\gamma = 1-\beta$ will be called the order of extension of the commutator bracket outside the light cone.

We note that this definition is obviously Lorentz invariant, the space S_x^α evidently being so.

We extend this definition to the finer specification given by an indicatrix function $g(p^2)$. Let

$$g(p^2) = \sum_{r=0}^{\infty} c_{2r} p^{2r}.$$

Then the space $C^{Bg}(\mathbf{R}^4)$ is defined as

$$C^{Bg}(\mathbf{R}^4) = [\phi : |\mathbf{x}^k \phi^{(q)}(x)| \leq C_k B^{|q|} c_q^{-q}]. \tag{10}$$

We modify this space to $C^g(\Omega_\mu)$ as before, and define

$$C_x^g = \bigcup_{\substack{\mu>0 \\ B<\infty}} C^{Bg}(\Omega_\mu).$$

Similarly we define the space $T_{n,j}(g) = C^g(\mathbf{R}^{4(n-2)}) \otimes C_x^g$, and extend the Definition 1 to

Definition 2: If for each n and j we may extend $C_{n,j}$ from $C^g(\mathbf{R}^{4(n-1)})$ to $T_{n,j}(g)$, where g may be chosen independently of n and j, then the smallest g which may be chosen defines the nature of decrease of the extension of the commutator outside the light cone.

To see this nature of decrease in detail, we extend Lemma 1 to

Lemma 2: Any function $\phi \in C_x^g$ has an analytic continuation to the whole of $\mathbf{C}^4$ and increases at infinity at most like $\Pi_{i=1}^4 G(B|x_i|)$, when

$$G(x) = \sum_{n\geq 0} c_n^{-n}[(Bx)^n/n!]$$

has an infinite radius of convergence, and B depends on ϕ.

The proof of this lemma follows that of Lemma 1 almost identically, where we assume that g does not satisfy (1), so we are dealing with the nonlocalizable situation. It need not be the case that any $\phi(x) \in C_x^g$ has an *analytic* continuation to all $\mathbf{C}^4$ in general, since ϕ need at most be quasianalytic in Ω_μ for some μ. That is why we need to impose the condition on the infinite radius of convergence of the series $G(x)$. However, Definition 2 may be used even if it is only quasianalytic but the extension process can only be performed infinitesimally outside the light cone. It does not seem possible to specify the rate of decrease of the commutator in this more general situation. As examples we note that if $c_n = n^{-\alpha}$ we return to the S^α spaces, whilst if $c_n = n^{-1}\log n$ it has an infinite radius of convergence. However, if $c_n = n^{-1}(\log n)^{-1}$ the function $G(x)$ has zero radius of convergence, so that analytic continuation of any $\phi \in C_x^g$ need not be possible to arbitrary spacelike points.

We can enlarge the above approach to include the *range* of decrease of the commutator bracket outside the light cone if we consider the various Wightman functions in terms of their invariant variables; the discussion is so similar to the above, except for replacement of 4-vector variables by invariants, that we need not give that discussion here.

3. FUNCTIONS OF THE MASSLESS FREE FIELD

Let us turn now to specific examples to indicate how the above formulation of noncausality actually applies. We consider in this section an infinite series in normal-ordered powers of the free massless scalar field

$$\phi(x) = \sum_{n=0} (d_n/n!) :\phi_0^n(x):, \tag{11}$$

where : : denotes the normal ordering and the d_n are real coefficients. Then in the notation of Sec. 4 of Ref. 5,

$$C_{n,k}(x_1, \ldots, x_n) = \sum_{r_{ij}=0} \prod_{\substack{i=1 \\ i\neq k,k+1}}^{r} d_{R_i} \; T^{R/r_{k,k+1}} (R!/r_{k,k+1}!)^{-1}$$
$$\times 2\pi i \epsilon(x_0)_{r_{k,k+1}} \sum_{=0}^{\infty} \frac{d_{R_k} d_{R_{k+1}} - 1)^{r_{k,k+1}} \delta^{(r_{k,k+1}-1)}(\xi_k^2)}{r_{k,k+1}!\,(r_{k,k+1}-1)!\,(4\pi^2)^{r_{k,k+1}}} \tag{12}$$

where

$$T^{R/r_{k,k+1}} = \prod_{\substack{1\leq i\leq j\leq n \\ (i,j)\neq(k,k+1)}} [i^{-1}\Delta + (x_i - x_j)]^{r_{ij}}. \tag{13}$$

We see from (12) in this case that if

$$d_n^2 = \chi^n \Gamma(1+n\alpha), \tag{14}$$

then $C_{n,k}$ would appear to be in $T_{n,k}(\beta)$ provided $\beta < 2-\alpha$; the nonlocalizable situation corresponds to $\alpha > 1$ so the β can be chosen less than one to satisfy this. However, there is a difficulty in this approach which we will clarify for the particular case of the 2-point, $n=2$. For then, using (14),

$$\begin{aligned} C_2(x_1, x_2) &= \sum_{n\geq 0} (d_n^2/n!)[\Delta_+(x_1-x_2)^n - \Delta_-(x_1-x_2)^n], \\ &= \sum_{n\geq 0} \int_0^\infty [(t^{n\alpha}e^{-t})/n!]\, dt\, [\Delta_+(x_1-x_2)^n - \Delta_-(x_1-x_2)^n], \\ &= \int_0^\infty dt\, e^{-t}\{\exp[t^\alpha \Delta_+(x_1-x_2)] - \exp[t^\alpha \Delta_-(x_1-x_2)]\}. \end{aligned} \tag{15}$$

We may form $\langle C_2, \phi\rangle$, for ϕ analytic, by using that

$$\Delta_\pm(x) = \lim_{\epsilon\to 0}[(x_0 \mp i\epsilon)^2 - \mathbf{r}^2]^{-1},$$

with $x = (x_0, \mathbf{r})$ so that

FIG. 2. The final contour in the x_0 plane to which C of Fig. 1 may be shrunk after taking account of the singularities in $\langle C_2, \phi\rangle$.

$$\langle C_2, \phi\rangle = \lim_{\epsilon\to 0} \int_0^\infty dt\, e^{-t} \int_{C_\epsilon} dx_0 \int d^3r\, e^{t^\alpha/r^2}\, \phi(x), \tag{16}$$

where C_ϵ is a contour in the complex x_0 plane composed of two parts, as shown in Fig. 1.

Thus

$$\langle C_2, \phi\rangle = \lim_{\epsilon\to 0} \int_0^\infty dt\, e^{-t} \int_{C_\epsilon} dx_0 \int d^3\mathbf{r} \sum_{n\geq 0} (t^{n\alpha}/n!)[(x^2)^n]^{-1}\cdot \phi(x)$$

For $|\mathbf{r}| \neq 0$ the poles of $(x^2)^{-n}$ are at $x_0 = \pm|\mathbf{r}|$, and the contour C_ϵ may be shrunk to that of Fig. 2.

Thus we have

$$\langle C_2, \phi\rangle = \int_0^\infty dt\, e^{-t} \sum_n 2\pi i \frac{t^{n\alpha}}{(n!)^2} \int \frac{d^3\mathbf{r}}{|\mathbf{r}|^n} \times \left(\frac{\partial^n \phi}{\partial x_0^n}\bigg|_{x_0=|\mathbf{r}|} + (-1)^n \frac{\partial^n \phi}{\partial x_0^n}\bigg|_{x_0=|\mathbf{r}|} \right). \tag{17}$$

Provided that for all $\mathbf{r} \in \mathbf{R}^3$

$$|\mathbf{r}|^k \left|\partial^n \phi/\partial x_0^n\right|_{x_0=\pm|\mathbf{r}|} \leq C_k n^{n\beta} (B)^n. \tag{18}$$

then

$$|\langle C_2, \phi\rangle| \leq \text{const}\times \int_0^\infty dt\, e^{-t} \sum_n B^n [t^{n\alpha}/n^{n(2-\beta)}] < \infty$$

if $\alpha + \beta < 2$.

This is the same condition which arose above, but now we see the defect arising from both of these approaches. This is that the space integral in (17) is divergent for $n > 2$, owing to the factor $|\mathbf{r}|^{-n}$. Thus the condition (18) or the stronger condition $\phi \in S_x^\beta$ does not lead to a definite value for $\langle C_2, \phi\rangle$. Thus our discussion of functions of the free massless field is deficient at this point. Whilst this is unsatisfactory we will see that a similar situation does not arise in the massive case, so we turn to that now.

4. THE MASSIVE CASE

In this section we will consider two very simple examples of nonlocalizable fields which have been discussed already in the localizable case. The first of these is the generalized free field $\phi(x)$ for which the commutator bracket is

$$[A(x), A(y)]_- = \int_{m_0^2}^\infty d\kappa \rho(\kappa^2) \Delta(x-y, \kappa^2), \tag{19}$$

where $\Delta(x, \kappa^2)$ is the invariant propagator for mass κ. The function $\rho(\kappa^2)$ is allowed to increase for large κ with order of growth at most one for localizability. We can see this by considering the rhs of (19) when applied to a test function ϕ, by means of Fourier transforms, for it takes the value

$$\langle [A(x), A(0)]_-, \phi\rangle = \int_{m_0^2}^\infty d\kappa^2 \rho(\kappa^2) \int d^4x \Delta(x, \kappa^2)\phi(x), = \int_{m_0^2}^\infty d\kappa^2 \rho(\kappa^2) \Delta_\phi(\kappa^2), \tag{20}$$

where the functions Δ_ϕ of κ^2 is defined by

$$\Delta_\phi(\kappa^2) = \int d^4x \phi(x) \Delta(x, \kappa^2), = \int d^4p \tilde{\phi}(p) \delta(p^2 - \kappa^2) \epsilon(p_0), \tag{21}$$

where ϕ is the Fourier transform of ϕ. We now prove

Lemma 3: If $\phi \in S^{\alpha B}$ then $|\Delta_\phi(\kappa^2)| \leq e^{-(B\kappa)^{1/\alpha}}$. (const).

Proof: If

$$\phi \in S^{\alpha B} \Rightarrow \tilde{\phi} \in S_{\alpha B} \Rightarrow |\tilde{\phi}(p)| \leq e^{-(B\|p\|)^{1/\alpha}},$$

where $\|p\|$ is any norm on $\mathbf{R}^4$ consistent with the usual Euclidean topology. Let us take $\|p\|^{1/\alpha} = |p|^{1/\alpha} + |p_0|^{1/\alpha}$. Then

$$\Delta_\phi(\kappa^2) = \int (d^3p/2|p|)[\tilde{\phi}((p^2+\kappa^2)^{1/2}, p) - \tilde{\phi}(-(p^2+\kappa^2)^{1/2}, p)], \tag{22}$$

so

$$|\Delta_\phi(\kappa^2)| \leq \int (d^3p/2|p|) e^{-(B\|p\|)^{1/\alpha}} - [B(p^2+\kappa)^{1/2}]^{1/\alpha} \leq (\text{const}) \times e^{-(B\kappa)^{1/\alpha}}$$

as required. Thus the integral on the rhs of (20) is finite at infinity for all $\phi \in S^{\alpha, B+\epsilon}$ and $\epsilon > 0$, provided that

$$|\rho| \sim e^{(\kappa^2)^{1/2\alpha}} \quad \text{as } \kappa^2 \to \infty. \tag{23}$$

To take account of possible singularities in ρ we have to discuss the differentiability properties of $\Delta_\phi(\kappa^2)$.

Lemma 4: If $\phi \in S^{\alpha,B}(\mathbf{R}^4)$ then $\Delta_\phi(\kappa^2) \in S_{\alpha,B}[(0,\infty)]$, the Gelfand–Shilov $S_{\alpha,B}$-space, but defined on the open interval $(0, \infty)$.

Proof: This follows directly from (22), using the differentiability properties of ϕ:

$$|\Delta_\phi^{(q)}(\kappa^2)| \leq \int (d^3p/2|p|) \left| \left(2\sqrt{p^2+\kappa^2}\right)^{-1} \circ \frac{\partial}{\partial\sqrt{\kappa^2+p^2}} \right)^{(q)} \times [\tilde{\phi}(\sqrt{p^2+\kappa^2}, p) - \tilde{\phi}(-\sqrt{p^2+\kappa^2}, p)] | < \infty,$$

so that $\Delta_\phi(\kappa^2) \in C^\infty((0,\infty))$, the set of indefinitely differentiable functions on the open interval $(0, \infty)$. Including the results of Lemma 3 proves the lemma.

We have proved that if $\rho \in S_{\alpha B}(0,\infty)$ then the commutator brackets (as well as all the Wightman functions, as can easily be seen) are all in the appropriate $S^{\alpha B'}$ spaces of the relevant 4-vector variables.

When we turn to the extension problem, with $\alpha < 1$, we see that the previous discussion using Fourier transforms can no longer be given, since the test functions of S_x^α or S_x^β may increase too rapidly outside the light cone for their Fourier transforms even to be defined. Thus we need to rephrase the preceding discussion purely in terms of coordinate space. We do that in the following lemma.

Lemma 5: If $\phi \in S^{B\beta}(\Omega_\mu)$ for some B and μ, then $\Delta_\phi(\kappa^2) \in S_{B^{-1}\beta}(0,\infty) = [\phi : |x^k \phi^{(q)}(x)| \leq c_q B^k (c_k)^{-k}, x \in (0,\infty)]$.

Proof: Let us consider $\Delta_\phi(\kappa^2)$ of (21), and form

FIG. 3. The integration contour C in the complex x_0 plane over which the massive commutator $\langle C_2, \phi\rangle$ of Eq. (34) is evaluated.

$$\kappa^{2n}\Delta_\phi(\kappa^2) = \int d^4x \kappa^{2n}\Delta(x,\kappa^2)\phi(x) = \int d^4x(\Box^2)^n\Delta(x,\kappa^2)\phi(x),$$
$$= \int d^4x\Delta(x,\kappa^2)(\Box^2)^n\phi(x). \tag{24}$$

Using the standard representation for $\Delta(x,\kappa^2)$,

$$\Delta(x,\kappa^2) = (4\pi r)^{-1}\frac{\partial}{\partial r}\{[\epsilon(x_0)J_0(\kappa\sqrt{x^2})]\theta(x^2)\},$$

we obtain

$$\kappa^{2n}\Delta_\phi(\kappa^2) = \int_{-\infty}^{+\infty} dx_0\epsilon(x_0)\int_0^{|x_0|} r^2 dr \frac{\partial}{\partial r}[r^{-1}\hat{\phi}_n(x_0,r)]\cdot J_0(\kappa\sqrt{x_0^2-r^2}),$$

where

$$\hat{\phi}_n(x_0,r) = \int_{|\mathbf{r}|=r} d\Omega(\mathbf{r})(\Box^2)^n\phi,$$

$d\Omega(\mathbf{r})$ being the measure of the spherical surface $|\mathbf{r}|=r$. In terms of the variables $\lambda=\sqrt{x_0^2-\mathbf{r}^2}$ and r, we have

$$\kappa^{2n}\Delta_\phi(\kappa^2) = \int_0^\infty \lambda d\lambda \int_0^\infty r^2 dr(r^2+\lambda^2)^{-1/2}\frac{\partial}{\partial r}[r^{-1}\overline{\phi}_n^{(+)}(\lambda,r) - r^{-1}\overline{\phi}_n^{(-)}(\lambda,r)]J_0(\kappa\lambda), \tag{25}$$

where $\overline{\phi}_n^{\pm}(\lambda,r) = \hat{\phi}_n(\pm\sqrt{\lambda^2+r^2},r)$. From (10) we have

$$r^{-1}|\overline{\phi}_n^{+}(\lambda,r) - \overline{\phi}_n^{-}(\lambda,r)| \le (\lambda^2+r^2+1)^{-N}B^{2n}(c_{2n})^{-2n} \tag{26}$$

for a suitable positive integer N, so

$$|\kappa^{2n}\Delta_\phi(\kappa^2)| \le (\text{const})B^{2n}(c_{2n})^{-2n}. \tag{27}$$

The derivative $\kappa^{2n}\Delta_\phi^{(q)}(\kappa^2)$ can be handled exactly as above, with

$$\kappa^{2n}\Delta_\phi^{(q)}(\kappa^2) = \int_0^\infty d\lambda\lambda^{q+1}\int_0^\infty r^2 dr(\lambda^2+r^2)^{-1/2}(\partial/\partial r)r^{-1} \times[\overline{\phi}_n^{+}(\lambda,r) - \overline{\phi}_n^{-}(\lambda,r)]J_0^{(q)}(\lambda\kappa).$$

The bound (26) with $N \ge q+2$ will thus prove

$$|\kappa^{2n}\Delta_\phi^{(q)}(\kappa^2)| \le C_q B^{2n}(c_{2n})^{-2n}.$$

Thus $\Delta_\phi(\kappa^2)\in S_{B^{-1}g}((0,\infty))$. We may thus extend $[A(x),A(0)]_-$ as a generalized function-valued operator to the space S_x^g if $\rho\in S_g'$ (since $\cap_B S_{B^{-1}g}=S_g'$). We may specify the increase of $\rho(\kappa^2)$ in this case straightforwardly from (27), since then

$$|\Delta_\phi(\kappa^2)| \le \sup_{n\ge 0} C_0[B^{2n}(c_{2n})^{-2n}/\kappa^{2n}] = C_0\overline{G}(\kappa^2/B); \tag{28}$$

so the increase of ρ for larger κ^2 must be slower than $\overline{G}(\kappa^2)^{-1}$. In the case of $c_n=n^{-\alpha}$ we obtain the familiar value $\overline{G}(\kappa^2)^{-1}=\exp\kappa^{1/2\alpha}$. We have thus proved

Theorem 1: The generalized free field (19) has an extension whose order of decrease outside the light cone is specified by the smallest possible g for which $\rho\in S_g'$. The bound on ρ for large κ^2 is essentially $\overline{G}(\kappa^2)^{-1}$, where $\overline{G}(\kappa^2)$ is given by (28).

Let us now turn to another simple example of a nonlocalizable field, for the case

$$B(x) = g(\Box^2)A(x), \tag{29}$$

where $A(x)$ is a tempered field (though A could be localizable) and g is a suitable function. To consider this case we note that the Wightman functions for B are trivially related to those for A:

$$W_n^{(B)}(x_1,\ldots,x_n) = \prod_{i=1}^n g(\Box_i^2)W_n^{(A)}(x_1,\ldots,x_n).$$

Let us consider specifically the 2-point function

$$W_2^{(B)}(x_1,x_2) = g^2(\Box_{x_1x_2}^2)W_2^A(x_1,x_2).$$

We may write W_2^A by means of a Lehmann representation

$$W_2^{(A)}(x) = \int_{m_0^2}^\infty d\kappa^2\rho^{(A)}(\kappa^2)\Delta^{(+)}(x,\kappa^2), \tag{30}$$

so that $W_2^{(B)}$ also has such a representation, though now with weight function $\rho^{(B)}(\kappa^2=g^2(\kappa^2)\rho^{(A)}(\kappa^2)$. Thus if g is an entire function with exponential growth of order $(\frac{1}{2}\alpha)$ and $\rho^{(A)}$ is a measure then $\rho^{(B)}(\kappa^2)\in S_\alpha'$. By the discussion for the generalized free field the commutator

$$C_2^{(A)}(x) = \int_{m_0^2}^\infty d\kappa^2\rho^{(A)}(\kappa^2)\Delta(x,\kappa^2)$$

can be extended to some space S_x^α. For the general commutator

$$C_{n,j}^{(B)}(x_1,\ldots,x_n) = \prod_{i=1}^n g(\Box_{x_i}^2)C_{n,j}^{(A)}(x_1,\ldots,x_n) \tag{31}$$

the tempered distribution $C_{nj}^{(A)}$ vanishes outside the light cone in the variable $\xi_j=x_j,\ldots,x_{j+1}$. The generalized function $g(\Box_{\xi_j}^2)C_{nj}^{(A)}(\xi_j)$ will have an extension outside the light cone which can be determined immediately, since

$$\langle\prod_{i=1}^n g(\Box_i^2)C_{nj},\phi\rangle = \langle C_{nj},\prod_{i=1}^n g(\Box_i^2)\phi\rangle. \tag{32}$$

The details of the extension are given by

Lemma 6: If g is an entire function order of growth $\frac{1}{2}\alpha$, then

$$\prod_{i=1}^n g(\Box_i^2)\phi \in \mathcal{S}(\mathbf{R}^{4(n-1)})$$

for any $\phi\in S^\beta(\mathbf{R}^{4(n-1)})$, any $\beta<\alpha$, whilst if $\phi\in T_{nj}(\beta)$ then

$$\prod_{i=1}^n g(\Box_i^2)\phi \in \mathcal{S}_{nj}^x(R^{4(n-1)})$$

for any $\beta<\alpha$. Here

$$\mathcal{S}_{nj}^x(R^{4(n-1)}) = \bigcup_{\mu>0}\mathcal{S}(\mathbf{R}^{4(n-2)})\otimes\mathcal{S}(\Omega_\mu),$$

where $\mathcal{S}(\Omega_\mu)$ is in the variable ξ_j. The first part of this lemma follows immediately by Fourier transformation,

FIG. 4. The final contour in the x_0 plane to which C of Fig. 2 may be shrunk after taking account of the singularities in $\langle C_2,\phi\rangle$.

whilst both it and the second part can be derived by direct computation by using the defining properties of the spaces S^α and $\mathcal{S}$. Since the support of C_{nj} is the interior of the light cone in ξ_j then the right-hand side of (32) is defined for $\phi \in T_{nj}(\beta)$ for any $\beta < \alpha$. We have thus proved

Theorem 2: The massive field (29) has an extension of order of decrease α. It is evidently possible to obtain further results on this extension, such as its range or its order, in the case of S^δ spaces, but we will not do that here, because the results are not of essential interest. This is especially so for both of these examples in that the first, the generalized free field, has the trivial unit S matrix whilst the example (29) has the same S matrix elements as the field $A(x)$, as can easily be seen in momentum space. We leave these cases, then, and turn to functions of the free massive field. These also have trivial S matrix, but have a great deal more complexity.

5. FUNCTIONS OF THE FREE MASSIVE FIELD

As before we consider functions of the form

$$B(x) = \sum_{n\geq 0} (d_n/n!) : A(x)^n : , \tag{33}$$

where $A(x)$ is a free scalar field of mass m. We will only investigate the 2-point function here for simplicity, especially because this situation is already quite complicated. We have, for the massless field, and taking (14), that

$$\langle C_2, \phi\rangle = \int_0^\infty dt e^{-t} \lim_{\epsilon\to 0} \int_{C_\epsilon} dx_0 \int d^3x \exp[t^\alpha \Delta(x, m^2)]\phi(x) . \tag{34}$$

We have that $\Delta(x, m^2) = m(-x^2)^{-1/2} K_1[m(-x^2)^{1/2}]$, so that for $|\mathbf{r}| \neq 0$ the contour C_ϵ is to be taken as in Fig. 3. We can express the most singular part of Δ as

$$\Delta(x, m^2) = -(x^2)^{-1} - \tfrac{1}{2} m \log(\tfrac{1}{2} m\sqrt{-x^2}) + \log(-x^2) O(x^2) .$$

Thus

$$\exp[t^\alpha \Delta(x, m^2)] = \exp[-t^\alpha/x^2](t^2\sqrt{-x^2})^{-m/2} \exp[t^\alpha \log(-x^2) O(x^2)] . \tag{35}$$

The integration contour C_ϵ may be modified so as to include two small circles, one round $|\mathbf{r}|$, the other round $-|\mathbf{r}|$, together with the remainder, so giving the contours of Fig. 4. The contribution from the circles may be evaluated as for the massless case in Sec. 3 by expansion of the first factor in (35). Except for rather special values of m this will give no contribution at all, neglecting the third factor on the rhs of (25) very near $x = 0$. So the main contribution to (34) is completely different in the massive case from the massless situation; only the contribution from the contours outside $\pm|\mathbf{r}|$ in Fig. 4 are to be considered.

To obtain the extension of (34) we expand the exponential in (34) and use an integral representation[11] for powers of $\Delta(x, m^2)$:

$$\langle C_2, \phi\rangle = 2\int_0^\infty dt e^{-t} \int d^4x \phi(x) \sum_{n\geq 0} (t^{n\alpha}/n!) \int_{(nm)^2}^\infty d\kappa^2 \Delta \times (x, \kappa^2) \Omega_n^{(m)}(\kappa^2) , \tag{36}$$

where $\Omega_n^{(m)}(\kappa^2)$ is the phase space for n particles of mass m and total squared center of mass energy κ^2. We use the result of Lemma 5, so that if $\phi \in S_x^\delta$, with $\beta < 1$, then $\Delta_\phi(\kappa^2) \in S_{B\beta}(R_+)$, for some B. Using the bound[11]

$$|\Omega_n^{(m)}(\kappa^2)| \leq (\text{const}) \cdot [\kappa^{(n-3)/2}(\kappa - nm)^{(3n-s)/2}/\Gamma(2n)] \tag{37}$$

to within a function of slow increase in n and κ, which we can neglect without error,

$$|\langle C_2, \phi\rangle| \leq (\text{const}) \times \int_0^\infty dt e^{-t} \sum_{n\geq 0} \frac{t^{n\alpha}}{n!} \int_{(nm)^2}^\infty d\kappa^2 \times \kappa^{(n-3)/2}(\kappa - nm)^{(3n-s)/2} \exp[-(\beta\kappa)^{1/\beta}/[\Gamma(2n)] . \tag{38}$$

We may evaluate the κ^2 integral on the rhs of (38) by the change of variable $\kappa = nmx$ to give

$$(nm)^{2n-2} \int_1^\infty dx\, x^{(n-1)/2}(x-1)^{(3n-s)/2} \exp[-(bnmx)^{1/\beta}] . \tag{39}$$

Denoting by $g(x)x^{-3/2}$ the integrand of (39), we may put a bound on (39) by finding the positive of the maximum of g, which is at the solution of

$$g'(x) = \left(\frac{n+2}{2x} + \frac{(3n-s)}{2x-1} - \frac{(bnm)^{1/\beta}}{\beta} x^{1/\beta - 1}\right) g(x) = 0 .$$

The solution of this for large n is very close to $x = 1$ (in the range $1 \leq x \leq \infty$), and has value $x = 1+\epsilon$, $\epsilon \sim (3n-s) \times \beta/2(bnm)^{1/\beta}$, which is as small as we please for n large enough. Then (39) is bounded for all n by

$$(nm)^{2n-2}(1+\epsilon)^{(n+2)/2}\epsilon^{(3n-s)/2} \exp[-(bm)^{1/\beta}[n(1+\epsilon)]^{1/\beta} \times \int_1^\infty dx/x^{3/2} ,$$

with crucial contribution proportional to

$$n^{2n}[(3n-s)\beta/2]^{(3n-s)/2} n^{-[(3n-s)/2]/\beta} \exp[-(bnm)^{1/\beta}] ,$$

or

$$n^{(3n)/2(1-1/\beta)+2n} \exp[-(bm)^{1/\beta} n^{1/\beta}] .$$

Thus we have the bound on (38) given by

$$|\langle C_2, \phi\rangle| \leq \int_0^\infty dt e^{-t} \sum_{n\geq 0} (t^{n\alpha}/n!) n^{(3n)/2(1-1/\beta)} \times \exp[-(bmn)^{1/2}] \tag{40}$$

and this is finite if $\beta < 1$, since then the summation in (40) gives a function increasing at infinity slower than $\exp(t)$. Thus we need to choose any $\beta < 1$ for the extension of $\langle C_2, \phi\rangle$ from S^α to S_x^β to be possible. We have thus proved

Theorem 3: The 2-point commutator bracket of the function (33) of the massive free field has an extension of order β for any $\beta < 1$, where α is defined by (14) and $\alpha > 1$.

We note that the massless case does not give the same limitation if the above method is used here, but only the condition $\alpha < 1$; this approach only works in that case for a localizable theory. We see that the above method could be extended to the indicatrix spaces, and also to higher point functions, though we will not do the latter of those here since no further insight into the situation is expected to be gained. *However, we can sharpen the results of Theorem* 3 so as to relate to the discussion of Rieckers.[12] If we assume, with Rieckers, that for large n,

$$d_n^2/n! \sim \exp(n^{1/\alpha}), \tag{41}$$

then we have to replace the expression (36) by

$$\langle C_2, \phi\rangle = \sum_{n\geq 0} (d_n^2/n!) \int_{(nm)^2}^{\infty} \Delta_\phi(\kappa^2)\Omega_n^{(m)}(\kappa^2)d\kappa^2;$$

so

$$|\langle C_2, \phi\rangle| \leq \sum_n \exp(an^{1/\alpha}) \int_{(nm)^2}^{\infty} d\kappa^2 \frac{\kappa^{(n-3)/2}(\kappa - nm)^{(3n-s)/2}}{\Gamma(2n)} \times \exp[-(b\kappa)^{1/\beta}].$$

Using the previous method we obtain

$$|\langle C_2, \phi\rangle| \leq \sum_n \exp[an^{1/\alpha-(bmn)1/\beta} n^{(n/2)(1-3/\beta)}. \tag{42}$$

This is convergent for any $\beta < \alpha < 1$ (agreeing with Rieckers'[12] results for the Wightman functions), so proving

Theorem 4: The 2-point commutator bracket of the function (33) with coefficients d_n satisfying (41) has a decrease outside the light cone of order α, and is damped like $\exp(-|x^2|^{1/2\alpha})$ as $x^2 \to -\infty$.

6. DISCUSSION

We have obtained a prescription for describing how the commutator bracket of a nonlocalizable field extends outside the light cone, and shown that is is applicable to various models. There are two difficulties associated with this. The first is that we have not been able to show that our prescription does actually work for the case of functions of a massless free field. This is rather surprising since we expect that case to be simpler than the massive one. This problem is associated with that of defining $S^{(n)}(x^2)$ for $n > 2$, and of course related to the fact that in the massless case all the higher particle thresholds coalesce onto the single particle one. We do not at present see any way of satisfactorily treating this question, though feel it rather pressing especially because of all the work involved in applications of nonpolynomial Lagrangians which use the massless case.[2]

The second difficulty is that we have not been able to present a realistic model of a truly nonlocalizable field, that is, one for which the S matrix is not the same as that arising from some localizable one. Only if that can be done can we expect that there is any possible physical trace of noncausality. Indeed, we have discussed[13] recently the manner in which the notion of Borcher's equivalence classes of fields[14] can be extended to include nonlocalizable ones. What is needed is a proof that any nonlocalizable field is equivalent in this extended sense to some localizable field. This can, indeed, be done if indefinite metric localizable fields are allowed, but it is not known if such a theorem is true for positive metric fields.

Even if it is not possible to say whether or not nonlocalizable fields actually appear as such in nature can we say anything about the expected sizes of the range and order of the extension, if it exists? A natural range would be that determined by the radius of an interaction. There is far greater difficulty about a natural value for the order of the extension, it being a dimensionless quantity. However, the dimensionless quantities of interest are the coupling constants of the various interactions. But present evidence indicates that the order of the nonlocalizability depends heavily on the nature of the interaction; it may well be that among all equivalent interactions the least order of decrease is determined by the dimensionless strength.

We must realize, of course, that noncausality need not destroy many of the usual results which follow from causality, such as analyticity[14] and even polynomial boundedness of S-matrix elements may still be valid. There may be observable effects of nonpolynomiality in the behavior of form factors, as discussed in the localizable case by Jaffe.[15] We hope to discuss this and related questions in more detail elsewhere.

ACKNOWLEDGMENTS

One of us, F.C. would like to thank the Science Research Council of Great Britain for a grant and King's College, London, for its hospitality to enable this work to be achieved.

[1]A. Jaffe, Ann. Phys. (N.Y.) 32, 127 (1965).
[2]See, for example, A. Salam "Computation of the Renormalization Constants," in *Proceedings of the Coral Gables Conference*, edited by Dal Cin, Iverson, and Perlmutter (Gordon and Breach, New York, 1971) and J.G. Taylor "Non Polynomial Lagrangians, Particles, and Black Holes," Lectures at Toronto Summer School (June 1971),to appear in *Lectures on Particles and Fields*, edited by H.H. Aly and references quoted in these papers.
[3]S. Fels, Phys. Rev. D1, 2370 (1970).
[4]J.G. Taylor, Ann. Phys. (N.Y.) 68, 484 (1971).
[5]R. Blomer and F. Constantinescu, Nuovo Cimento A 6, 62 (1971).
[6]R.F. Streater and A.S. Wightman, *PCT and All That* (Benjamin, New York, 1964).
[7]I. Gelfand and G. Shilov, *Generalized Functions* (Academic, New York, 1964), Vols. 1 and 2.
[8]A. Jaffé, Phys. Rev. 158, 1454 (1967).
[9]H.J. Borchers and K. Pohlmeyer, Comm. Math. Phys. 8, 269 (1968).
[10]See Ref. 7, Vol. 2, Chap. IV, Sec. 3.2.
[11]M.K. Volkov, Ann. Phys. (N.Y.) 49, 202 (1968), Eq. (2.4).
[12]A. Rieckers, Int. Journ. Theor. Phys. 4, 55 (1971).
[13]F. Constantinescu and J.G. Taylor, "Equivalence between Nonlocalizable and Local Fields," King's College preprint (April 1972).
[14]H.J. Borchers, N. Cimento 15, 784 (1960).
[15]A.M. Jaffé, Phys. Rev. Lett. 17, 661 (1966).

Scattering of a plane longitudinal elastic wave by a large convex rigid object with a statistically corrugated surface. I. Perturbation solutions

W. Fan and Y. M. Chen

Department of Applied Mathematics and Statistics,
State University of New York at Stony Brook, New York 11790
(Received 1 March 1972)

The scattering of time-harmonic plane longitudinal elastic wave by a large convex rigid object with statistical surface irregularities is considered. The maximum deviation of the corrugated surface from the smooth one is assumed to be small, and hence the boundary-perturbation technique is utilized in this study. First, the scattering of longitudinal wave by a large rigid sphere with statistical surface irregularities is treated as a canonical problem in the general discussion. It is found that the higher-order solutions can be obtained from the zeroth-order solution in a straightforward manner. Due to the complexity of the problem, only the first-order solution and its asymptotic expansion are explicitly computed and carried out. A general recipe based on the zeroth-order solution is given for the treatment of the general problem. The asymptotic expressions of mean values of the scattered wave function and the scattered intensity are also given for the general problem.

1. INTRODUCTION

In recent years, an increasing amount of attention has been devoted to the study of the effect of statistical surface irregularities on propagation and scattering of various types of waves. Although the general problem of scattering waves from statistical corrugated surfaces appears to be difficult, a number of investigators have been able to make progress by applying either probability theory or perturbation theory to the problem of scattering of waves by statistically corrugated surfaces under a few suitable assumptions. For a systematic classification of existing theories developed for random irregular plane surfaces and a rather complete bibliography, readers are referred to an excellent text by Beckmann and Spizzichino.[1] Recently, upon using a boundary perturbation technique, Chen and Kim,[2] Chen,[3] and Chen and Fan[4] have studied the effects on the scattering of scalar waves by convex objects with statistical surface irregularities.

The effects of the deterministic surface irregularities of an infinite plane surface on the propagation of elastic waves have been studied by Dunkin and Eringen,[5] Handelman,[6] and Abubakar.[7,8] In the present paper, the problem of the scattering of time harmonic plane longitudinal elastic wave by a large convex rigid object with statistical surface irregularities is investigated. The problem permits scalarization such that the method used in the scalar cases (3) and (4) applies. Here the ratio of the maximum deviation of the corrugated surface from the unperturbed one to the local radius of the scatterer is assumed to be small. Hence the scattered wave can be determined by the boundary-perturbation technique,[9] which is based on the Taylor expansion of boundary conditions at the perturbed boundary and the representation of the field as a power series in the aforementioned ratio.

First, we shall treat the scattering of plane longitudinal elastic wave by a rigid sphere with a statistically corrugated surface as a canonical problem in our general discussion.[10] This has the advantage of illustrating the method without introducing extraneous geometrical details; it is also a case for which we can obtain the exact perturbation solution, assuming that the perturbation series converges. It is found that the higher-order perturbation solutions can be systematically obtained from the zeroth-order perturbation in a rather straightforward manner. The exact perturbation solution can be then asymptotically expanded for large $k_i a$ $(i = 1, 2)$. However, due to the complexity of the problem, only the first-order perturbation solution and its asymptotic representation are explicitly computed. Its expression contains the spherical surface integration over all contributions generated by the equivalent source S_1, arising from the interaction of the zeroth-order solution with surface irregularities.

Next, as for the case of scalar waves,[3] a general recipe based on the zeroth-order solution can be constructed from the geometrical theory of diffraction.[10,11,12] And this recipe is given in the case of scattering by a large convex rigid object with a statistically corrugated surface. Finally, the formulas of the asymptotic representation of mean values of the scattered wave function and the scattered intensity are given for the general problem. It is found that if the surface is statistically homogeneous, the mean value of the surface fluctuation is zero, i.e., $\langle\langle f \rangle\rangle = 0$, and then the mean value of the first-order perturbation is zero. In this case, the explicit computation of the second-order perturbation solution is needed. However, as long as the surface is not statistically homogeneous, the effect of the random rough surface on the scattering of elastic waves can be sufficiently explained by knowing the first-order perturbation solution explicitly.

In a sequel to this paper, the aforementioned surface integral of the first-order perturbation solution containing S_1 will be evaluated asymptotically for the field point far away from the scatterer. Then the effective reflection and diffraction coefficients will be deduced from it. Finally, the effective scattering cross section will be given.

2. GENERAL FORMULATION

The propagation of waves in an homogeneous isotropic elastic medium is governed by the equation

$$(\lambda + 2\mu)\nabla(\nabla \cdot \mathbf{w}) - \mu\nabla \times \nabla \times \mathbf{w} = \rho \frac{\partial^2 \mathbf{w}}{\partial t^2}, \tag{2.1}$$

where $\mathbf{w}$ is the displacement vector, ρ is the density of the medium, λ is the Lame's parameter, μ is the shear modulus of the medium and $\lambda + \frac{2}{3}\mu$ is its compression modulus of the isotropic elastic medium.

By assuming that $\mathbf{w} = \mathbf{u}(\mathbf{r})e^{iwt}$, Eq. (2.1) can be rewritten as

$$\alpha^2\nabla(\nabla\cdot\mathbf{u}) - \beta^2\nabla\times\nabla\times\mathbf{u} + \omega^2\mathbf{u} = 0 \tag{2.2}$$

where $\alpha = (\lambda + 2\mu/\rho)^{1/2}$ and $\beta = (\mu/\rho)^{1/2}$ are the velocities of compressional and shear waves respectively in the medium.

Here the scatterer is assumed to be rigid and infinitely dense. Hence u satisfies the boundary condition

$$\mathbf{u} = 0 \tag{2.3}$$

on the surface of the scatterer and a certain radiation condition for $r \to \infty$ to insure the uniqueness of $\mathbf{u}$.

Let Φ and Ψ be two scalar functions which satisfy equations

$$\nabla^2\Phi + K_1^2\Phi = 0 \tag{2.4}$$

and

$$\nabla^2\Psi + K_2^2\Psi = 0, \tag{2.5}$$

where

$$K_1 = \omega/\alpha \quad \text{and} \quad K_2 = \omega/\beta \tag{2.6}$$

Then the three independent solutions of (2.2) in the spherical coordinate system are

$$\mathbf{l} = \nabla\Phi, \tag{2.7}$$

$$\mathbf{m} = \nabla\times\mathbf{a}\Psi, \tag{2.8}$$

$$\mathbf{n} = 1/K_2\nabla\times\mathbf{m}, \tag{2.9}$$

where $\mathbf{a}$ is any constant unit vector. From Stratton,[13] one obtains the following set of characteristic wave functions

$$\mathbf{l}_{\substack{e\\o}mn}(\mathbf{r}) = \frac{\partial}{\partial r} Z_n(k_1r)P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix} m\varphi\,\mathbf{1}_r + \frac{1}{r}Z_n(k_1r)\frac{\partial}{\partial\theta}P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix} m\varphi\,\mathbf{1}_\theta \mp \frac{m}{r\sin\theta}Z_n(k_1r)P_n^m(\cos\theta)\begin{matrix}\sin\\\cos\end{matrix} m\varphi\,\mathbf{1}_\varphi,$$

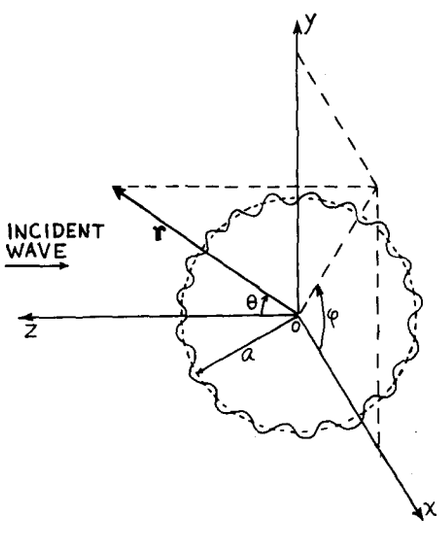


FIG. 1. The geometry of a statistically corrugated sphere is shown.

$$\mathbf{m}_{\substack{e\\o}mn}(\mathbf{r}) = \mp\frac{m}{\sin\theta}Z_n(k_2r)P_n^m(\cos\theta)\begin{matrix}\sin\\\cos\end{matrix} m\varphi\,\mathbf{1}_\theta - Z_n(k_2r)\frac{\partial}{\partial\theta}P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix} m\varphi\,\mathbf{1}_\varphi,$$

$$\mathbf{n}_{\substack{e\\o}mn}(\mathbf{r}) = \frac{n(n+1)}{k_2r}Z_n(k_2r)P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix} m\varphi\,\mathbf{1}_r + \frac{1}{k_2r}\frac{\partial}{\partial r}[rZ_n(k_2r)]\frac{\partial}{\partial\theta}P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix} m\varphi\,\mathbf{1}_\theta \mp \frac{m}{k_2r\sin\theta}\frac{\partial}{\partial r}[rZ_n(k_2r)]P_n^m(\cos\theta)\begin{matrix}\sin\\\cos\end{matrix} m\varphi\,\mathbf{1}_\varphi,$$

where $\mathbf{1}_r, \mathbf{1}_\theta, \mathbf{1}_\varphi$ are unit base vectors for the spherical coordinate system; $Z_n(k_ir)$ can be any one of the spherical Bessel functions $j_n(k_ir), n_n(k_ir), h_n^{(1)}(k_ir)$ or $h_n^{(2)}(k_ir)$; $P_n^m(\cos\theta)$ is the associated Legendre polynomial; and the subscript e or o refers to the even or odd nature of the function of φ. These functions form a complete set with the l functions corresponding to the compression wave and the m and n functions corresponding to the shear waves.

3. BOUNDARY PERTURBATION FOR A ROUGH SPHERE

Let the position vector be $\mathbf{r} = (r, \theta, \varphi)$. If the random surface S has only a small deviation from the sphere with radius a we may define the surface S (Fig. 1) by

$$\tilde{\mathbf{r}} = a[1 + \epsilon f(\theta, \varphi, q)], \tag{3.1}$$

where ϵ is a small parameter such that $|\epsilon f| < 1$; $f(\theta, \varphi, g)$ is a smooth continuous function of θ and φ such that $\left|\frac{\partial f}{\partial\theta}\right| < 1$ and $\left|\frac{\partial f}{\partial\varphi}\right| < 1$; and g is a random variable over the probability space Q in which a probability density function $P(g)$ is defined such that the average value of a random function $V(\mathbf{r}, g)$ is defined as

$$\langle\langle V\rangle\rangle = \int_Q V(\mathbf{r}, g)P(g)dg. \tag{3.2}$$

The perturbation technique consists of setting the scattered wave $\mathbf{u}_s$ in the form

$$\mathbf{u}_s(\mathbf{r}, g) = \mathbf{u}_{s0}(\mathbf{r}) + \sum_{j=1}^{\infty}\epsilon^j\mathbf{u}_{sj}(\mathbf{r}, g) \tag{3.3}$$

and expanding $\mathbf{u}_i$ and $\mathbf{u}_{sj}$ in Taylor series near $r = a$. Upon setting $r_0 = (a, \theta, \varphi)$, we obtain

$$\mathbf{u}_i(\mathbf{r}) = \mathbf{u}_i(\mathbf{r}_0) + \sum_{j=1}^{\infty}\frac{(\epsilon af)^j}{j!}\mathbf{u}_i^{(j)}(\mathbf{r}_0) \tag{3.4}$$

and

$$\mathbf{u}_s(\mathbf{r}) = \mathbf{u}_{s0}(\mathbf{r}_0) + \epsilon\left[af\left.\frac{\partial\mathbf{u}_{s0}}{\partial r}\right|_{r=a} + \mathbf{u}_{s1}(\mathbf{r}_0, g)\right] + \epsilon^2\left[\frac{a^2f^2}{2}\left.\frac{\partial^2\mathbf{u}_{s0}}{\partial r^2}\right|_{r=a} + af\left.\frac{\partial\mathbf{u}_{s1}}{\partial r}\right|_{r=a} + \mathbf{u}_{r2}(\mathbf{r}_0, g)\right] + O(\epsilon^3). \tag{3.5}$$

Next, we substitute these expressions into the boundary condition (2.3). By collecting coefficients of $\epsilon^0, \epsilon, \epsilon^2, \cdots$ and equating each one to zero, we obtain

$$\mathbf{u}_i(\mathbf{r}_0) + \mathbf{u}_{s0}(\mathbf{r}_0) = \mathbf{0}, \tag{3.6}$$

$$af\left[\left.\frac{\partial\mathbf{u}_i}{\partial r}\right|_{r=a} + \left.\frac{\partial\mathbf{u}_{s0}}{\partial r}\right]_{r=a} + \mathbf{u}_{s1}(\mathbf{r}_0, g) = 0, \tag{3.7}$$

$$\frac{a^2f^2}{2}\left[\frac{\partial^2\mathbf{u}_i}{\partial r^2}\bigg|_{r=a} + af\frac{\partial^2\mathbf{u}_{s0}}{\partial r^2}\bigg|_{r=a}\right] + af\frac{\partial\mathbf{u}_{s1}}{\partial r}\bigg|_{r=a} + \mathbf{u}_{s2}(\mathbf{r}_0,g) = \mathbf{0}. \quad (3.8)$$

As it has been pointed out by Chen,[3] the essential effect of this boundary-perturbation technique is to transform the original boundary value problem to an equivalent boundary value problem with unperturbed body as the scatterer on which equivalent sources are induced. These source functions arise from the interactions between the surface irregularity and all the lower order solutions.

4. PERTURBATION SOLUTION OF A ROUGH SPHERE

Consider the case where a unit plane compression wave propagating along the negative direction of z axis and impinging on the aforementioned rough sphere. Since the incident wave is independent of φ and the fact that

$$\mathbf{l}_{o0n}(\mathbf{r}) = \mathbf{0}, \quad n = 0,1,2,\cdots. \quad (4.1)$$

This incident wave can be expressed as

$$\mathbf{u}_i(\mathbf{r}) = \sum_{n=0}^{\infty}\alpha_n\mathbf{l}_{e0n}(\mathbf{r}), \quad (4.2)$$

where

$$\alpha_n = 1/k_1(2n+1)\exp[i\pi/2(n+1)], \quad (4.3)$$

since $\mathbf{u}_i(0) < \infty$, we let $Z_n(k_1r) = j_n(k_1r)$ in $\mathbf{l}_{e0n}(\mathbf{r})$ of (4.2).

The zeroth-order scattered wave can be represented as the linear combination of $\mathbf{l}, \mathbf{m}$ and $\mathbf{n}$. Since $\mathbf{u}_i(\mathbf{r})$ is independent of φ and the sphere is symmetric with respect to φ, $\mathbf{u}_{s0}(\mathbf{r})$ should be also independent of φ, i.e., the subscript m is equal to zero. Observing the fact that

$$\mathbf{l}_{o0n}(\mathbf{r}) = m_{o0n}(\mathbf{r}) = \mathbf{n}_{o0n}(\mathbf{r}) = 0, \quad n = 0,1,2,\cdots, \quad (4.4)$$

we can represent the zeroth-order scattered wave as

$$\mathbf{u}_{s0}(\mathbf{r}) = \sum_{n=0}^{\infty}[A_n\mathbf{l}_{e0n}(\mathbf{r}) + B_n\mathbf{m}_{e0n}(\mathbf{r}) + C_n\mathbf{n}_{e0n}(\mathbf{r})], \quad (4.5)$$

since $\mathbf{u}_s(\mathbf{r},g)|_{r\to\infty} \simeq (1/r)e^{-ik_1r}$, we let $Z_n(k_ir) = h_n^{(2)}(k_ir)$ in all scattered waves. By matching the boundary condition (3.6), we obtain

$$\alpha_n\mathbf{l}_{e0n}(\mathbf{r}_0) + A_n\mathbf{l}_{e0n}(\mathbf{r}_0) + B_n\mathbf{m}_{e0n}(\mathbf{r}_0) + C_n\mathbf{n}_{e0n}(\mathbf{r}_0) = \mathbf{0}, \quad n = 0,1,2,\cdots, \quad (4.6)$$

which is equivalent to

$$\alpha_nk_1j_n'(k_1a) + A_nk_1h_n^{(2)\prime}(k_1a) + C_n\frac{n(n+1)}{k_2a}h_n^{(2)}(k_2a) = 0, \quad (4.7)$$

$$\alpha_nj_n(k_1a) + A_nh_n^{(2)}(k_1a) + C_n\frac{1}{k_2}[h_n^{(2)}(k_2a) + k_2ah_n^{(2)\prime}(k_2a)] = 0, \quad (4.8)$$

$$B_n = 0. \quad (4.9)$$

Upon solving (4.7), (4.8), and (4.9), we obtain the zeroth-order solution

$$\mathbf{u}_{s0}(\mathbf{r}) = u_{0pr}(\mathbf{r})\mathbf{1}_r + u_{0p\theta}(\mathbf{r})\mathbf{1}_\theta + u_{0sr}(\mathbf{r})\mathbf{1}_r + u_{0s\theta}(\mathbf{r})\mathbf{1}_\theta, \quad (4.10)$$

where

$$\mathbf{u}_{0pr}(\mathbf{r}) = \sum_{n=0}^{\infty}A_nk_1h_n^{(2)\prime}(k_1r)P_n(\cos\theta), \quad (4.11)$$

$$\mathbf{u}_{0p\theta}(\mathbf{r}) = \sum_{n=0}^{\infty}A_n\frac{1}{r}h_n^{(2)}(k_1r)\frac{\partial}{\partial\theta}P_n(\cos\theta), \quad (4.12)$$

$$\mathbf{u}_{0sr}(\mathbf{r}) = \sum_{n=0}^{\infty}C_n\frac{n(n+1)}{k_2r}h_n^{(2)}(k_2r)P_n(\cos\theta), \quad (4.13)$$

and

$$\mathbf{u}_{0sr}(\mathbf{r}) = \sum_{n=0}^{\infty}C_n\frac{1}{k_2r}\frac{\partial}{\partial r}[rh_n^{(2)}(k_2r)]\frac{\partial}{\partial\theta}P_n(\cos\theta), \quad (4.14)$$

with

$$A_n = \frac{2n+1}{k_1}\frac{[n(n+1)/k_2a]j_n(k_1a)h_n^{(2)}(k_2a) - (k_1/k_2)j_n'(k_1a)[h_n^{(2)}(k_2a) + k_2ah_n^{(2)\prime}(k_2a)]}{[n(n+1)/k_2a]h_n^{(2)}(k_1a)h_n^{(2)}(k_2a) - (k_1/k_2)h_n^{(2)\prime}(k_1a)[h_n^{(2)}(k_2a) + k_2ah_n^{(2)\prime}(k_2a)]}e^{i(\pi/2)(n-1)} \quad (4.15)$$

and

$$C_n = \frac{2n+1}{k_1^2a^2}\left(\frac{n(n+1)}{k_2a}h_n^{(2)}(k_1a)h_n^{(2)}(k_2a) - \frac{k_1}{k_2}h_n^{(2)1}(k_1a)[h_n^{(2)}(k_2a) + k_2ah_n^{(2)1}(k_2a)]\right)^{-1}e^{i(\pi/2)n}. \quad (4.16)$$

Here u_{0pr} and $u_{0p\theta}$ are the r and the θ components of zeroth-order reflected P (pressure) wave, respectively. Similarly, u_{0sr} and $u_{0s\theta}$ are the r and the θ components of zeroth-order reflected S (shear) wave respectively.

The first-order solution can be represented by the complete expansion,

$$\mathbf{u}_{s1}(\mathbf{r},g) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}[A_{mn}\mathbf{l}_{emn}(\mathbf{r}) + B_{mn}\mathbf{l}_{omn}(\mathbf{r}) + C_{mn}\mathbf{m}_{emn}(\mathbf{r}) + D_{mn}\mathbf{m}_{omn}(\mathbf{r}) + E_{mn}\mathbf{n}_{emn}(\mathbf{r}) + F_{mn}\mathbf{n}_{omn}(\mathbf{r})]. \quad (4.17)$$

The boundary condition (3.7) becomes

$$\mathbf{u}_{s1}(\mathbf{r}_0,g) = \sum_{n=0}^{\infty}\sum_{m=0}\Big[A_{mn}\mathbf{l}_{emn}(\mathbf{r}_0) + B_{mn}\mathbf{l}_{omn}(\mathbf{r}_0) + C_{mn}\mathbf{m}_{emn}(\mathbf{r}_0) + D_{mn}\mathbf{m}_{0mn}(\mathbf{r}_0) + E_{mn}\mathbf{n}_{emn}(\mathbf{r}_0) + F_{mn}\mathbf{n}_{omn}(\mathbf{r}_0)\Big] = -af(\theta,\varphi,g)\left[\frac{\partial\mathbf{u}_i}{\partial r}\bigg|_{r=a} + \frac{\partial\mathbf{u}_{s0}}{\partial r}\bigg|_{r=a}\right] \equiv \mathbf{S}_1. \quad (4.18)$$

To find the coefficients A_{mn} through F_{mn}, we take the scalar products of (4.18) with $\mathbf{l}_{emn}(\mathbf{r}_0)$ through $\mathbf{n}_{0mn}(\mathbf{r}_0)$ successively. Here the scalar product of $\mathbf{l}$ and $\mathbf{n}$ is defined as

$$\langle\mathbf{l}\cdot\mathbf{n}\rangle = \int_0^{2\pi}\int_0^{\pi}\mathbf{l}\cdot\mathbf{n}\sin\theta\,d\theta d\varphi, \quad (4.19)$$

where $\sin\theta d\theta d\varphi$ is the element of surface on the unit sphere. Upon using the orthogonality relations given in Appendix (1), we get

$$A_{mn}\|\mathbf{l}\|^2 + E_{mn}\langle\mathbf{l}\cdot\mathbf{n}\rangle = \langle\mathbf{s}_1\cdot\mathbf{l}_{emn}(\mathbf{r}_0)\rangle, \quad (4.20)$$

$$B_{mn}\|\mathbf{l}\|^2 + F_{mn}\langle\mathbf{l}\cdot\mathbf{n}\rangle = \langle\mathbf{s}_1\cdot\mathbf{l}_{omn}(\mathbf{r}_0)\rangle, \quad (4.21)$$

$$C_{mn}\|\mathbf{m}\|^2 = \langle\mathbf{s}_1\cdot\mathbf{m}_{emn}(\mathbf{r}_0)\rangle, \quad (4.22)$$

$$D_{mn}\|\mathbf{m}\|^2 = \langle\mathbf{s}_1\cdot\mathbf{m}_{omn}(\mathbf{r}_0)\rangle, \quad (4.23)$$

$$A_{mn}\langle\mathbf{l}\cdot\mathbf{n}\rangle + E_{mn}\|\mathbf{n}\|^2 = \langle\mathbf{s}_1\cdot\mathbf{n}_{emn}(\mathbf{r}_0)\rangle, \quad (4.24)$$

$$B_{mn}\langle\mathbf{l}\cdot\mathbf{n}\rangle + F_{mn}\|\mathbf{n}\|^2 = \langle\mathbf{s}_1\cdot\mathbf{n}_{omn}(\mathbf{r}_0)\rangle. \quad (4.25)$$

By solving this system of six linear equations with six unknowns, we obtain the coefficients

$$A_{mn} = G_{mn}[\|\mathbf{n}\|^2\langle \mathbf{s}_1\cdot\mathbf{l}_{emn}(\mathbf{r}_0)\rangle - \langle \mathbf{l}\cdot\mathbf{n}\rangle\langle \mathbf{s}_1\cdot\mathbf{n}_{emn}(\mathbf{r}_0)\rangle], \quad (4.26)$$

$$B_{mn} = G_{mn}[\|\mathbf{n}\|^2\langle \mathbf{s}_1\cdot\mathbf{l}_{omn}(\mathbf{r}_0)\rangle - \langle \mathbf{l}\cdot\mathbf{n}\rangle\langle \mathbf{s}_1\cdot\mathbf{n}_{omn}(\mathbf{r}_0)\rangle], \quad (4.27)$$

$$C_{mn} = \|\mathbf{m}\|^{-2}\langle \mathbf{s}_1\cdot\mathbf{m}_{emn}(\mathbf{r}_0)\rangle, \quad (4.28)$$

$$D_{mn} = \|\mathbf{m}\|^{-2}\langle \mathbf{s}_1\cdot\mathbf{m}_{omn}(\mathbf{r}_0)\rangle, \quad (4.29)$$

$$E_{mn} = G_{mn}[\|\mathbf{l}\|^2\langle \mathbf{s}_1\cdot\mathbf{n}_{emn}(\mathbf{r}_0)\rangle - \langle \mathbf{l}\cdot\mathbf{n}\rangle\langle \mathbf{s}_1\cdot\mathbf{l}_{emn}(\mathbf{r}_0)\rangle], \quad (4.30)$$

$$F_{mn} = G_{mn}[\|\mathbf{l}\|^2\langle \mathbf{s}_1\cdot\mathbf{n}_{omn}(\mathbf{r}_0)\rangle - \langle \mathbf{l}\cdot\mathbf{n}\rangle\langle \mathbf{s}_1\cdot\mathbf{l}_{omn}(\mathbf{r}_0)\rangle], \quad (4.31)$$

where

$$G_{mn} = \left((1+\delta_m)\frac{2\pi}{2n+1}\frac{(n+m)!}{(n-m)!}\sqrt{n(n+1)}\,\frac{1}{a}\right)^{-2} \times \left(\frac{n(n+1)}{k_2 a}h_n^{(2)}(k_1a)h_n^{(2)}(k_2a) - \frac{k_1}{k_2}h_n^{(2)1}(k_1a)h_n^{(2)}(k_2a) - k_1 a h_n^{(2)1}(k_1a)h_n^{(2)1}(k_2a)\right)^{-2}. \quad (4.32)$$

The second-order and higher-order solutions can be obtained exactly in the same manner.

5. ZEROTH-ORDER SOLUTION IN THE EXTERIOR OF A LARGE SPHERE

The solution obtained in the preceding section is quite general and valid as long as the roughness of the surface satisfies the prescribed conditions. However, for the purpose of numerical evaluations and physical interpretation, these series converge fast enough only for $a < L$, where L is the wave length of the incident wave.

Thus when $k_1 a > 1$, we must seek another representation of the solution. Some years ago, Negase[14] obtained the high-frequency asymptotic solution for the problem of diffraction of elastic waves by a spherical cavity imbedded in an infinite elastic medium. Later, the high-frequency asymptotic solution for the problem of diffraction of elastic waves by a smooth convex rigid cylinder and a smooth convex soft cylinder were derived by Gilbert and Knopoff.[12,15] More recently, Christiansen[16] has obtained the high-frequency asymptotic solution for the case of an elastic cylinder imbedded in a different elastic medium. Since the zeroth-order solution for the case of smooth rigid sphere and its asymptotic solution has not appeared in the existing literature, it is worth the effort to derive it in a complete manner here.

In this paper we shall follow the treatment of Chen,[3] Chen and Fan[4] rather closely. By means of Poisson summation formula [Appendix (2)] and from the relations between spherical Bessel functions and Bessel functions [Appendix (3)], Eqs. (4.11)-(4.14) can be rewritten as

$$\mathbf{u}_{op\,r}(\mathbf{r}) = \left(\frac{\pi}{2k_1 r}\right)^{1/2}\sum_{l=-\infty}^{\infty}\int_0^{\infty} H_\nu^{(2)1}(k_1 r)P_{\nu-1/2}(\cos\theta) \times \exp(-i2\pi l\nu + i\pi l + i\tfrac{1}{2}\pi\nu - i\tfrac{3}{4}\pi)\nu d\nu + \left(\frac{\pi}{2k_1 r}\right)^{1/2}\sum_{l=-\infty}^{\infty}\int_0^{\infty}\frac{N_0}{D_0}\frac{H_\nu^{(1)}(k_1a)H_\nu^{(2)1}(k_1r)}{H_\nu^{(2)}(k_1a)}P_{\nu-1/2}(\cos\theta) \times \exp(-i2\pi l\nu + i\pi l + i\tfrac{1}{2}\pi\nu - i\tfrac{3}{4}\pi)\nu d\nu, \quad (5.1)$$

$$u_{0p\theta}(\mathbf{r}) = \frac{1}{k_1 r}\left(\frac{\pi}{2k_1 r}\right)^{1/2}\sum_{l=-\infty}^{\infty}\int_0^{\infty} H_\nu^{(2)}(k_1 r)\frac{\partial}{\partial\theta}P_{\nu-1/2}(\cos\theta) \times \exp(-i2\pi l\nu + i\pi l + i\tfrac{1}{2}\pi\nu - i\tfrac{3}{4}\pi)\nu d\nu + \frac{1}{k_1 r}\left(\frac{\pi}{2k_1 r}\right)^{1/2}\sum_{l=-\infty}^{\infty}\int_0^{\infty}\frac{N_0}{D_0}\frac{H_\nu^{(1)}(k_1a)H^{(2)}(k_1r)}{H_\nu^{(2)}(k_1a)}\frac{\partial}{\partial\theta} \times P_{\nu-1/2}(\cos\theta)\exp(-i2\pi l\nu + i\pi l + i\pi\nu - i\tfrac{3}{4}\pi)\nu d\nu, \quad (5.2)$$

$$u_{0sr}(\mathbf{r}) = \frac{2}{k_1^3 a^3 k_2 r}\left(\frac{2k_1a^2}{\pi r}\right)^{1/2}\sum_{l=-\infty}^{\infty}\int_0^{\infty}\frac{1}{D_0}\frac{H_\nu^{(2)}(k_2r)}{H_\nu^{(2)}(k_1a)H_\nu^{(2)}(k_2a)} \times P_{\nu-1/2}(\cos\theta)\exp(-i2\pi l\nu + i\pi l + i\tfrac{1}{2}\pi\nu - i\tfrac{1}{4}\pi)\nu d\nu, \quad (5.3)$$

$$u_{0s\theta}(\mathbf{r}) = \frac{2}{k_1^3 a^3}\left(\frac{2k_1a^2}{\pi r}\right)^{1/2}\sum_{l=-\infty}^{\infty}\int_0^{\infty}\frac{1}{D_0}\frac{H_\nu^{(2)1}(k_2r)}{H_\nu^{(2)}(k_1a)H_\nu^{(2)}(k_2a)}\frac{\partial}{\partial\theta} \times P_{\nu-1/2}(\cos\theta)\exp(-i2\pi l\nu + i\pi l + i\tfrac{1}{2}\nu - i\tfrac{1}{4}\pi)\nu d\nu, \quad (5.4)$$

where

$$N_0 = \frac{\nu^2}{k_1 k_2 a_2} - \frac{H_\nu^{(1)\prime}(k_1a)}{H_\nu^{(1)}(k_1a)}\frac{H_\nu^{(2)\prime}(k_2a)}{H_\nu^{(2)}(k_2a)} \quad (5.5)$$

and

$$D_0 = \frac{\nu^2}{k_1 k_2 a^2} - \frac{H_\nu^{(2)\prime}(k_1a)}{H_\nu^{(2)}(k_1a)}\frac{H_\nu^{(2)\prime}(k_2a)}{H_\nu^{(2)}(k_2a)}. \quad (5.6)$$

A. Geometric optics wave

The criterion for the proper identification of the geometric optics wave from the asymptotic evaluation of equations (5.1)-(5.4) is the existence of real saddle points ν_0 such that $0 < \nu_0 < k_1 a$.

Define

$$Q_{\nu-1/2}^{(2)} = \tfrac{1}{2}(P_{\nu-1/2} + i\frac{2}{\pi}Q_{\nu-1/2}) \quad (5.7)$$

and

$$Q_{\nu-1/2}^{(2)} = \tfrac{1}{2}(P_{\nu-1/2} - i\frac{2}{\pi}Q_{\nu-1/2}), \quad (5.8)$$

where $Q_{\nu-1/2}$ is the Legendre function of the second kind. Then we may substitute

$$P_{\nu-1/2} = Q_{\nu-1/2}^{(1)} + Q_{\nu-1/2}^{(2)} \quad (5.9)$$

into (5.1)-(5.4). After using the proper asymptotic forms [Appendix (4), (5)] and the saddle-point method [Appendix (6)], it can easily be shown that there is no real saddle point for terms with $Q_{\nu-1/2}^{(2)}$ and for terms with $Q_{\nu-1/2}$ the real saddle points exist only when $l = 0$.

The corresponding saddle-point equation of the first terms of and $u_{0p\theta}$ is

$$\cos^{-1}(\nu_0/k_1 r) = \theta - (\pi/2). \quad (5.10)$$

It has a unique real solution only for the field point r lying in the shadow region (Fig. 2) and the solution is

$$\nu_0 = k_1 r\sin\theta. \quad (5.11)$$

Upon substituting (5.11) into (5.10), one obtains the correct geometrical relation

$$\theta = \theta. \quad (5.12)$$

The corresponding saddle point equation of the second terms of $u_{op\,r}$ and $u_{op\theta}$ is

$$\cos^{-1}(\nu_0/k_1 r) - 2\cos^{-1}(\nu_0/k_1 a) = \theta - (\pi/2). \quad (5.13)$$

It has a unique real solution only for the field point r lying in the lit region (Fig. 3) and its solution can be expressed in terms of physical entities as

$$\nu_0 = k_1 a \sin \eta_{0p} = k_2 a \sin \lambda_{0p} = k_1 r \sin \xi_{0p}. \quad (5.14)$$

By substituting (5.14) into (5.13), one obtains the correct geometrical relation (Fig. 3) as

$$2\eta_{0p} - \xi_{0p} = \theta. \quad (5.15)$$

The saddle point equation of u_{osr} and $u_{os\theta}$ is

$$\cos^{-1}\frac{\nu_0}{k_2 r} - \cos^{-1}\frac{\nu_0}{k_1 a} - \cos^{-1}\frac{\nu_0}{k_2 a} = \theta - \frac{\pi}{2}. \quad (5.16)$$

It has a unique real solution only for the field point lying in the lit region (Fig. 4) and its solution can be expressed in terms of physical entities as

$$\nu_0 = k_1 a \sin \eta_{0s} = k_2 a \sin \lambda_{0s} = k_2 r \sin \zeta_{0s}. \quad (5.17)$$

By substituting (5.17) into (5.16), one obtains the correct geometrical relation (Fig. 4) as

$$\eta_{0s} + \lambda_{0s} - \zeta_{0s} = \theta. \quad (5.18)$$

The saddle-point path C is shown in Fig. 5. It is found that the end-point contribution of integrals is asymptotically small in comparison with the saddle-point contribution; hence it can be neglected. Finally, the saddle-point contributions of each component of u_{so} are

$$u^G_{0pr}(\mathbf{r}) \approx \cos \xi_{0p} D_{0p} R_{0p} \exp\left[-ik_1(r\cos \xi_{0p} - 2a\cos\eta_{0p})\right]$$
in the lit region,
$$\approx \cos\theta \exp[ik_1 r\cos\theta] \text{ in the shadow region,} \quad (5.19)$$

$$u^G_{0p\theta}(\mathbf{r}) \simeq \sin \xi_{0p} D_{0p} R_{0p} \exp\left[-ik_1(r\cos \xi_{0p} - 2a\cos \eta_{0p})\right]$$
in the lit region,
$$\simeq -\sin\theta \exp[ik_1 r\cos\theta] \quad \text{in the shadow region,} \quad (5.20)$$

$$u^G_{0sr}(\mathbf{r}) \simeq N\sin \zeta_{0s} D_{0s} R_{0s} \times \exp\left[-ik_2(r\cos \zeta_{0s} - a\cos \lambda_{0s} - (1/N)\, a\cos \eta_{0s})\right]$$
in the lit region,
$$\simeq 0 \quad \text{in the shadow region,} \quad (5.21)$$

$$u^G_{0s\theta}(\mathbf{r}) \approx -N\cos \zeta_{0s} D_{0s} R_{0s} \times \exp\left[-ik_2(r\cos \zeta_{0s} - a\cos \lambda_{0s} - (1/N)\, a\cos \eta_{0s})\right]$$
in the lit region,
$$\simeq 0 \quad \text{in the shadow region,} \quad (5.22)$$

with

$$D_{0p} = \left(\frac{a^2 \cos \eta_{0p} \sin \eta_{0p}}{r \sin\theta (2r\cos \xi_{0p} - a\cos \eta_{0p})}\right)^{1/2}, \quad (5.23)$$

$$R_{0p} = \frac{\cos(\eta_{0p} + \lambda_{0p})}{\cos(\eta_{0p} - \lambda_{0p})}, \quad (5.24)$$

$$D_{0s} = \left(\frac{a^2 \cos \lambda_{0s} \sin \eta_{0s}}{r \sin\theta [r\cos \zeta_{0s} + (\cos \eta_{0s}/N\cos \lambda_{0s})(r\cos \zeta_{0s} - a\cos \lambda_{0s})]}\right)^{1/2}, \quad (5.25)$$

$$R_{0s} = \frac{2\cos \eta_{0s} \sin \lambda_{0s}}{\cos(\eta_{0s} - \lambda_{0s})}, \quad (5.26)$$

$$N = (k_2/k_1) > 1, \quad (5.27)$$

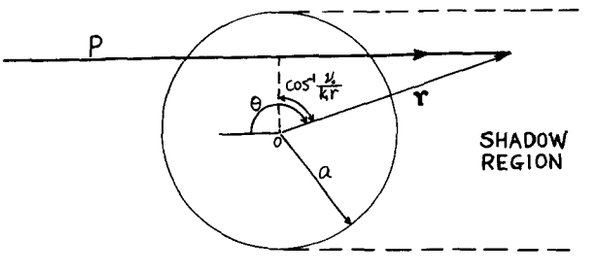


FIG. 2. The geometry of Eq. (5.10) is shown.

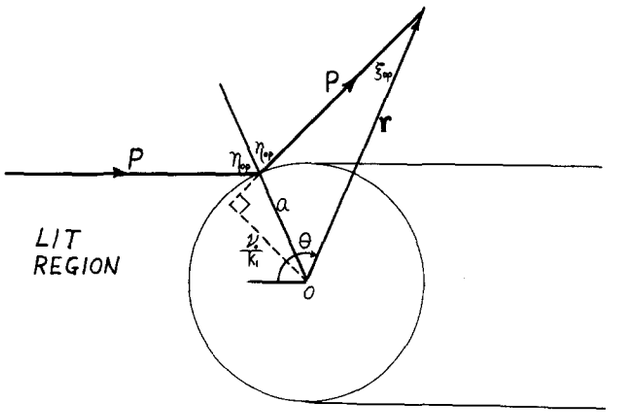


FIG. 3. The geometry of Eq. (5.13) is shown.

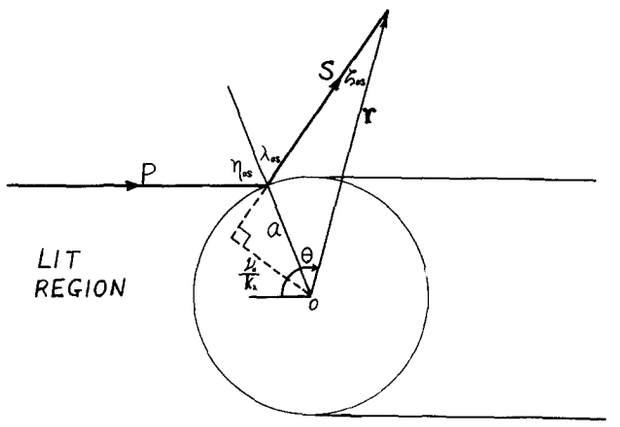


FIG. 4. The geometry of Eq. (5.16) is shown.

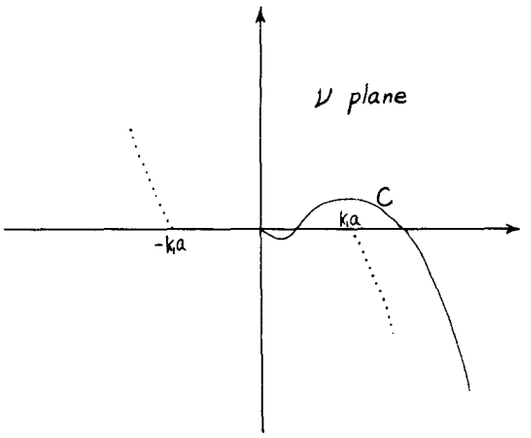


FIG. 5. The saddle-point path C is shown schematically.

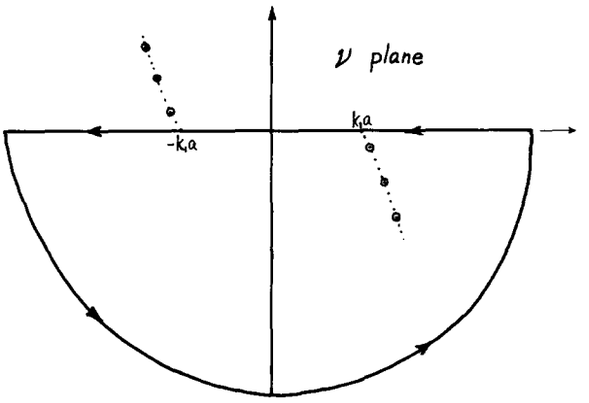


FIG. 6. This figure shows schematically the positions of the poles of the integrand of (5.31)–(5.34), as well as the negative path of integration in the ν plane.

where D_{0p} and R_{0p} are the divergence factor and the reflection coefficient[10] of the zeroth-order reflected compression wave, respectively; D_{0s} and R_{0s} are the divergence factor and the reflection coefficient of the zeroth-order reflected shear wave, respectively. Hence from (4.10), (5.19)-(5.22), we have

$$\begin{aligned} u_{s0}(\mathbf{r}) &\approx (\cos\zeta_{0p}\mathbf{1}_r + \sin\xi_{0p}\mathbf{1}_\theta)D_{0p}R_{0p} \\ &\quad\times \exp[-ik_1(r\cos\xi_{0p} - 2a\cos\eta_{0p})] \\ &\quad + N(\sin\zeta_{0s}\mathbf{1}_r - \cos\zeta_{0s}\mathbf{1}_\theta)D_{0s}R_{0s} \\ &\quad\times \exp[-ik_2(r\cos\zeta_{0s} - a\cos\lambda_{0s} - (a/N)\cos\eta_{0s})] \\ &\qquad\qquad \text{in the lit region,} \qquad (5.28) \\ &\approx (\cos\theta\mathbf{1}_r - \sin\theta\mathbf{1}_\theta)\exp[ik_1 r\cos\theta] \\ &\qquad\qquad \text{in the shadow region.} \end{aligned}$$

The unit plane compression incident wave, which propagates along the negative direction of Z axis, has the simple form

$$\begin{aligned} \mathbf{u}_i(\mathbf{r}) &= -\mathbf{1}_Z\exp(ik_1 z) \\ &= (-\cos\theta\mathbf{1}_r + \sin\mathbf{1}_\theta)\exp(ik_1 r\cos\theta). \qquad (5.29) \end{aligned}$$

Finally, combination of (5.28) and (5.29) yields the total zeroth-order geometric optics wave,

$$\begin{aligned} \mathbf{u}_i(\mathbf{r}) + \mathbf{u}^G_{s0}(\mathbf{r}) &\approx (-\cos\theta\mathbf{1}_r + \sin\theta\mathbf{1}_\theta)\exp[ik_1 r\cos\theta] \\ &\quad + (\cos\xi_{0p}\mathbf{1}_r + \sin\xi_{0p}\mathbf{1}_\theta)D_{0p}R_{0p} \\ &\quad\times \exp[-ik_1(r\cos\xi_{0p} - 2a\cos\eta_{0p})] \\ &\quad + (\sin\zeta_{0s}\mathbf{1}_r - \cos\zeta_{0s}\mathbf{1}_\theta)ND_{0s}R_{0s} \\ &\quad\times \exp[-ik_2(r\cos\zeta_{0s} - a\cos\lambda_{0s} - (a/N)\cos\eta_{0s})] \\ &\qquad\qquad \text{in the lit region,} \qquad (5.30) \\ &\approx 0 \quad \text{in the shadow region.} \end{aligned}$$

B. Diffracted Wave

Contributions to the integrals (5.1)–(5.4) also arise from the poles of the integrands. In order to make the asymptotic evaluation, we first extend the integrals along the positive real ν axis to the entire real ν axis.

It is easy to see that D_0 has zeros of first order, which in turn, are the simple poles of the integrands. With the aid of the Appendix (8), the approximate positions of the poles ν_λ are determined by the equation

$$\frac{\nu_\lambda^2}{k_1k_2a^2} - i\left(1 - \frac{\nu_\lambda^2}{k_2^2a^2}\right)^{1/2}\left(\frac{6}{k_1a}\right)^{1/3} e^{i\pi/3}\,\frac{\mathrm{Ai}'(t_\lambda)}{\mathrm{Ai}(t_\lambda)} = 0 \qquad (5.31)$$

and have the form

$$\nu_\lambda = k_1a + t_\lambda[(k_1a/6)]^{1/3}e^{-i\pi/3} + \cdots \qquad (5.32)$$

where $\mathrm{Ai}(t_\lambda)$ is the Airy functions of t_λ.

After closing the contour in the lower half complex ν plane (Fig. 6), evaluation of the residues and neglecting terms of $O[(k_ia)^{-1}]$, $i = 1, 2$, we obtain the zeroth-order diffracted wave,

$$\begin{aligned} \mathbf{u}^D_{s0}(\mathbf{r}) &\approx \sum_{\nu_\lambda}\Bigg\{\left[\left(1 - \frac{a^2}{r^2}\right)^{1/2}\mathbf{1}_r + \frac{a}{r}\mathbf{1}_\theta\right] \\ &\quad\times \exp[-i\nu_\lambda(2\pi + \theta - \tfrac{1}{2}\pi) + i\tfrac{3}{2}\pi] \\ &\quad + \left[\left(1 - \frac{a^2}{r^2}\right)^{1/2}\mathbf{1}_r - \frac{a}{r}\mathbf{1}_\theta\right]\exp[-i\nu_\lambda(2\pi - \theta - \tfrac{1}{2}\pi) + i\tfrac{1}{2}\pi]\Bigg\} \\ &\quad\times \mathfrak{D}_{0p}\left(\frac{a}{r^2\sin\theta}\right)^{1/2}\left(1 - \frac{a^2}{r^2}\right)^{-1/4}(1 + e^{-i2\pi\nu_\lambda})^{-1} \\ &\quad\times \exp\left[-ik_1(r^2 - a^2)^{1/2} + i\nu_\lambda\cos^{-1}\frac{a}{r}\right] \\ &\quad + \sum_{\nu_\lambda}\Bigg\{\left[\frac{a}{Nr}\mathbf{1}_r + \left(1 - \frac{a^2}{N^2r^2}\right)^{1/2}\mathbf{1}_\theta\right]\exp[-i\nu_\lambda(2\pi + \theta - \tfrac{1}{2}\pi)] \\ &\quad + \left[\frac{a}{Nr}\mathbf{1}_r - \left(1 - \frac{a^2}{N^2r^2}\right)^{1/2}\mathbf{1}_\theta\right] \\ &\quad\times \exp[-i\nu_\lambda(2\pi - \theta - \tfrac{1}{2}\pi) - i\tfrac{1}{2}\pi]\Bigg\} \\ &\quad\times \mathfrak{D}_{0s}\left(\frac{a}{r^2\sin\theta}\right)^{1/2}\left(1 - \frac{a^2}{N^2r^2}\right)^{-1/4}(1 + e^{-i2\pi\nu_\lambda})^{-1} \\ &\quad\times \exp\left[-ik_2\left(r^2 - \frac{a^2}{N^2}\right)^{1/2} + ik_2\left(a^2 - \frac{a^2}{N^2}\right)^{1/2}\right. \\ &\quad \left. + i\nu_\lambda\cos^{-1}\frac{a}{Nr} - i\nu_\lambda\cos^{-1}\frac{1}{N}\right] \\ &\qquad\qquad \text{in the lit region} \end{aligned}$$

$$\simeq \sum_{\nu_\lambda}\left\{\left[\left(1-\frac{a^2}{r^2}\right)^{1/2}\mathbf{1}_r+\frac{a}{r}\,\mathbf{1}_\theta\right]\exp\left[-i\nu_\lambda\left(\theta-\frac{\pi}{2}\right)+i\frac{\pi}{2}\right]\right.$$
$$+\left[\left(1-\frac{a^2}{r^2}\right)^{1/2}\mathbf{1}_r-\frac{a}{r}\,\mathbf{1}_\theta\right]\exp\left[-i\nu_\lambda\left(2\pi-\theta-\frac{\pi}{2}\right)+i\pi\right]$$
$$\times \mathfrak{D}_{0p}\left(\frac{a}{r^2\sin\theta}\right)^{1/2}\left(1-\frac{a^2}{r^2}\right)^{-1/4}(1+e^{-i2\pi\nu_\lambda})^{-1}$$
$$\times\exp\left[-ik_1(r^2-a^2)^{1/2}+i\nu_\lambda\cos^{-1}\frac{a}{r}\right]$$
$$+\sum_{\nu_\lambda}\left\{\left[\left(\frac{a}{Nr}\mathbf{1}_r+\left(1-\frac{a^2}{N^2r^2}\right)^{1/2}\mathbf{1}_\theta\right]\exp\left[-i\nu_\lambda\left(\theta-\frac{\pi}{2}\right)+i\pi\right]\right.\right.$$
$$+\left[\frac{a}{Nr}\mathbf{1}_r-\left(1-\frac{a^2}{N^2r^2}\right)^{1/2}\mathbf{1}_\theta\right]$$
$$\times\exp\left[-i\nu_\lambda\left(2\pi-\theta-\frac{\pi}{2}\right)+i\frac{3}{2}\pi\right]\Bigg\}$$
$$\times\mathfrak{D}_{0s}\left(\frac{a}{r^2\sin\theta}\right)^{1/2}\left(1-\frac{a^2}{N^2r^2}\right)^{-1/4}(1+e^{-i2\pi\nu_\lambda})^{-1}$$
$$\times\exp\left[-ik_2\left(r^2-\frac{a^2}{N^2}\right)^{1/2}+ik_2\left(a^2-\frac{a^2}{N^2}\right)^{1/2}\right.$$
$$\left.+i\nu_\lambda\cos^{-1}\frac{a}{Nr}-i\nu_\lambda\cos^{-1}\frac{1}{N}\right]\quad\text{in the shadow region,}\tag{5.33}$$

where the coefficients $\mathfrak{D}_{0p}$ and $\mathfrak{D}_{0s}$ are defined as

$$\mathfrak{D}_{0p}=\tfrac{1}{2}(N^2-1)^{1/2}(2\pi/k_1)^{1/2}\left[\mathrm{Ai}\left(t_\lambda e^{-i\frac{2\pi}{3}}\right)\Big/\mathrm{Ai}(t_\lambda)\right]e^{-i\frac{5}{12}\pi}\tag{5.34}$$

for the zeroth-order diffracted P-wace, and

$$\mathfrak{D}_{0s}=2\pi N^2\left(1-\frac{1}{N^2}\right)^{3/4}\left(\frac{1}{k_1a}\right)^{1/2}\left(\frac{k_1a}{6}\right)^{1/3}\frac{1}{\mathrm{Ai}(t_\lambda)}\,e^{-i\frac{\pi}{3}}\tag{5.35}$$

for the zeroth-order diffracted S wave.

The detailed physical interpretation of the zeroth-order diffracted wave has been given for scalar and vector waves before.[11,12] Hence only few simple geometric interpretation are shown in Figs. 7 and 8.

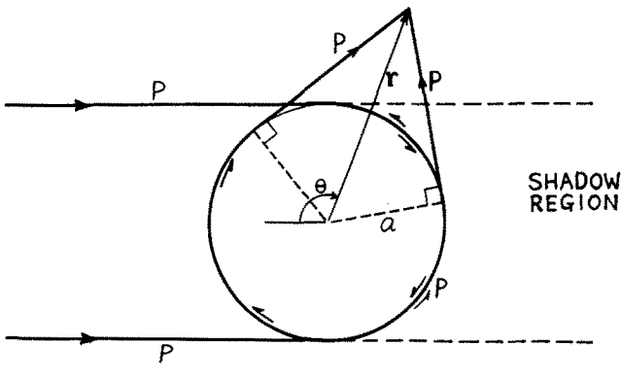


FIG. 7. P-P-P ray paths.

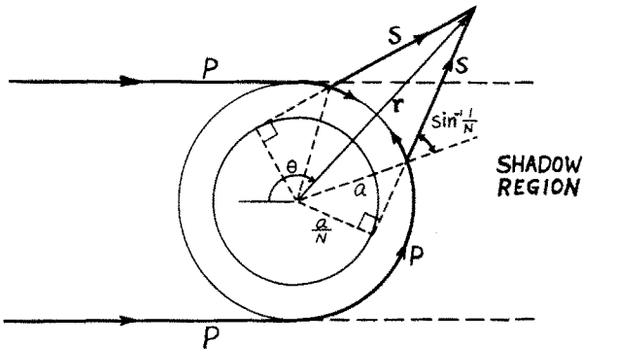


FIG. 8. P-P-S ray paths.

6. FIRST ORDER SOLUTION IN THE EXTERIOR OF A LARGE ROUGH SPHERE

Upon substituting (4.26)–(4.31) into (4.17), we have

$$\mathbf{u}_{s1}(\mathbf{r},g)=\mathbf{u}_{11}(\mathbf{r},g)+\mathbf{u}_{12}(\mathbf{r},g)+\mathbf{u}_{13}(\mathbf{r},g)+\mathbf{u}_{14}(\mathbf{r},g)+\mathbf{u}_{15}(\mathbf{r},g),\tag{6.1}$$

where

$$\mathbf{u}_{11}(\mathbf{r},g)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}G_{mn}\|\mathbf{n}\|^{2}[\langle\mathbf{S}_1\cdot\mathbf{l}_{emn}(\mathbf{r}_0)\rangle\mathbf{l}_{emn}(\mathbf{r})+\langle\mathbf{S}_1\cdot\mathbf{l}_{emn}(\mathbf{r}_0)\rangle\mathbf{l}_{omn}(\mathbf{r})],\tag{6.2}$$

$$\mathbf{u}_{12}(\mathbf{r},g)=-\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}G_{mn}\langle\mathbf{l}\cdot\mathbf{n}\rangle[\langle\mathbf{S}_1\cdot\mathbf{n}_{emn}(\mathbf{r}_0)\rangle\mathbf{l}_{emn}(\mathbf{r})+\langle\mathbf{S}_1\cdot\mathbf{n}_{omn}(\mathbf{r}_0)\rangle\mathbf{l}_{omn}(\mathbf{r})],\tag{6.3}$$

$$\mathbf{u}_{13}(\mathbf{r},g)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\|\mathbf{m}\|^{-2}[\langle\mathbf{S}_1\cdot\mathbf{m}_{emn}(\mathbf{r}_0)\rangle\mathbf{m}_{emn}(\mathbf{r})+\langle\mathbf{S}_1\cdot\mathbf{m}_{omn}(\mathbf{r}_0)\rangle\mathbf{m}_{omn}(\mathbf{r})],\tag{6.4}$$

$$\mathbf{u}_{14}(\mathbf{r},g)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}G_{mn}\|\mathbf{l}\|^{2}[\langle\mathbf{S}_1\cdot\mathbf{n}_{emn}(\mathbf{r}_0)\rangle\mathbf{n}_{emn}(\mathbf{r})+\langle\mathbf{S}_1\cdot\mathbf{n}_{omn}(\mathbf{r}_0)\rangle\mathbf{n}_{omn}(\mathbf{r})],\tag{6.5}$$

$$\mathbf{u}_{15}(\mathbf{r},g)=-\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}G_{mn}\langle\mathbf{l}\cdot\mathbf{n}\rangle[\langle\mathbf{S}_1\cdot\mathbf{l}_{emn}(\mathbf{r}_0)\rangle\mathbf{n}_{emn}(\mathbf{r})+\langle\mathbf{S}_1\cdot\mathbf{l}_{omn}(\mathbf{r}_0)\rangle\mathbf{n}_{omn}(\mathbf{r})].\tag{6.6}$$

Here, the decomposition of $\mathbf{u}_{s1}$ is purely for the convenience in calculation.

A. Geometric optics wave

The criterion for the proper identification of the geometric optics wave from the asymptotic evaluation of (6.1) is the same as that of the zeroth-order solution.

Upon substituting (4.32) into (6.2) and utilizing the orthogonality relations of $\mathbf{l}$, $\mathbf{m}$, $\mathbf{n}$ and the addition formula for the Legendre polynomials [Appendix (9)], we obtain

$$\mathbf{u}_{11}(\mathbf{r},g)=\int_{\omega'}[(I_{l'r'lr}+I_{l'\theta'lr})\mathbf{1}_r+(I_{l'r'l\theta}+I_{l'\theta'l\theta})\mathbf{1}_\theta+(I_{l'r'l\varphi}+I_{l'\theta'l\varphi})\mathbf{1}_\varphi]d\omega',\tag{6.7}$$

where the subscripts of $I_{a'b'cd}$ indicate the origin of the terms, with $a'b'cd$ standing for the product composed of the scalar product of $\mathbf{s}_1$ with the b' component of a' and the d component of c;

$$\int_{\omega'}\cdots d\omega'\equiv\int_0^{2\pi}\int_0^{\pi}\cdots\sin\theta'd\theta'd\varphi',\tag{6.8}$$

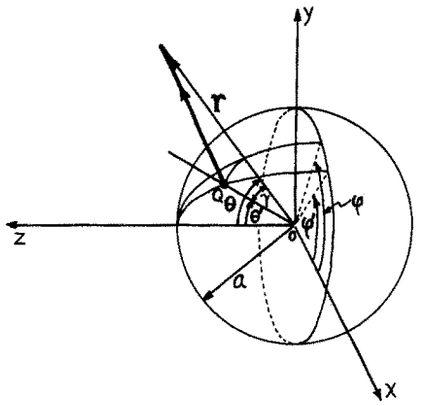


FIG. 9. The geometry of the angle γ is shown.

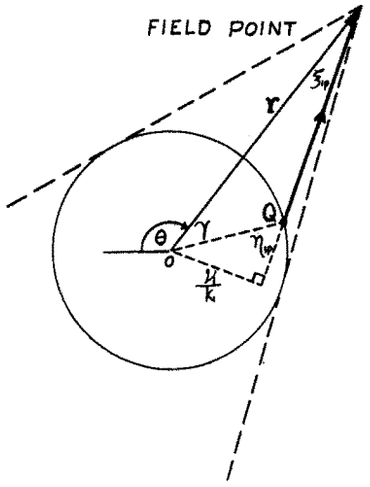


FIG. 10 The geometric ray from the equivalent source at a point Q on the surface to the field point is shown, and so is the geometrical relation between γ and η_{1p}.

with (a, θ', φ') being a point on the spherical surface:

$$I_{l'r'lr} = \sum_{n=0}^{\infty} s_{1r'} k_1^2 h_n^{(2)'}(k_1 a) h_n^{(2)'}(k_1 r) H_{1n} P_n(\cos\gamma), \tag{6.9}$$

$$I_{l'\theta'lr} = \sum_{n=0}^{\infty} s_{1\theta'} \frac{\partial\gamma}{\partial\theta'} \frac{k_1}{a} h_n^{(2)}(k_1 a) h_n^{(2)'}(k_1 r) H_{1n} \frac{\partial}{\partial\gamma} P_n(\cos\gamma), \tag{6.10}$$

$$I_{l'r'l\theta} = \sum_{n=0}^{\infty} s_{1r'} \frac{\partial\gamma}{\partial\theta} \frac{k_1}{r} h_n^{(2)'}(k_1 a) h_n^{(2)}(k_1 r) H_{1n} \frac{\partial}{\partial\gamma} P_n(\cos\gamma), \tag{6.11}$$

$$I_{l'\theta'l\theta} = \sum_{n=0}^{\infty} s_{1\theta'} \frac{\partial\gamma}{\partial\theta} \frac{\partial\gamma}{\partial\theta'} \frac{1}{ar} h_n^{(2)}(k_1 a) h_n^{(2)}(k_1 r) H_{1n} \frac{\partial^2}{\partial\gamma^2} (\cos\gamma), \tag{6.12}$$

$$I_{l'r'l\varphi} = \frac{1}{\sin\theta} \left(\frac{\partial\gamma}{\partial\varphi} \Big/ \frac{\partial\gamma}{\partial\theta} \right) I_{l'r'l\theta}, \tag{6.13}$$

and

$$I_{l'\theta'l\varphi} = \frac{1}{\sin\theta} \left(\frac{\partial\gamma}{\partial\varphi} \Big/ \frac{\partial\gamma}{\partial\theta} \right) I_{l'\theta'l\theta}, \tag{6.14}$$

with the spherical angle (Fig. 9) defined by

$$\cos\gamma \equiv \sin\theta \sin\theta' \cos(\varphi - \varphi') + \cos\theta \cos\theta', \tag{6.15}$$

$$H_{1n} = \frac{[n + (1/2)]}{2\pi} a^2 \left(\frac{n^2 + n + 1}{k_2^2 a^2} [h_n^{(2)}(k_2 a)]^2 + \frac{2}{k_2 a} h_n^{(2)}(k_2 a) h_n^{(2)'}(k_2 a) + [h_n^{(2)'}(k_2 a)]^2 \right) \times \left(\frac{n(n+1)}{k_2 a} h_n^{(2)}(k_1 a) h_n^{(2)}(k_2 a) - \frac{k_1}{k_2} h_n^{(2)'}(k_1 a) h_n^{(2)}(k_2 a) - k_1 a h_n^{(2)'}(k_1 a) h_n^{(2)'}(k_2 a) \right)^{-2} \tag{6.16}$$

and $s_{1r'}$ and $s_{1\theta'}$ being the components of $\mathbf{s}_1$ in the direction of $\mathbf{1}_{r'}$ and $\mathbf{1}_{\theta'}$, respectively.

To find the contribution of the waves radiated directly from the equivalent $\mathbf{s}_1$ [as a point (a, θ', φ') on the spherical surface] to the field point $\mathbf{r}$, we have to evaluate I's asymptotically by the saddle-point method. After using the proper asymptotic forms, it is found that all I's have a proper real saddle-point only when (a, θ', φ') is in the lit region of $\mathbf{r}$ and $t = 0$. The saddle-point equation is

$$\cos^{-1}\frac{\nu_1}{k_1 r} - \cos^{-1}\frac{\nu_1}{k_1 a} = \gamma. \tag{6.17}$$

Its solution can be expressed in terms of physical entities as

$$\nu_1 = k_1 a \sin\eta_{1p} = k_2 a \sin\lambda_{1p} = k_1 r \sin\xi_{1p} \tag{6.18}$$

and then (6.17) becomes

$$\eta_{1p} - \xi_{1p} = \gamma, \tag{6.19}$$

which gives the correct geometrical relation (Fig. 10).

Since the end-point contribution of I's is asymptotically small in comparison with the saddle-point contribution, the major contributions of I's are

$$I^G_{l'r'lr} \simeq D_{1p} R_{1pl} s_{1r'} \cos\xi_{1p} \cos\eta_{1p} \exp[-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})], \tag{6.20}$$

$$I^G_{l'\theta'lr} \simeq D_{1p} R_{1pl} s_{1\theta'} \frac{\partial\gamma}{\partial\theta'} \cos\xi_{1p} \sin\eta_{1p} \times \exp[-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})], \tag{6.21}$$

$$I^G_{l'r'l\theta} \simeq D_{1p} R_{1pl} s_{1r'} \frac{\partial\gamma}{\partial\theta} \sin\xi_{1p} \cos\eta_{1p} \times \exp[-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})], \tag{6.22}$$

$$I^G_{l'\theta'l\theta} \simeq D_{1p} R_{1pl} s_{1\theta'} \frac{\partial\gamma}{\partial\theta} \frac{\partial\gamma}{\partial\theta'} \sin\xi_{1p} \sin\eta_{1p} \times \exp[-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})], \tag{6.23}$$

$$I^G_{l'r'l\varphi} = \frac{1}{\sin\theta} \left(\frac{\partial\gamma}{\partial\varphi} \Big/ \frac{\partial\gamma}{\partial\theta} \right) I^G_{l'r'l\theta} \tag{6.24}$$

and

$$I^G_{l'\theta'l\varphi} = \frac{1}{\sin\theta} \left(\frac{\partial\gamma}{\partial\varphi} \Big/ \frac{\partial\gamma}{\partial\theta} \right) I^G_{l'\theta'l\theta}. \tag{6.25}$$

Hence,

$$\mathbf{u}^G_{11}(\mathbf{r}, g) \simeq \int\int_{\mathcal{L}} D_{1p} R_{1pl} \left(s_{1r'} \cos\eta_{1p} + s_{1\theta'} \frac{\partial\gamma}{\partial\theta'} \sin\eta_{1p} \right) \times \left(\cos\xi_{1p} \mathbf{1}_r + \sin\xi_{1p} \frac{\partial\gamma}{\partial\theta} \mathbf{1}_\theta + \frac{\sin\xi_{1p}}{\sin\theta} \frac{\partial\gamma}{\partial\varphi} \mathbf{1}_\varphi \right) \times \exp[-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})] d\omega', \tag{6.26}$$

where $\mathcal{L}$ means that the integration is carried out only on the portion of spherical surface in the lit region of r:

$$D_{1p} = \frac{ik_1 a}{2\pi}\left[\frac{a^2 \sin\eta_{1p}}{r\cos\gamma(r\cos\xi_{1p} - a\cos\eta_{1p})}\right]^{1/2} \quad (6.27)$$

and

$$R_{1pl} = \frac{\cos\eta_{1p}\cos 2\lambda_{1p}}{\cos^2(\eta_{1p} - \lambda_{1p})}. \quad (6.28)$$

Similarly, we have

$$\mathbf{u}^G_{12}(\mathbf{r},g) \simeq \iint_{\mathcal{L}} D_{1p}R_{1pn}\left(s_{1r'}\cos\lambda_{1p} - s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\sin\lambda_{1p}\right) \times \left(\cos\xi_{1p}\mathbf{1}_r + \sin\xi_{1p}\frac{\partial\gamma}{\partial\theta}\mathbf{1}_\theta + \frac{\sin\xi_{1p}}{\sin\theta}\frac{\partial\gamma}{\partial\varphi}\mathbf{1}_\varphi\right) \times \exp(-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})]d\omega', \quad (6.29)$$

where R_{1pn} is defined as

$$R_{1pn} = \frac{\cos\eta_{1p}\sin(\eta_{1p} + \lambda_{1p})}{\cos^2(\eta_{1p} - \lambda_{1p})}. \quad (6.30)$$

In a similar manner, we obtain the saddle-point equation of $\mathbf{u}^G_{13}(\mathbf{r}, q)$,

$$\lambda_{1s} - \zeta_{1s} = \gamma, \quad (6.31)$$

where the angles λ_{1s} and ζ_{1s} satisfy the equation

$$k_2 a\sin\lambda_{1s} = k_2 r\sin\zeta_{1s}. \quad (6.32)$$

And the solution is

$$\mathbf{u}^G_{13}(\mathbf{r},g) \simeq \iint_{\mathcal{L}} D_{1s}s_{1\theta'}\frac{\partial\gamma}{\partial\varphi'}\frac{\cos\lambda_{1s}}{\sin\theta}\left(-\frac{\partial\gamma}{\partial\varphi}\frac{1}{\sin\theta}\mathbf{1}_\theta + \frac{\partial\gamma}{\partial\theta}\mathbf{1}_\varphi\right) \exp[-ik_2(r\cos\zeta_{1s} - a\cos\lambda_{1s})]d\omega', \quad (6.33)$$

where the coefficient D_{1s} is defined as

$$D_{1s} = \frac{ik_2 a}{2\pi}\left(\frac{a^2\sin\lambda_{1s}}{r\sin\gamma(r\cos\zeta_{1s} - a\cos\lambda_{1s})}\right)^{1/2}. \quad (6.34)$$

In a similar manner, we obtain

$$\mathbf{u}^G_{14}(\mathbf{r},g) \simeq \iint_{\mathcal{L}} D_{1s}R_{1sn}\left(s_{1r'}\sin\lambda_{1s} - s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\cos\lambda_{1s}\right) \times \left(-\sin\zeta_{1s}\mathbf{1}_r + \cos\zeta_{1s}\frac{\partial\gamma}{\partial\theta}\mathbf{1}_\theta + \frac{\cos\zeta_{1s}}{\sin\theta}\frac{\partial\gamma}{\partial\varphi}\mathbf{1}_\varphi\right) \times \exp(-ik_2(r\cos\zeta_{1s} - a\cos\lambda_{1s})]d\omega', \quad (6.35)$$

where the coefficient R_{1sn} is defined as

$$R_{1sn} = \frac{\cos 2\eta_{1s}\cos\lambda_{1s}}{\cos^2(\eta_{1s} - \lambda_{1s})}. \quad (6.36)$$

Finally, $\mathbf{u}^G_{15}$ is obtained as

$$\mathbf{u}^G_{15}(\mathbf{r},g) \simeq \iint_{\mathcal{L}} D_{1s}R_{1sl}\left(s_{1r'}\cos\eta_{1s} + s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\sin\eta_{1s}\right) \times (\sin\zeta_{1s}\mathbf{1}_r - \cos\zeta_{1s}\frac{\partial\gamma}{\partial\theta}\mathbf{1}_\theta - \frac{\cos\zeta_{1s}}{\sin\theta}\frac{\partial\gamma}{\partial\varphi}\mathbf{1}_\varphi \times \exp[-ik_2(r\cos\zeta_{1s} - a\cos\lambda_{1s})]d\omega', \quad (6.37)$$

where the coefficient R_{1sl} is defined as

$$R_{1sl} = \frac{\cos\lambda_{1s}\sin(\eta_{1s} + \lambda_{1s})}{\cos^2(\eta_{1s} - \lambda_{1s})}. \quad (6.38)$$

Upon substituting (6.26), (6.29), (6.33), (6.35) and (6.37) into (6.1), we obtain the first-order geometric optics elastic wave,

$$\begin{aligned}\mathbf{u}^G_{s1}(\mathbf{r},g) \simeq{}& \iint_{\mathcal{L}} D_{1p}\Big(R_{1pl}s_{1r'}\cos\eta_{1p} + R_{1pl}s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\sin\eta_{1p} \\ &+ R_{1pn}s_{1r'}\cos\lambda_{1p} - R_{1pn}s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\sin\lambda_{1p}\Big) \\ &\times\left(\cos\xi_{1p}\mathbf{1}_r + \sin\xi_{1p}\frac{\partial\gamma}{\partial\theta}\mathbf{1}_\theta + \frac{\sin\xi_{1p}}{\sin\theta}\frac{\partial\gamma}{\partial\varphi}\mathbf{1}_\varphi\right) \\ &\times\exp[-ik_1(r\cos\xi_{1p} - a\cos\eta_{1p})]d\omega' \\ &+ \iint_{\mathcal{L}} D_{1s}s_{1\theta'}\frac{\partial\gamma}{\partial\varphi'}\frac{\cos\lambda_{1s}}{\sin\theta}\left(-\frac{\partial\gamma}{\partial\varphi}\frac{1}{\sin\theta}\mathbf{1}_\theta + \frac{\partial\gamma}{\partial\theta}\mathbf{1}_\varphi\right) \\ &\times\exp[-ik_2(r\cos\zeta_{1s} - a\cos\lambda_{1s})]d\omega' \\ &+ \iint_{\mathcal{L}} D_{1s}\Big(R_{1sl}s_{1r'}\cos\eta_{1s} + R_{1sl}s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\sin\eta_{1s} \\ &- R_{1sn}s_{1r'}\sin\lambda_{1s} + R_{1sn}s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}\cos\lambda_{1s}\Big) \\ &\times\left(\sin\zeta_{1s}\mathbf{1}_r - \cos\zeta_{1s}\frac{\partial\gamma}{\partial\theta}\mathbf{1}_\theta - \frac{\cos\zeta_{1s}}{\sin\theta}\frac{\partial\gamma}{\partial\varphi}\mathbf{1}_\varphi\right) \\ &\times\exp[-ik_2(r\cos\zeta_{1s} - a\cos\lambda_{1s})]d\omega'. \end{aligned} \quad (6.39)$$

$\mathbf{u}^G_{s1}$ represents mainly the total contribution from the elastic waves radiated directly from the equivalent source $\mathbf{s}_1$ everywhere on the portion of the spherical surface which is in the lit region of $\mathbf{r}$ to the field point $\mathbf{r}$. Its surface integral will be asymptotically evaluated and the corrections to reflection coefficients of the zeroth-order solution in far field region will be derived in a sequel to this paper.

B. Diffracted wave

Similar to the zeroth-order solution, contributions to the integrals (6.2)–(6.6) also arise from the poles of the integrands. In order to make the asymptotic evaluations, we again extend the integrals along the positive real ν axis to the entire real ν axis.

Note that in the integral with the domain of integration $\mathcal{L}$, the lit region of $\mathbf{r}$, we have to subtract the term ($t = 0$) which gives the geometric optics wave; whereas in the integral with the domain of integration $\mathcal{L}'$, the shadow region of $\mathbf{r}$, the summation remains the same.

Since $D_1 = D_0^2$, the positions of the poles are exactly the same as those of the zeroth-order diffracted wave, determined by (5.35) and (5.36) and they are poles of second order.

After closing the contour in the lower half complex ν plane, and evaluation the residues with poles of second

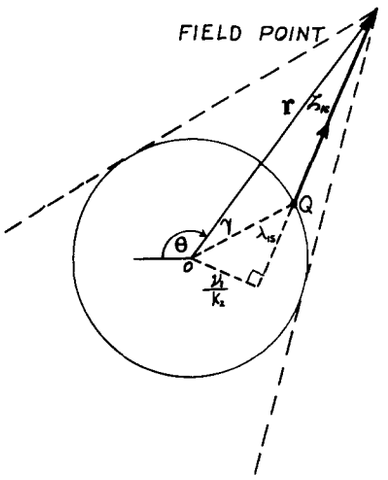


FIG. 11. The geometric ray from the equivalent source at a point Q on the surface to the field point is shown, and so is the geometrical relation between γ and λ_{1s}.

order, we obtain the first-order diffracted wave

$$\mathbf{u}_{s1}^{D}(\mathbf{r},g) \simeq -\sum_{\nu_\lambda}\iint_{\omega'}\Bigg\{\frac{i}{N^2-2}\left(J_p-\frac{a}{r}J_s\right)\times\left(s_{1r'}\sqrt{N^2-1}\,E_m+s_{1\theta'}\frac{\partial\gamma}{\partial\theta'}E_p\right)\mathbf{1}_r$$
$$+\Bigg[\left(\frac{2a}{r}k_{pm}J_p-i\frac{a}{r}\frac{\sqrt{N^2-1}}{N^2-2}J_p+i\frac{\sqrt{N^2-1}}{N^2-2}\sqrt{N^2-\frac{a^2}{r^2}}J_s\right)\times s_{1r'}E_p\frac{\partial\gamma}{\partial\theta}+\left(\frac{2a}{r}\sqrt{N^2-1}J_p k_{pp}-i\frac{a}{r}\frac{2N^2-3}{N^2-2}J_p+i\frac{\sqrt{N^2-a^2/r^2}}{N^2-2}J_s\right)s_{1\theta'}E_m\frac{\partial\gamma}{\partial\theta}\frac{\partial\gamma}{\partial\theta'}\Bigg]\mathbf{1}_\theta$$
$$+\frac{1}{\sin\theta}\Bigg[\left(\frac{2a}{r}k_{pm}J_p-i\frac{a}{r}\frac{\sqrt{N^2-1}}{N^2-2}J_p+i\frac{\sqrt{N^2-1}}{N^2-2}\sqrt{N^2-\frac{a^2}{r^2}}J_s\right)\times s_{1r'}E_p\frac{\partial\gamma}{\partial\varphi}+\left(\frac{2a}{r}\sqrt{N^2-1}J_p k_{pp}-i\frac{a}{r}\frac{2N^2-3}{N^2-2}J_p+i\frac{\sqrt{N^2-a^2/r^2}}{N^2-2}J_s\right)s_{1\theta'}E_m\frac{\partial\gamma}{\partial\varphi}\frac{\partial\gamma}{\partial\theta'}\Bigg]\mathbf{1}_\varphi\Bigg\}d\omega, \quad (6.40)$$

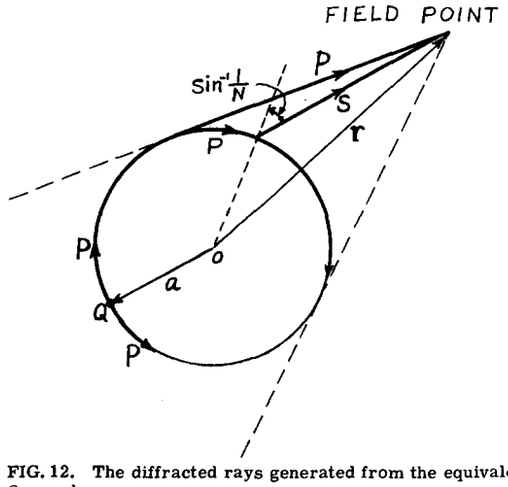


FIG. 12. The diffracted rays generated from the equivalent source at Q are shown.

where

$$J_p = \tfrac{1}{2}(N^2-1)^{1/2}(N^2-2)(k_1a/6)^{1/3}e^{-i\frac{\pi}{3}}[\mathrm{Ai}(t_\lambda)]^{-1}\times a^2/r^2\sin\gamma)^{1/2}(1-a^2/r^2)^{1/4}(1+e^{-i2\pi\nu_\lambda})^{-1}\times\exp[-ik_1(r^2-a^2)^{1/2}+i\nu_\lambda\cos^{-1}(a/r), \quad (6.41)$$

$$J_s = \tfrac{1}{2}(2k_1a/r)^{1/2}(N^2-2)(1-1/N^2)^{1/4}e^{-i(\pi/4)}\times(a^2/r^2\sin\gamma)^{1/2}(1-a^2/N^2r^2)^{1/4}(1+e^{-i2\pi\nu_\lambda})^{-1}\times\exp(-ik_2(r^2-a^2/N^2)^{1/2}+ik_2(a^2-a^2/N^2)^{1/2}+i\nu_\lambda\cos^{-1}(a/Nr)-i\nu_\lambda\cos^{-1}(1/N)], \quad (6.42)$$

$$k_{p\,m}^{p} = -i2\pi+i\pi e^{-i\pi\nu_\lambda}\sec(\pi\nu_\lambda)+i\cos^{-1}(a/r)\pm\gamma\begin{matrix}\cot\\ \tan\end{matrix}(\gamma\nu_\lambda+\pi/4)\quad\text{in }\mathcal{L}$$
$$= -i\pi+i\pi e^{-i\pi\nu_\lambda}\sec(\pi\nu_\lambda)+i\cos^{-1}(a/r)\mp(\pi-\gamma)\begin{matrix}\tan\\ \cot\end{matrix}(\pi\nu_\lambda-\gamma\nu_\lambda-\pi/4)\quad\text{in }\mathcal{L}', \quad (6.43)$$

$$k_{s\,m}^{p} = -i2\pi+i\pi e^{-i\pi\nu_\lambda}\sec(\pi\nu_\lambda)+i\cos^{-1}(a/Nr)-i\cos^{-1}\left(\frac{1}{N}\right)\pm\gamma\begin{matrix}\cot\\ \tan\end{matrix}(\gamma\nu_\lambda+\pi/4)\quad\text{in }\mathcal{L}$$
$$= -i\pi+i\pi e^{-i\pi\nu_\lambda}\sec(\pi\nu_\lambda)+i\cos^{-1}(a/Nr)-i\cos^{-1}(1/N)\mp(\pi-\gamma)\begin{matrix}\tan\\ \cot\end{matrix}(\pi\nu_\lambda-\gamma\nu_\lambda-\pi/4)\quad\text{in }\mathcal{L}', \quad (6.44)$$

$$E_m^p = \exp[-i\nu_\lambda(2\pi-\gamma)+i\pi/2]\pm\exp[-i\nu_\lambda(2\pi+\gamma)]\quad\text{in }\mathcal{L}$$
$$=\exp[-i\nu_\lambda(2\pi-\gamma)+i\pi/2]\mp\exp[-i\nu_\lambda\gamma]\quad\text{in }\mathcal{L}'. \quad (6.45)$$

$\mathbf{u}_{s1}^{D}$ represents mainly the total contribution from the waves which are radiated from the equivalent source $\mathbf{s}_1$, travel along the surface certain angular distances with phase velocity close to wk_1^{-1} while delay exponentially, and then either leaves the surface tangentially as a compression wave or leaves the surface by critical refraction angle as a shear wave toward the field point $\mathbf{r}$ (see Fig. 12). Again, its surface integral will be asymptotically evaluated and the corrections to reflection coefficients of the zeroth-order solution in far field region will be derived in a sequel to this paper.

The higher-order solutions can be systematically constructed in a similar manner. Their general expressions are the same as that of the first-order solution with $\mathbf{s}_1$ replaced by $\mathbf{s}_j (j = 2, 3, 4, \cdots)$.

7. PERTURBATION SOLUTION IN THE EXTERIOR OF A LARGE CONVEX ROUGH SCATTERER, MEAN WAVE FUNCTION AND MEAN INTENSITY

The zeroth-order solution is the solution for the special case of a large convex smooth and deterministic scatterer for which an asymptotic solution in the exterior of the scatterer can be constructed from the geometrical theory of diffraction[10,11,12] as mentioned in the Introduction. The higher-order solutions can be simply constructed by following the recipe for the scattering of scalar waves.[3]

The mean value of the wave function and intensity in the exterior of the scatterer are, respectively,

$$\langle\langle\mathbf{u}_i+\mathbf{u}_s\rangle\rangle = \mathbf{u}_i+\mathbf{u}_{s0}+\sum_{j=1}^{\infty}\epsilon^j\langle\langle\mathbf{u}_{sj}\rangle\rangle \quad (7.1)$$

and

$$\langle\langle|\mathbf{u}_s|^2\rangle\rangle = |\mathbf{u}_{s0}|^2+2\epsilon\mathrm{Re}(\mathbf{u}_{s0}^{*}\cdot\langle\langle\mathbf{u}_{s1}\rangle\rangle)+\epsilon^2[2\mathrm{Re}(\mathbf{u}_{s0}^{*}\cdot\langle\langle\mathbf{u}_{s2}\rangle\rangle)+\langle\langle|\mathbf{u}_{s1}|^2\rangle\rangle]+O(\epsilon^3). \quad (7.2)$$

Upon examining our problem, it is found that the statistics of this problem are contained only in the equivalent

sources $\mathbf{s}_j$ ($j = 1, 2, 3, \cdots$). Since $\mathbf{s}_j$ contain various combinations of f in sums and products, once all of the moments of these various combinations are known, then the mean values of the scattered wave function and the scattered intensity are determined. Furthermore, if the surface is statistically homogeneous, the mean value of the surface fluctuation is zero, i.e., $\langle\langle f\rangle\rangle = 0$, and then the mean value of the first-order perturbation solution is zero. In this case, the explicit computation of the second-order perturbation solution is needed. However, as long as the surface is not statistically homogeneous, the effect of the random rough surface on the scattering of elastic waves can be sufficiently explained by knowing the first-order perturbation solution explicitly.

APPENDIX

(1) Orthogonality relations between **l**, **m**, and **n**:

$$\|\mathbf{l}\|^2 = \int_0^{2\pi}\int_0^{\pi} \mathbf{l}_{\substack{e\\o}mn}(r_0)\cdot\mathbf{l}_{\substack{e\\o}mn}(r_0)\,\sin\theta\,d\theta\,d\varphi$$
$$= (1+\delta_m)\frac{2\pi}{2n+1}\frac{(n+m)!}{(n-m)!}\frac{1}{a^2}\{k_1^2a^2[h_n^{(2)\prime}(k_1a)]^2 + n(n+1)[h_n^{(2)}(k_1a)]^2\},$$

$$\|\mathbf{m}\|^2 = \int_0^{2\pi}\int_0^{\pi} \mathbf{m}_{\substack{e\\o}mn}(\mathbf{r}_0)\cdot\mathbf{m}_{\substack{e\\o}mn}(\mathbf{r}_0)\,\sin\theta\,d\theta\,d\varphi$$
$$= (1+\delta_m)\frac{2\pi}{2n+1}\frac{(n+m)!}{(n-m)!}n(n+1)[h_n^{(2)}(k_2a)]^{-2},$$

$$\|\mathbf{n}\|^2 = \int_0^{2\pi}\int_0^{\pi} \mathbf{n}_{\substack{e\\o}mn}(\mathbf{r}_0)\cdot\mathbf{n}_{\substack{e\\o}mn}(\mathbf{r}_0)\,\sin\theta\,d\theta d\varphi$$
$$= (1+\delta_m)\frac{2\pi}{2n+1}\frac{(n+m)!}{(n-m)!}n(n+1)\left\{\frac{n^2+n+1}{k_2^2a^2}[h_n^{(2)}(k_2a)]^2 + \frac{2}{k_2a}h_n^{(2)}(k_2a)h_n^{(2)\prime}(k_2a) + [h_n^{(2)\prime}(k_2a)]^2\right\},$$

$$\langle\mathbf{l}\cdot\mathbf{n}\rangle = \int_0^{2\pi}\int_0^{\pi} \mathbf{l}_{\substack{e\\o}mn}(\mathbf{r}_0)\cdot\mathbf{n}_{\substack{e\\o}mn}(\mathbf{r}_0)\,\sin\theta\,d\theta d\varphi$$
$$= (1+\delta_m)\frac{2\pi}{2n+1}\frac{(n+m)!}{(n-m)!}n(n+1)\frac{1}{a}\times\left[\frac{1}{k_2a}h_n^{(2)}(k_1a)h^{(2)}(k_2a) + h_n^{(2)}(k_1a)h_n^{(2)\prime}(k_2a) + \frac{k_1}{k_2}h_n^{(2)\prime}(k_1a)h_n^{(2)}(k_2a)\right],$$

where $\delta_m = 1$ if $m = 0$, $= 0$ if $m > 0$,

all other scalar products are zero.

(2) Poisson summation formula:

$$\sum_{n=0}^{\infty} g(n) = \sum_{l=-\infty}^{\infty}\int_0^{\infty} g(\nu)\,\exp(-i2\pi l\nu)d\nu.$$

(3) Relations between spherical Bessel functions and Bessel functions:

$$jn(\rho) = \sqrt{\pi/2\rho}\,J_{n+1/2}(\rho), \qquad h_n^{(2)}(\rho) = \sqrt{\pi/2\rho}H_{n+1/2}^{(2)}(\rho),$$
$$jn'(\rho) \simeq \sqrt{\pi/2\rho}\,J'_{n+1/2}(\rho), \qquad h_n^{(2)\prime}(\rho) \simeq \sqrt{\pi/2\rho}H_{n+1/2}^{(2)\prime}(\rho).$$

(4) Debye asymptotic forms for large argument and index, with $|\nu| < z$, are

$$H_\nu^{(1)}(z) \approx \sqrt{2/\pi}(z^2-\nu^2)^{1/4}\exp[i(z^2-\nu^2)^{1/2} - i\nu\cos^{-1}(\nu/z) - i\pi/4],$$
$$H_\nu^{(2)}(z) \approx \sqrt{z/\pi}(z^2-\nu^2)^{-1/4}\exp[-i(z^2-\nu^2)^{1/2} + i\nu\cos^{-1}(\nu/z) + i\pi/4],$$
$$H_\nu^{(1)\prime}(z) \approx i(1-\nu^2/z^2)^{1/2}H_\nu^{(1)}(z),$$
$$H_\nu^{(2)\prime}(z) \approx -i(1-\nu^2/z^2)^{1/2}H_\nu^{(2)}(z).$$

(5) Asymptotic forms of $Q_{\nu-1/2}^{(1)}$ and $Q_{\nu-1/2}^{(2)}$ for $0 < \theta < \pi$:

$$Q_{\nu-1/2}^{(1)}(\cos\theta) \simeq (2\pi\nu\,\sin\theta)^{-1/2}\exp(-i\nu\theta + i\pi/4),$$
$$Q_{\nu-1/2}^{(2)}(\cos\theta) \simeq (2\pi\nu\,\sin\theta)^{-1/2}\exp(i\nu\theta - i\pi/4),$$
$$Q_{\nu-1/2}^{(1)}(-\cos\theta) \simeq (2\pi\nu\,\sin\theta)^{-1/2}\exp(i\nu\pi - i\nu\theta - i\pi/4),$$
$$Q_{\nu-1/2}^{(2)}(-\cos\theta) \simeq (2\pi\nu\,\sin\theta)^{-1/2}\exp(-i\nu\pi + i\nu\theta + i\pi/4).$$

(6) Saddle point method for line integral:

$$\int_\alpha^\beta f(t)\,\exp[ixh(t)]dt \simeq [2\pi/xh''(t_0)]^{1/2}f(t_0)\exp[ixh(t_0) + i\pi/4],\quad x\to\infty,$$

where t_0 is the saddle point and satisfies

$$h'(t_0) = 0, \qquad h''(t_0) \neq 0.$$

(7) Relations between Legendre functions and Hankel functions:

$$P_{\nu-1/2}(\cos\theta) = -ie^{i\pi\nu}P_{\nu-1/2}(-\cos\theta) + (1+e^{i2\pi\nu})Q_{\nu-1/2}^{(1)}(\cos\theta),$$
$$P_{-\nu-1/2}(-\cos\theta) = P_{\nu-1/2}(-\cos\theta),$$
$$H_{-\nu}^{(1)}e^{-i(\pi/2)\nu} = H_\nu^{(1)}e^{i(\pi/2)\nu}.$$

(8) Airy function representations for the Hankel functions for argument and index both large, and $\nu \simeq z$:

$$H_\nu^{(1)}(k_1a) \simeq (2/\pi)ze^{-i(2\pi/3)}\mathrm{Ai}(te^{-i(2\pi/3)}),$$
$$H_\nu^{(1)\prime}(k_1a) \simeq -(2/\pi)z^2e^{-i(4\pi/3)}\mathrm{Ai}'(te^{-i(2\pi/3)}),$$
$$H_\nu^{(2)}(k_1a) \simeq (2/\pi)z\,\mathrm{Ai}(t),$$
$$H_\nu^{(2)\prime}(k_1a) \simeq (2/\pi)z^2\mathrm{Ai}'(t),$$
$$\frac{\partial}{\partial\nu}H_\nu^{(2)}(k_1a) \simeq (2/\pi)z^2\mathrm{Ai}'(t),$$
$$\frac{\partial}{\partial\nu}H_\nu^{(2)\prime}(k_1a) \simeq (2/\pi)(z^3/3)t\,\mathrm{Ai}(t),$$

where

$$z = (6/k_1a)^{1/3}e^{i\frac{\pi}{3}}, \qquad t = z(\nu - k_1a).$$

(9) The addition formula for the Legendre polynomials:

$$P_n(\cos\gamma) = P_n(\cos\theta)P_n(\cos\theta') + 2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!}\times P_n^m(\cos\theta)P_n^m(\cos\theta')\cos m(\varphi-\varphi'),$$

where

$$\cos\gamma \equiv \sin\theta\,\sin\theta'\cos(\varphi-\varphi') + \cos\theta\,\cos\theta'.$$

[1] P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces* (Macmillan, New York, 1963).
[2] Y. M. Chen and S. J. Kim, J. Acoust. Soc. Am. **42**, 1 (1967).
[3] Y. M. Chen, J. Math. Phys. **9**, 439 (1968).
[4] Y. M. Chen and W. Fan, J. Acoust. Soc. Am. (to appear).
[5] J. W. Dunkin and A. C. Eringen, Proc. Fourth U. S. Nat. Congr. Appl. Mech. **1**, 143 (1962).
[6] G. H. Handelman, "Surface Waves over a Slightly Curved Elastic Half-space," in *Progress in Applied Mechanics* (Macmillan, New York, 1963), pp. 43–54.
[7] I. Abubakar, Proc. Camb. Phil. Soc. **58**, 136 (1962).
[8] I. Abubakar, Proc. Camb. Phil. Soc. **59**, 231 (1963).
[9] P. M. Morse and H. Feshback, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. II.
[10] J. B. Keller, "A Geometrical Theory of Diffraction", Symp. Microwave Optics, Eaton Electronics Res. Lab., McGill Univ. Montreal, Canada, 1953.
[11] J. B. Keller, J. Opt. Soc. Am. **52**, 116 (1962).
[12] F. Gilbert, J. Acoust. Soc. Am. **32**, 841 (1960).
[13] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).
[14] M. Nagase, J. Phys. Soc., Japan, **11**, 279 (1956).
[15] F. Gilbert and L. Knopoff, J. Acoust. Soc. Am. **31**, 1169 (1959).
[16] P. L. Christiansen, "Scattering of Short Electromagnetic and Elastic Waves by a Circular Cylinder in which Both Longitudinal and Transverse Modes are Excited," Tech. Rep., Lab. Math. Phys., The Technical University of Denmark (1965).

The mixture of quantum states

William M. Cornette

Weapons Development Department, U.S. Naval Weapons Center, China Lake, California 93555

Stanley P. Gudder

Department of Mathematics, University of Denver, Denver, Colorado 80210
(Received 2 January 1974)

Convex prestructures, a mathematical framework that extends the usual concept of convex subset of a real linear space, are employed to generalize methods used in the study of axiomatic quantum mechanics. A brief summary of the mathematical framework of convex prestructures is given. Convex prestructures are classified, and those which are isomorphic to a convex subset of a real linear space are characterized. The operational quantum mechanics of Davies and Lewis is generalized within the framework of convex prestructures and the existence of a physically motivated orthomodular poset is given. Mielnik's beams and filters are also discussed within the framework of convex prestructures. An error in Mielnik's formulation of beam mixtures is pointed out and it is shown that his beam mixtures and one classification of convex prestructures are equivalent. Also his concept of a filter is generalized in the framework of convex prestructures and geometric requirements needed on the set of normalized states so that they may correspond to a physical system are investigated. Finally, Mackey's axioms are discussed and reformulated in the language of P-convex structures.

1. INTRODUCTION

Axiomatic quantum mechanics has been developed from several viewpoints. Two of the most widely accepted approaches have been the C^*-algebra approach[1,2] and the "quantum logic" approach.[3-6] However, recently several researchers have approached the development of axiomatic quantum mechanics using convex set methods to study the geometric properties of the set of normalized state S_1. This "convex set" or "operational" approach has been studied by Ludwig *et al.*[7-11], Gunson,[12] Mielnik,[13,14] Davies and Lewis,[15,16] and others.[17-21]

The authors have recently generalized the usual concept of a convex subset of a real linear space from a viewpoint that seems to be reasonably physically motivated.[22,23] In this paper, we shall use this framework to generalize and unify the formalism of operational and convex set methods used by Davies and Lewis,[15,16] and Mielnik.[13,14] This generalization of convexity involves only the set of normalized states of a physical system with an operation of the formation of "mixtures" of quantum states.

2. CONVEX PRESTRUCTURES

Let $S_1=\{p, q, r, \cdots\}$ be the set of normalized states for some physical system. For generality, we shall not specify any particular form for these states but take them to be undefined, primitive elements. In different axiomatic models for quantum mechanics the states take various forms. In the conventional model, the states are positive trace-class operators with trace one[4-6]; in the quantum logic model, the states are probability measures on an orthomodular lattice[3-6]; in the algebraic model, the states are positive, normal, linear functionals on a C^*-algebra[1,2]; in the operational model, the states are positive elements of an ordered Banach space.[14,16,18] Although in these models the normalized states form a convex subset of a vector space, we can formulate an axiomatic framework without any linear structure whatsoever. Thus our theory not only generalizes the usual models but leaves open the possibility of nonlinear structures for quantum mechanics.[22,23]

We shall assume a concept of a mixture of quantum states by postulating a map

$$T_1:[0,1]\times S_1\times S_1\to S_1$$
$$(\lambda,p,q)\to\langle\lambda,p,q\rangle$$

where $\langle\lambda,p,q\rangle$ is a *mixture* of the normalized states p and q. Physically, $\langle\lambda,p,q\rangle$ can be interpreted as a mixture of $(1-\lambda)$ parts p and λ parts q. If S_1 is a convex subset of a linear space, it shall be understood that $\langle\lambda,p,q\rangle=(1-\lambda)p+\lambda q$. There exist many instances in which S_1 cannot be considered as a convex subset of a real linear space; for example, S_1 can be finite or countable. The structure (S_1,T_1) will be called *a convex prestructure*. No real restriction is placed on a set S_1 by postulating such a map T_1 unless some axioms are placed on T_1. Some of the more useful axioms are:

(i) $\langle\lambda,p,q\rangle=\langle1-\lambda,q,p\rangle$ (c-commutativity)

(ii) $\langle0,p,q\rangle=p$ (0-endpoint condition)

(iii) $\langle\lambda,p,p\rangle=p$ (point convexity)

(iv) $\langle\lambda,p,\langle\mu,p,q\rangle\rangle=\langle\lambda\mu,p,q\rangle$ (weak associativity)

(v) $\langle\lambda,p,\langle\mu,q,r\rangle\rangle$
$=\langle\lambda\mu,\langle\lambda(1-\mu)(1-\lambda\mu)^{-1},p,q\rangle,r\rangle$ $(\lambda\mu\neq1)$ (c-associativity).

Two *measurement axioms* are also useful:

(M1) $\langle\lambda,p,q\rangle=p$ implies $\lambda=0$ or $p=q$

(M2) $\langle\lambda,p,q\rangle=\langle\lambda,p,r\rangle$ implies $\lambda=0$ or $q=r$.

Convex prestructures with some of the above axioms have been studied in detail by the authors and are covered elsewhere.[22] All that shall be attempted in this section is a brief summary of the main results that will be needed later in this paper.

Three sets are useful in the study of convex prestructures:

$[p,q]=\{r\in S_1:\langle\lambda,p,q\rangle=r$ for some $\lambda\in[0,1]\}$ (line),

$\langle p,q\rangle=\{r\in S_1:\langle\lambda,p,q\rangle=r$ for some $\lambda\in(0,1)\}$,

$(p,q)=[p,q]-\{p,q\}$ (interior).

An *affine map* Φ from (S_1,T_1) to (S_1',T_1') satisfies

$\Phi(\langle\lambda,p,q\rangle)=\langle\lambda,\Phi(p),\Phi(q)\rangle'$. The set of all affine maps from S_1 to S_1' is denoted by $Af(S_1,S_1')$. If $S_1=S_1'$, $Af(S_1,S_1)\equiv Af(S_1)$ and if $S_1'=R$, $Af(S_1,R)\equiv S_1^*$. S_1^* is *total* if $p\neq q$ implies there exists $\Phi\in S_1^*$ such that $\Phi(p)\neq\Phi(q)$. If we require that the map $\lambda\to\langle\lambda,p,q\rangle$ be affine, then T_1 must be c-commutative, weak associative, and have the 0-endpoint condition. Such a structure is referred to as a convex *line-based structure*. If we require T_1 to be c-associative, rather than weak associative, in addition to being c-commutative and having the 0-end point condition so that we can deal meaningfully with mixtures of more than two states, we obtain a *convex structure*; and if we also require point convexity, we obtain a *P-convex structure*. A P-convex structure that satisfies (M1) is *dense* and if it satisfies (M2), it is *full*.

Theorem 2.1: If (S_1,T_1) is a convex line-based structure, then all lines $[p,q]$ belong to one of the following classes:

Class I: $(p,q)=\phi$ and $\langle p,q\rangle=p$ or q;

Class II: $(p,q)=\langle p,q\rangle=r$ for some $r\in S_1$;

Class III: $[p,q]$ is isomorphic to $[0,1]$ where isomorphic means there exists a bijective affine map from $[p,q]$ onto $[0,1]$.

Using point convexity and c-associativity, it is easily seen that a P-convex structure is a convex line-based structure. Examples can be given[22] which show that a convex line-based structure need not be a convex structure and a convex structure need not be a P-convex structure.

We shall now define a "distance" between normalized states. Let

$$\sigma(q,p)=\inf\{0\leq\lambda<1:\langle\lambda,p,p_1\rangle=\langle\lambda,q,q_1\rangle,\ p_1,q_1\in S_1\}$$

and

$$\rho(p,q)=\sigma(p,q)[1-\sigma(p,q)]^{-1}.$$

If (S_1,T_1) is a convex structure, then ρ is a semimetric (called the *intrinsic semimetric*) and $0\leq\rho(p,q)\leq 1$. In fact, if S_1 is a bounded convex subset of R^n then ρ is topologically equivalent to the Euclidean metric.

Let S_0 be a convex set in a real vector space V with intrinsic semimetric ρ. S_0 is *absorbing* if for any $x\in V$ there is a $\delta(x)>0$ such that $\lambda x\in S_0$ for all λ with $|\lambda|\leq\delta(x)$. S_0 is *balanced* if $\lambda x\in S_0$, for all $x\in S_0$, $|\lambda|\leq 1$. Let $D=\{cp-dq:0\leq c,d\leq 1;\ p,q\in S_0\}$; then D is a convex, balanced absorbing subset of X, the subspace of V generated by S_0, and $0\in D$. For $x\in X$, let $|x|=\inf\{\lambda>0:x\in\lambda D\}$. Then $|\cdot|$ is the Minkowski functional[21] for D in X. $|\cdot|$ is a seminorm induced by the intrinsic semimetric ρ.

Theorem 2.2: (a) For a convex prestructure, the following are equivalent:

(i) (S_1,T_1) is a P-convex structure that satisfies the (M2) measurement axiom;

(ii) S_1^*, the set of all affine functionals into the reals, is total;

(iii) S_1 is isomorphic to a convex subset S_0 of a real linear space.

(b) For a P-convex structure (S_1,T_1), ρ is a metric if and only if S_1 is isomorphic to a bounded convex subset S_0 of a real linear space.

For physical applications, it is important to consider the set $S\equiv S_1^+=\{(\alpha,p):\alpha\geq 0$ and $p\in S_1\}$. We shall define $(\alpha,p)=(\beta,q)$ if $\alpha=\beta\neq 0$ and $p=q$ and $(0,p)=(0,q)\equiv 0$ for all $p,q\in S_1$. For convenience we shall write αp instead of (α,p). If S_1 corresponds to the set of normalized states of a physical system, then S corresponds to the states. For $\alpha p\in S$ and $\beta\geq 0$, we define $\beta(\alpha p)=(\beta\alpha)p$. A convex prestructure (S,T) can then be generated on S from the convex prestructure (S_1,T_1) on S_1 in the following manner:

$$\langle\lambda,\alpha p,\beta q\rangle=0 \text{ if } (1-\lambda)\alpha+\lambda\beta=0$$

and

$$\langle\lambda,\alpha p,\beta q\rangle=[(1-\lambda)\alpha+\lambda\beta]\langle\lambda\beta[(1-\lambda)\alpha+\lambda\beta]^{-1},\ p,q\rangle.$$

Theorem 2.3: The generated convex prestructure (S,T) satisfies the axioms (i) through (v) and (M1) and (M2) if and only if the convex prestructure, or generator, (S_1,T_1) does.

We define a binary operation, $+$ on (S,T) by $p+q=2\langle\frac{1}{2},p,q\rangle$.

Theorem 2.4: Let (S_1,T_1) be a P-convex structure. Then the operation $+$ on the generated P-convex structure (S,T) satisfies the following:

(i) $X+0=X$,

(ii) $X+Y=X$ implies $Y=0$,

(iii) $X+Y=0$ implies $X=Y=0$,

(iv) $X+Y=Y+X$,

(v) $X+(Y+Z)=(X+Y)+Z$.

(Note: For proofs of the above results and additional results concerning convex prestructures, the reader is referred to other works by the authors.[22,23])

3. OPERATIONAL QUANTUM MECHANICS

In this section we briefly describe a generalization of Davies' and Lewis' operational quantum mechanics[16] in our convex structure framework. The set S_1 of normalized states is assumed to be a convex structure and the set of *states* S is defined as $S=\{\alpha p:\alpha\geq 0,\ p\in S_1\}$. As we have seen S can be made into a convex structure (Theorem 2.3) and S_1 can be thought of as a convex substructure of S. Now $\tau\in S^*$ is the *distinguished affine functional* defined as $\tau(\alpha p)=\alpha$ for all $\alpha p\in S$. We endow S^* with the weak*-topology; that is, if $f_0\in S^*$ a neighborhood of f_0 is a set of the form $N(f_0;p_1,\ldots,p_n;\epsilon)=\{f\in S^*:|f(p_i)-f_0(p_i)|<\epsilon,\ i=1,\ldots,n\}$ where $p_1,\ldots,p_n\in S$, $\epsilon>0$. We also endow $Af(S)$ with the topology given by neighborhoods of the form

$$N(A_0;p_1,\ldots,p_n;\epsilon)=\{A\in Af(S):\rho(Ap_i,A_0p_i)<\epsilon,\ i=1,\ldots,n\}$$

where $p_1,\ldots,p_n\in S$, $\epsilon>0$ and ρ is the intrinsic semimetric on S.

An *observable* is a triple $(X,\mathcal{A},a)$, where X is a set, $\mathcal{A}$ is a σ-algebra of subsets of X and $a:\mathcal{A}\to S^*$ satisfies:

(i) $0\leq a(E)\leq a(X)$ for all $E\in\mathcal{A}$;

(ii) $a(X)=\tau$;

(iii) if E_i are mutually disjoint sets in $\mathcal{A}$ then $a(\sum_1^\infty E_i) = \sum_1^\infty a(E_i)$ where the convergence is in the topology of S^*.

An *instrument* is a triple $(X, \mathcal{A}, \mathcal{E})$ consisting of a Borel space $(X, \mathcal{A})$ and a map $\mathcal{E}: \mathcal{A} \to Af(S)$ satisfying:

(i) $\tau[\mathcal{E}(X)p] = \tau(p)$ for every $p \in S$:

(ii) if E_i are mutually disjoint sets in $\mathcal{A}$, then $\mathcal{E}(\cup_1^\infty E_i) = \sum_1^\infty \mathcal{E}(E_i)$ where the convergence is in the topology of $Af(S)$.

Now to every instrument $\mathcal{E}$ on $(X, \mathcal{A})$ there is a unique observable a on $(X, \mathcal{A})$ such that $\tau[\mathcal{E}(E)p] = a(E)p$ for every $p \in S$, $E \in \mathcal{A}$. Indeed, if we define $a(E)p = \tau[\mathcal{E}(E)p]$ then it is easily checked that a is an observable. Conversely, any observable a on $(X, \mathcal{A})$ is determined in the above manner by at least one instrument. Indeed, let $p \in S_1$ and define $\mathcal{E}: \mathcal{A} \to Af(S)$ by $\mathcal{E}(E)q = [a(E)q]p$. Then $\mathcal{E}$ is an instrument and $\tau[\mathcal{E}(E)q] = a(E)p$. More generally, if $p_i \in S_1$ and E_i are mutually disjoint, $i = 1, \ldots, n$, then define $\mathcal{E}(E)q = \sum [a(E \cap E_i)q]p_i$.

If $\mathcal{E}$ and $\mathcal{J}$ are instruments on Borel spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, respectively, and there exists an instrument G on $(X \times Y, \mathcal{A} \times \mathcal{B})$ such that $\mathcal{G}(E \times F) = \mathcal{E}(E)\,\mathcal{J}(F)$ for all Borel sets $E \in \mathcal{A}$, $F \in \mathcal{B}$, we call $\mathcal{G}$ the *composition* of $\mathcal{E}$ following $\mathcal{J}$ and use the notation $\mathcal{G} = \mathcal{E} \circ \mathcal{J}$. It is clear that $\mathcal{G}$ is unique if it exists. Davies and Lewis[16] give an existence proof in their framework; we shall not discuss existence here.

Davies and Lewis give the following interpretation of their model. A state p corresponds to a "beam" of non-interacting copies of a system and $\tau(p)$ is a measure of the strength of the beam. An instrument $\mathcal{E}: (X, \mathcal{A}) \to Af(S)$ measuring a physical quantity taking values in X corresponds to a family of filters indexed by the subsets $\mathcal{A}$; the filter corresponding to $\mathcal{E}(E)$ allows a copy of the system to pass if its observed value lies in E and in so doing transforms the input state p into the output state $\mathcal{E}(E)p$. Now $E \to \tau[\mathcal{E}(E)p]/\tau(p)$ is a probability measure on $(X, \mathcal{A})$; the right-hand side is the ratio of the strength of the output beam to the strength of the input and hence is the probability that in the state p the physical quantity observed takes a value in $E \in \mathcal{A}$. Successive applications of two families of filters correspond to the composition of instruments.

For $A \in Af(S)$ we define the linear operator $A^*: S^* \to S^*$ by $(A^*f)(p) = f(Ap)$ for all $f \in S^*$, $p \in S$.[23] Let $(X, \mathcal{E})$ and $(Y, \mathcal{J})$ be instruments which determine observables (X, a) and (Y, b), respectively. The S^*-valued measure $F \to \mathcal{E}(X)^* b(F)$ is an observable called the observable b *conditioned by the measure* a *with instrument* $\mathcal{E}$. If $\mathcal{J} \circ \mathcal{E}$ exists the observable $c(M) = (\mathcal{J} \circ \mathcal{E})(M)^* \tau$ is called the *joint distribution of* $\mathcal{J}$ *following* $\mathcal{E}$. Let us check that the marginal distributions are $c(Y \times E) = a(E)$, $c(F \times X) = \mathcal{E}(X)^* b(F)$. Indeed,

$$\begin{aligned} c(Y \times E)p &= [(\mathcal{J} \circ \mathcal{E})(Y \times E)^*]p \\ &= \tau[(\mathcal{J} \circ \mathcal{E})(Y \times E)p] = \tau[\mathcal{J}(Y)\mathcal{E}(E)p] \\ &= (\tau(E)p) = a(E)p \end{aligned}$$

and

$$\begin{aligned} c(F \times X)p &= \tau[(\mathcal{J} \circ \mathcal{E})(F \times X)p] = \tau[\mathcal{J}(F)\mathcal{E}(X)p] \\ &= b(F)[\mathcal{E}(X)p] = [\mathcal{E}(X)^* b(F)]p. \end{aligned}$$

Thus a joint distribution gives the desired marginal distributions.

Let (S, T) be a convex structure satisfying the (M2) measurement axiom. We define a *proposition* as an element $A \in Af(S)$ satisfying:

(1) $A^2 = A$;

(2) there exists $A' \in Af(S)$ such that $A'^2 = A'$, $A'A = AA' = 0$, and $A + A' = I$.

Notice that A' is unique, since if $A + B = I$ then

$$2\langle 2^{-1}, Ap, Bp\rangle = Ap + Bp = (A + B)p = (A + A')p = Ap + A'p = 2\langle 2^{-1}, Ap, A'p\rangle$$

so $Bp = A'p$ for all $p \in S$. Let P be the set of propositions. For $A, B \in P$ we define $A \leq B$ if $AB = BA = A$.

Theorem 3.1: P is an orthomodular poset.

Proof: Notice $0, I \in P$ where $0(p) = 0$, $I(p) = p$ for all $p \in S$. Clearly $0 \leq A \leq I$ and $A \leq A$ for all $A \in P$. If $A \leq B$ and $B \leq A$ then $A = AB = B$ and if $A \leq B$, $B \leq C$ then $AC = (AB)C = A(BC) = AB = A$ so $A \leq C$. Thus P is a poset with universal bounds $0, I$. We now show $A \to A'$ is an orthocomplementation. Clearly $A'' = A$. If $A \leq B$, then since $I = A + A'$ we have $B' = B'A + B'A'$. But $B'A = B'BA = BA = 0$ so $B' = B'A'$. Similarly, $B' = A'B'$ so $B' \leq A'$. Now if $A \perp B$ then $A + B \in P$. Indeed since $B = AB + A'B = AB + B$ we have $AB = 0$ and similarly $BA = 0$ so $(A + B)^2 = A + B$ and (1) is satisfied. We now show $(A + B)' = B'A'$ which would prove (2) for $A + B$. In fact, since $B + A'B' = A'B + A'B' = A'$ we have $B'A'B' = B'A'$ and since $B + B'A' = BA' + B'A' = A'$ we have $B'A'B' = A'B'$. Thus $(A'B')^2 = A'B'$ and $(A + B)A'B' = 0 = A'B'(A + B)$. Finally, since $B + B'A' = A'$ we have $A + B + B'A' = A + A' = I$. Clearly $A, B \leq A + B$ and if $A, B \leq C$ then $C(A + B) = CA + CB = A + B = (A + B)C$ so $A + B \leq C$ and hence $A \wedge B = A + B$. Clearly $A \wedge A' = I$ for all $A \in P$. Finally, if $A \leq B$ then $B = BA + BA' = A + BA'$. Since $A \perp B'$ from before, we have $BA' = (A + B')' \in P$. We thus see that $B = A \vee (BA')$ which completes the proof.

The set of propositions from the basis for the quantum logic approach of Jauch–Piron[4,24] and Mackey.[5] It should be mentioned that Davies and Lewis[16] argue that the order we have used in Theorem 3.1 is not the correct order for physical implication while Mielnik[14] points out that condition (2) for propositions may not be a realistic physical assumption. Thus there might be propositions which do not satisfy (2) so the set of propositions may not have as rich a structure as an orthomodular poset. We shall not argue these points here but merely set down the facts so the reader may judge for himself.

Returning to our general convex structure of states S, suppose there are enough observables to distinguish between states or equivalently that S^* is total. It then follows from Theorem 2.2 that S_1 is isomorphic to a convex set in a real linear space V. Then S is a cone with base S_1 and $X = S - S$ is a subspace of V generated by S. The intrinsic semimetric ρ is a metric and so we

can form the completion $\tilde{S}_1$ of S_1. Then $\tilde{S}_1$ is a base for a cone $\tilde{S}$ which generates the linear space $\tilde{X}=\tilde{S}-\tilde{S}$. If $|\cdot|$ is the intrinsic norm on $\tilde{X}$ then $\tilde{S}_1$ is normalized and hence $(\tilde{X}, |\cdot|)$ is a Banach space with closed generating cone $\tilde{S}$ (Ref. 23, Theorem 3.7). Defining τ as above we see that $\tau(x)=|x|$ for every $x\in\tilde{S}$ and the triple (X,S,τ) becomes a complete base normed space (or state space), the basic framework for the operational quantum mechanics of Davies and Lewis. The norm $|\cdot|$ is equivalent to the natural norm $|\cdot|_1$ used by Davies and Lewis and our definitions of observables, instruments, joint distributions, etc., reduce to theirs (Ref. 23, Theorem 35). We thus see that in the case of a separating set of observables our framework reduces to that of Davies and Lewis and hence gives a simple, axiomatic motivation for their theory.

4. QUANTUM INFORMATION SYSTEMS: MIXTURE OF BEAMS

In the next three sections we compare our theory to that of Mielnik, the latter being similar to and representative of the framework of Ludwig, Gunson, and others. As has been pointed out previously, it is possible that (pure) quantum states do not form a linear space. For example, if the superposition principle does not necessarily hold then the states may form a nonlinear structure. As a step in this direction Mielnik[13,14] has considered simple examples of physical systems in which the pure quantum states cannot be represented by vectors in a Hilbert space.

In Mielnik's development of axiomatic quantum mechanics,[14] a *general information system* (D, T, B) is defined as three sets: D, a set of *detectors*; B, a set of *beams*; and T, a set of *transmitters* such that

1. There exists a map $D\times B\to R$, $(d,x)\to dx$ such that
   (a) $\exists!\ 0\in B$ such that $d0=0$, $\forall\, d\in D$,
   (b) $\exists!\ 0\in D$ such that $0x=0$, $\forall\, x\in B$.
2. There exists a map $B\times B\to B$, $(x,y)\to x+y$ such that
   (a) $x+y=y+x$,
   (b) $x+(y+z)=(x+y)+z$,
   (c) $x+0=x$,
   (d) $x+y+z=x$ implies $y=z=0$.
3. T is a non-abelian semigroup of operators in B. The image dx defined in condition 1 is the *average intensity of the beam x measured by the detector d*; the image $x+y$ defined in condition 2 the *combination of x and y*. From condition 1 above, each $d\in D$ determines a *detection functional* on the set of beams and each $x\in B$ determines a *beam functional* on the set of detectors.

Mielnik extends the definition of combinations of beams by defining $nx=x+x+\cdots+x$ (n times) for $n\in N$ and then assumes that there exists a map $R^+\times B\to B$, where $R^+=\{\lambda\in R:\lambda\geq 0\}$, such that $\forall\,\lambda,\mu\in R^+$ and $\forall\, x,y\in B$

1. $1x=x$, $0x=0$,
2. $\lambda(\mu x)=(\lambda\mu)x$,
3. $(\lambda+\mu)x=\lambda x+\mu x$,
4. $\lambda(x+y)=\lambda x+\lambda y$,
5. $nx=x+x+\cdots+x$ (n times), $n\in N$.

This will be called *multiplication by nonnegative scalars* and it is closely related to average beam intensity. Mielnik notes that B, with combinations and multiplication by nonnegative scalars, does not give one a linear space; however, he does state

> ... B admits a convenient representation as a convex cone in a real linear space X so that the linear combination with positive coefficients defined in B becomes a special case of the general linear combination in X. The space X can be constructed as the set of formal differences $x-y$ of beams $x,y\in B$ with the assumed identity: $x-x=0$ .... The embedding of B in X is done by assigning to each $x\in B$ an element $x-0\in X$. The set B then becomes a *positive cone* in X (Ref. 14, p. 7).

In the above, Mielnik implicitly assumes that it is possible to define an inverse; he does this through the irreversibility axiom $2d$.[25] The above axioms do not imply that B can be embedded in a linear space, as the following example will show.

Let $S_1=\{p,q\}$, $p\neq q$, and consider the cartesian product $R^+\times S$; we shall write λp and μq for (λ,p) and (μ,q), respectively. If $\lambda\neq 0\neq\mu$, we define $\lambda p=\mu p$ iff $\lambda=\mu$ and $p=q$; $0p=0q$. This is an equivalence relation on (R^+,S). Also assume that $1x=x$ and define $\lambda(\mu x)x=(\lambda\mu)x$ for $\lambda\geq 0$ and $x\in\{p,q\}$. Define $\lambda p+\mu q$ to be $(\lambda+\mu)p$ if $\lambda\neq 0$ or μq if $\lambda=0$; further, define $(\lambda+\mu)x=\lambda x+\mu x$ and $\lambda(x+y)=\lambda x+\lambda y$ for $x,y\in\{p,q\}$. One can easily see that $x+y$ and λx so defined do satisfy all of Mielnik's axioms on combinations and multiplication by nonnegative numbers. Such a structure does not, however, generate a linear space; indeed, if $\exists\,(-p)$ such that $p+(-p)=0$, then $(p+q)-p=(p+p)-p$ which implies $p=q$, a contradiction. Therefore, Mielnik's statement that B can be represented as a positive cone in a linear space is incorrect. We shall now show that the above structure is a P-convex structure.

Let (S_1,T_1) be a convex prestructure and let (S,T) be the generated convex prestructure as defined in Sec. 2.

Theorem 4.1: If (S_1,T_1) is a convex line-based (convex, P-convex, dense P-convex, full P-convex) structure, then (S,T) is a convex line-based (convex, P-convex, dense P-convex, full P-convex) structure.

Proof: It has been proved that if (S_1,T_1) either is C-commutative, or is C-associative, or satisfies the 0-endpoint condition, then (S,T) either is C-commutative, or is C-associative, or satisfies the 0-endpoint condition (Ref. 23, Lemma 2.7). Therefore, we need to prove similar results for weak associativity, the point convexity condition, the M1 measurement axiom, and the M2 measurement axiom. First, we show weak associativity.

If $(1-\mu)\sigma+\mu\tau\neq 0$ and $\lambda\sigma\neq 0$, then

$$\langle\lambda,\sigma p,\langle\mu,\sigma p,\tau q\rangle\rangle$$
$$=\left\langle\lambda,\sigma p,[(1-\mu)\sigma+\mu\tau]\left\langle\frac{\mu\tau}{(1-\mu)\sigma+\mu\tau},\,p,q\right\rangle\right\rangle$$

$$= [(1-\lambda)\sigma + \lambda[(1-\mu)\sigma + \mu\tau]]\left\langle \frac{\lambda[(1-\mu)\sigma+\mu\tau]}{(1-\lambda\mu)\sigma+\lambda\mu\tau}, p, \left\langle \frac{\mu\tau}{(1-\mu)\sigma+\mu\tau}, p, q \right\rangle \right\rangle$$

$$= [(1-\lambda\mu)\sigma + \lambda\mu\tau]\left\langle \frac{\lambda\mu\tau}{(1-\lambda\mu)\sigma+\lambda\mu\tau}, p, q \right\rangle$$

$$= \langle \lambda\mu, \sigma p, \tau q \rangle.$$

If $(1-\mu)\sigma + \mu\tau = 0$, then

$$\langle \lambda, \sigma p, \langle \mu, \sigma p, \tau q \rangle \rangle = \langle \lambda, \sigma p, 0 \rangle$$

so that (S, T) is weakly associative. Also,

$$\langle \lambda, \mu p, \mu p \rangle = \mu \langle \lambda, p, p \rangle = \mu p,$$

so (S, T) satisfies the point convexity condition. Next, we prove the M1 measurement axiom. Assume that

$$\langle \lambda, \mu p, \nu q \rangle = \nu q.$$

If $(1-\lambda)\mu + \lambda\nu = 0$, then either $\lambda = 1$ or $\mu = 0$, so $\mu p = 0$.

If $(1-\lambda)\mu + \lambda\nu \neq 0$, then

$$\langle \lambda, \mu p, \nu q \rangle = [(1-\lambda)\mu + \lambda\mu]\left\langle \frac{\lambda\nu}{(1-\lambda)\mu+\lambda\mu}, p, q \right\rangle = \nu q.$$

The latter occurs only if $(1-\lambda)\mu + \lambda\nu = \nu$, so either $\mu = \nu$ or $\lambda = 1$. If $\mu = \nu$, then $\mu\langle \lambda, p, q \rangle = \mu q = \langle \lambda, p, q \rangle$ so that $p = q$. To prove M2 measurement axiom, assume

$$\langle \lambda, \mu p, \nu q \rangle = \langle \lambda, \mu p, \sigma r \rangle.$$

If $(1-\lambda)\mu + \lambda\nu = 0$ or $(1-\lambda)\mu + \lambda\sigma = 0$, then follow steps similar to those in the proof of the M1 axiom above. Otherwise,

$$[(1-\lambda)\mu + \lambda\nu]\left\langle \frac{\lambda\nu}{(1-\lambda)\mu+\lambda\nu}, p, q \right\rangle = [(1-\lambda)\mu + \lambda\sigma]\left\langle \frac{\lambda\sigma}{(1-\lambda)\mu+\lambda\sigma}, p, r \right\rangle,$$

so that

$$(1-\lambda)\mu + \lambda\nu = (1-\lambda)\mu + \lambda\sigma.$$

If $\lambda \neq 0$, then $\sigma = \nu$ and from that we easily obtain $q = r$.

From this we shall show that Mielnik's set of beams with combinations and multiplication is equivalent to (S, T) for some P-convex structure (S_1, T_1). A detector or transmitter is *c-linear* if $\phi(x+y) = \phi(x) + \phi(y)$ and $\phi(\lambda x) = \lambda\phi(x) \forall \lambda \geq 0$, $\forall x, y \in S$. A *standard quantum detector* e is a c-linear detector such that (1) $ex \geq 0$ $\forall x \in B$ and (2) $ex = 0$ implies $x = 0$. For a given standard quantum detector e, we can define the *statistical figure* $S_e = \{x \in B : ex = 1\}$. Since e is c-linear, $(1-\lambda)x + \lambda y \in S_e \forall x, y \in S_e$ $\forall \lambda \in [0, 1]$. Consider

$$T_e : [0,1] \times S_e \times S_e \to S_e,$$
$$(\lambda, p, q) \to (1-\lambda)p + \lambda q.$$

Lemma 4.2: (S_e, T_e) is a P-convex structure.

Proof:

(a) 0-endpoint condition: $(1-0)p + 0 = 1p + 0 = p$;

(b) C-commutativity: $(1-\lambda)p + \lambda q = [1-(1-\lambda)]q + (1-\lambda)p$;

(c) C-associativity: $(1-\lambda)p + \lambda[(1-\mu)q + \mu r]$
$= [(1-\lambda)p + \lambda(1-\mu)q] + \lambda\mu r$;

(d) Point convexity condition: $(1-\lambda)p + \lambda p$
$= (1-\lambda+\lambda)p = p$.

Lemma 4.3: $B = (S_e)^*$.

Proof: Obviously, $(S_e)^* \subseteq B$. Let $x \in B$. If $e(x) = 0$, then $x = 0 \in (S_e)^*$. If $e(x) \neq 0$, then $[e(x)]^{-1} x \in S_e$, so $x \in (S_e)^*$.

Define $p + q = 2\langle \frac{1}{2}, p, q \rangle$ for the convex prestructure (S, T).

Lemma 4.4: For a generated P-convex structure (S, T):

(a) if $[x, y]$ is Class III and $x + y = x$, then $y = 0$;

(b) if $[x, y]$ is Class I or II and $x + y = x$, then $x = y = 0$.

Proof: (a) If $[x, y]$ is Class III, then $CH(\{x, y, 0\})$ is full. Thus $\langle \lambda, p, q \rangle = \langle \lambda, p, r \rangle$ implies $\lambda = 0$ or $q = r$, for any $p, q, r \in CH(\{x, y, 0\})$, and $\langle \frac{1}{2}, x, y \rangle = \frac{1}{2}x = \langle \frac{1}{2}, x, 0 \rangle$ implies $y = 0$.

(b) If $[x, y]$ is Class I, then either (i) $x = \langle x, y \rangle$ or (ii) $y = \langle x, y \rangle$. If (i) holds, $x + y = x = 2x$. If (ii) holds, $x + y = x = 2y = 3y$. If $[x, y]$ is Class II, then $x + y = 2z = x = 3x/4$ for some $z \in CH(\{x, y, 0\})$ so that $x = y = 0$.

Lemma 4.5: If (S, T) is a generated convex line-based structure, then $\forall x, y \in S$, $x + y = 0$ implies $x = y = 0$.

Proof: $x + y = 2\langle \frac{1}{2}, x, y \rangle = 0$, so $\langle \frac{1}{2}, x, y \rangle = 0$ and $\langle \lambda, x, 0 \rangle = \langle \lambda/2, x, y \rangle$, $\forall \lambda \in [0, 1]$. Therefore, $\forall \lambda \in [0, 1]$ and $\forall \mu > 0$

$$\langle \lambda, x, \mu 0 \rangle = (1+\lambda\mu)\left\langle \frac{\lambda\mu}{1+\lambda\mu}, x, 0 \right\rangle.$$

Let $\mu = 1/(2-\lambda)$; then

$$\frac{2}{2-\lambda}\langle \tfrac{1}{2}\lambda, x, 0 \rangle = \langle \lambda, x, 0 \rangle = \langle \tfrac{1}{2}\lambda, x, y \rangle.$$

If $\lambda = 1$, then $2\langle \frac{1}{2}, x, 0 \rangle = 0 = x + 0 - 0$ and $y = 0$.

Theorem 4.6: There exist a one-to-one correspondences between statistical figures and P-convex structures and between the class of sets of beams and the class of generated P-convex structures.

Proof: By Lemmas 4.2 and 4.3, if S_e is a statistical figure, then it is a P-convex structure and the set of beams is a generated P-convex structure. To show the converse, it is sufficient to prove that every generated P-convex structure (S, T) satisfies the axioms on combinations and multiplication by nonnegative numbers, because the map e given by $e(\lambda x) = \lambda$ for $x \in S_1$ is a standard quantum detector and $S_e = S_1$, where (S_1, T_1) is the generator of (S, T). It has been shown [Ref. 23, Lemma 2.9] that a generated convex structure given by $p + q = 2\langle \frac{1}{2}, p, q \rangle$ satisfies

(a) $p + q = q + p$,

(b) $p + 0 = p$,

(c) $\lambda p + \lambda q = \lambda(p + q)$,

(d) $p + (q + r) = (p + q) + r$.

By definition, $1p=p$, $0p=0$, and $\lambda(\mu p)=(\lambda\mu)p$. To show that $(\lambda+\mu)p=\lambda p+\mu q$, one simply applies the definition of $p+q$. If $\lambda=\mu=0$, then the result is trivial. If λ and μ are not both zero, then

$$\lambda p+\mu p=2\langle\tfrac{1}{2},\lambda p,\mu p\rangle$$
$$=2\,\frac{\lambda+\mu}{2}\left\langle\frac{\mu}{\lambda+\mu},p,p\right\rangle$$
$$=(\lambda+\mu)p.$$

All that remains to be proved is that $p+q+r=p$ implies $q=r=0$; This follows immediately from Lemmas 4.4 and 4.5.

5. FILTERS

The concept of a filter, yes—no, experiment or question is a crucial one in axiomatic quantum mechanics. Mackey states as an axiom that the set of all questions under a given ordering is isomorphic to the poset of all closed subspaces of a separable, infinite-dimensional Hilbert space.[5] As Mackey indicated, this assumption is made simply because it gives a theory that "explains physical phenomena and successfully predicts the results of experiments" (Ref. 5, p. 72). Mielnik has defined a filter for a general information system (D,T,B) if (1) B can be embedded in a linear space and (2) there exists a standard quantum detector $e\in D$.[14] Such a system will be called a *quantum system* (e,D,T,B). If we have a general information system (D,T,B) with a standard quantum detector e, then (e,D,T,B) will be a *quantal information system*.

For a quantum system (e,D,T,B), Mielnik defines a *filter* $a\in T$ as a linear transmitter such that

1. $eax\leq ex$, $\forall x\in B$;
2. $eax=ex$ implies $ax=x$;
3. if $b\in\{c\in T: c$ is linear and $cx=x$ if $ax=x\}$ and $ebx\leq eax\,\forall x\in B$, then $eb=ea$.[14]

Obviously, a filter a can be defined for a quantal information system (e,D,T,B) if $a\in T$ is a C-linear transmitter such that $a(\lambda x+\mu y)=\lambda ax+\mu ay\ \forall\lambda,\mu\geq 0$ and $\forall x,y\in B$. In terms of the statistical figure S_e, condition e above simply states that if $x\in S_e$ and $ax\in S_e$, then $ax=x$. For a filter $a\in T$, if $b\in\{c\in T: c$ is linear and $cx=x$ if $ax=x\}$ implies $ea<eb$, then a is an *absolute filter*. Mielnik has shown that for a quantum system and for any filter a, the sets $B_{a,I}=\{x\in B: ax=x\}$ and $B_{a,0}=\{x\in B: ax=0\}$ are closed extreme subsets of B in the weakest topology in which all detectors are continuous; further, he has shown that $S_{a,I}=\{x\in S_e: ax=x\}$ and $S_{a,0}=\{x\in S_e: ax=0\}$ are closed extreme subsets of S_e in the same topology (Ref. 14 Proposition 3). In this discussion, we shall deal solely with full P-convex structures. A convex prestructure (S,T) is *complete* if for every $x,y\in S$, $\exists p,q\in S$ that form a maximal line $[p,q]$ such that $[x,y]\subseteq[p,q]$. Such a maximal line will be denoted by $\overline{[x,y]}$.

Obviously, if (S,T) is a full P-convex structure then it is possible to embed S into a complete full P-convex structure $(\bar{S},\bar{T})$. $(\bar{S},\bar{T})$ is a *completion* of (S,T).

Lemma 5.1: For a quantal information system (e,D,T,B) and a filter a, if $x\in S_{a,0}$ and $\exists\, y,z\in S$ such that $x\in\langle y,z\rangle$, then $\overline{[y,z]}\subseteq S_{a,0}$.

Proof: If $x\in\langle y,z\rangle$, $\exists\,\lambda\in(0,1)$ such that $x=\langle\lambda,y,z\rangle$. Thus $ax=(1-\lambda)ay+\lambda az=0$ and by Theorem 4.6 $(1-\lambda)ay=0=\lambda az$, so $ay=az=0$ and $[y,z]\subseteq S_{a,0}$. If $p\in\overline{[y,z]}-[y,z]$, assume $p\in R_{y,z}$ so that $z=\langle\mu,p,y\rangle$ for some $\mu\in(0,1)$. Then $az=(1-\lambda)ap+\lambda ay=0$ and either $(1-\lambda)ap=0$ or $ap=0$; therefore $\overline{[y,z]}\subseteq S_{a,0}$.

Lemma 5.2: Let (e,D,T,B) be a quantal information system such that the statistical figure S_e is a full P-convex structure. For a filter $a\in T$, if $x\in S_{a,I}$ and $\exists\, y,z\in S$ such that $x\in\langle y,z\rangle$, then $\overline{[y,z]}\subseteq S_{a,I}$.

Proof: Assume $ax=x$, $ay=\mu p$, and $az=\nu q$ for some $p,q\in S$ and $\mu,\nu\geq 0$. Since $ear<ear\ \forall r\in B$, $\mu,\nu\in[0,1]$ so that $\langle\lambda,y,z\rangle=x$ for some $\lambda\in(0,1)$ and $\langle\lambda,y,z\rangle=\langle\lambda,\mu p,\nu q\rangle$. If $(1-\lambda)\mu+\lambda\nu=0$, then $x=0$, a contradiction since $0\notin S$. Thus

$$\langle\lambda,y,z\rangle=[(1-\lambda)\mu+\lambda\nu]\left\langle\frac{\lambda\nu}{(1-\lambda)\mu+\lambda\nu},p,q\right\rangle$$

so that $(1-\lambda)\mu+\lambda\nu=1$. Since $\lambda\in(0,1)$, $\mu=\nu=1$ so that $ay=p$ and $az=q$; but a is a filter, so $y=p$ and $z=q$. Then $ay=y$ and $az=z$, so $y,z\in S_{a,I}$. Let $t\in\overline{[y,z]}$ and assume $t\in R_{y,z}$. If $t=\langle\lambda,y,z\rangle$ for some $\lambda\in[0,1]$, then $at=a\langle\lambda,y,z\rangle=\langle\lambda,ay,az\rangle=\langle\lambda,y,z\rangle=t$. If $y=\langle\lambda,t,z\rangle$ for some $\lambda\in[0,1]$, then $y=ay=\langle\lambda,at,az\rangle=\langle\lambda,t,x\rangle$ and since S is full, $ap=p$ so that $\overline{[x,y]}\subseteq S_{a,I}$.

Lemma 5.3: If (e,D,T,B) is a quantal information system and the statistical figure is a P-convex structure and a is a filter, then $x\in S_{a,0}$ and $y\in S_{a,I}$ imply $[x,y]=\overline{[x,y]}$.

Proof: Assume $p\in\overline{[x,y]}$. If $x=\langle\lambda,p,y\rangle$ for some $\lambda\in(0,1)$, then $ax=\langle\lambda,ap,ay\rangle$ so that $0=(1-\lambda)ap+\lambda y$. By Theorem 4.6, $y=ap=0$, a contradiction as $y\in S$ and $0\notin S_e$. If $y=\langle\lambda,p,x\rangle$ for some $\lambda\in(0,1)$, then $y=(1-\lambda)ap$, a contradiction since $e[(1-\lambda)ap]<1$ and $ey=1$. Thus $[x,y]=\overline{[x,y]}$.

Let ϕ be the isomorphism that maps the full P-convex structure (S,T) onto a convex subset of a real linear space V. A *flat* is a subset $A\subseteq S$ of a P-convex structure (S,T) such that $\phi(A)=C\cap\phi(S)$ where C is a subspace of V. A flat A is *extremal* if $\phi(A)$ is an extreme subset of S. From the above three lemmas, we obtain the following:

Theorem 5.4: If (e,D,T,B) is a quantal information system with a statistical figure S that is a full P-convex structure, then for any filter $a\in T$, $S_{a,I}$ and $S_{a,0}$ are extremal flats.

Up to this point, we have not used the "minimal entropy" condition: if $b\in T$ such that $eb<ea$ and $ax=x$ implies $bx=x$, then $eb=ea$. We shall now use this condition to prove a very important result. If (e,D,T,B) is a quantal information system with a statistical figure that is a complete, full P-convex structure, then (e,D,T,B) is *discriminating*. Obviously, if (e,D,T,B) is a discriminating quantal information system, then $D_I=\{b\in B: eb\leq 1\}$ is a complete, full P-convex structure and any filter maps D_I into D_I. If $x\in S$ and a is a filter, then $ax=\lambda y$ for some $\lambda\in[0,1]$ and some $y\in S$. A filter is idempotent (Ref. 14, Proposition 2), so

$a^2x = a(\lambda y) = \lambda y$ and $ay = y$ if $\lambda \neq 0$; thus $y \in S_{a,I}$.

Theorem 5.5: If (e, D, T, B) is a discriminating quantal information system with the statistical figure S, then for any filter a

$$S = CH(S_{a,I} \cup S_{a,0}).$$

Proof: Let $x \in S$. For a filter a, $ax = \lambda y$ for some $\lambda \in [0,1]$ and some $y \in S$. If $\lambda = 0$, then $x \in S_{a,0}$, if $\lambda = 1$, then $x \in S_{a,I}$; thus we need only concern ourselves with $\lambda \in (0,1)$. S is complete, so by Lemma 5.3 $\exists p \in S$ such that $\overline{[x,y]} = [p,y]$. Thus $\exists \sigma \in (0,1)$ such that $x = \langle \sigma, p, y \rangle$, so that $ax = \langle \sigma, ap, ay \rangle$ or $\langle \sigma, ap, y \rangle = (1-\sigma)ap + \sigma y$ since a is idempotent. It is sufficient to show that $\lambda = \sigma$, for then Theorem 4.6 implies $(1-\lambda)ap = 0$, so that $ap = 0$ and $p \in S_{a,0}$. From the definition of a standard quantum detector, one obtains $e\lambda y = e(1-\sigma)ap + e\sigma y$. Then either $\lambda = (1-\sigma)eap + \sigma$ or

$$eap = \frac{\lambda - \sigma}{1 - \sigma} \geq 0;$$

thus $\lambda \geq \sigma$. Define the map $b : S \to S$ such that

(a) $br = r$ if $r \in a_n = \{q \in S : aq = q\}$

(b) $br = 0 bp = 0$ if $r \in ECH(S_{a,0} \cup \{p\})$

and extend b so that it is affine on $CH(a_n \cup S_{a,0} \cup \{p\})$; then $bx = x$. Further define $b(\lambda q) = \lambda bq \; \forall \lambda \geq 0 \; \forall q \in S$ so that b is linear on B. By minimal entropy condition on a, $ea = eb$; therefore, $eap = ebp = 0$, so that $\lambda = \sigma$. The following theorem is now clear.

Theorem 5.6: The set of extremal flats, ordered by set inclusion, has a subset that is isomorphic to the set of filters ordered by the standard quantum detector e.

6. REDUCTION MAPS

A map ϕ is a *reduction map* on a convex prestructure (S, T) if $\phi \in Af(S)$ and $\phi^2 = \phi$, where $\phi^2(a)$ means $\phi(\phi(a))$. If $\phi(a) = a \; \forall a \in S$, ϕ is the *trivial reduction map*; if $\phi(a) = x \; \forall a \in S$ and for some fixed $x \in S$, then ϕ is a *total reduction*. Obviously, there are as many total reductions in $Af(S)$ as there are points in S. The main motivation behind the definition of a reduction map is the concept of a projection in a linear space. In general, however, there does not exist a zero element in a convex prestructure; for this and other reasons, a reduction map is more general than a projection in a linear space. One can easily define a partial ordering on the set $R(S)$ of reduction maps of S by $\phi \leq \psi$ iff $\phi(S) \subseteq \psi(S)$ for $\phi, \psi \in R(S)$. Obviously, $R(S)$ is not a lattice, as $\phi \wedge \psi$ may not exist. For example, for two distinct total reduction maps ϕ_x and ϕ_y, $\phi_x \wedge \phi_y$ does not exist.

Lemma 6.1: If $\phi \in R(S)$ for a P-convex structure (S, T) and if $\phi(b) = b$ and $\phi(c) = c$ for some $b, c \in S$, then $\phi(a) = a \; \forall a \in [b, c]$.

Proof: Assume $a = \langle \lambda, b, c \rangle$ for some $\lambda \in [0,1]$. Then $\phi(a) = \langle \lambda, \phi(b), \phi(c) \rangle = \langle \lambda, b, c \rangle = a$.

Lemma 6.2: If $\phi \in R(S)$ for a full P-convex structure (S, T) and if $\phi(b) = b$ and $\phi(c) = c$ for some $b, c \in S$, then $\phi(x) = x \; \forall x \in \overline{[b, c]}$.

Proof: By Lemma 6.1 $\phi(x) = x \; \forall x \in [b, c]$. If $x \in \overline{[b, c]} - [b, c]$, assume $\exists \lambda \in (0,1)$ such that $b = \langle \lambda, x, c \rangle$; then $\phi(b) = \langle \lambda, \phi(x), \phi(c) \rangle = \langle \lambda, \phi(x), c \rangle = \langle \lambda, x, c \rangle$, so $x = \phi(x)$. The proof is similar if $\exists \lambda \in (0,1)$ such that $c = \langle \lambda, x, b \rangle$.

Two sets that are of some interest in P-convex structures are $CH(A)$ and $ECH(A)$ where, for $A \subseteq S$, $CH(A)$ is the smallest convex substructure containing A and $ECH(A)$ is the smallest flat containing A. Therefore, given a full P-convex structure (S, T) and a reduction map ϕ on S, there corresponds a set $A = \phi(S)$ such that $A = ECH(A)$. If $A = ECH(A)$, then A is a *flat*. The following theorem is immediate.

Theorem 6.3: For a P-convex structure (S, T), if A and B are flats, $A \cap B$ $[ECH(A \cup B)]$ is the largest (smallest) flat contained in (containing) A and B. The set $F(S)$ of flats of a dense P-convex structure forms a lattice under set inclusion. Therefore, $R(S)$ is isomorphic to a subset of the lattice $F(S)$ of flats for a full P-convex structure (S, T).

Corollary 6.4: The lattice $F(S)$ of flats of a dense P-convex structure (S, T) is atomic.

Proof: $ECH(\{p\}) = \{p\}$, so every point is a flat; thus the atoms of a dense P-convex structure are the atoms of $F(S)$.

7. MACKEY'S AXIOMS OF QUANTUM MECHANICS

A standard approach to axiomatic quantum mechanics is that used by Mackey.[5] As will be shown, Mackey's system carries over naturally to that of a full P-convex structure. Mackey is concerned with the set B of all Borel subsets of the real line R and with two undefined sets: S, the set of all states, and O, the set of all observables. There exists a map

$$p : O \times S \times B \to [0,1],$$
$$(A, x, E) \to x_A(E)$$

satisfying the axioms below.

1. $x_A(\emptyset) = 0$ and $x_A(R) = 1 \; \forall x \in S$ and $\forall A \in O$, and $x_A(\cup_1^\infty E_i) = \sum_1^\infty x_A(E_i)$ whenever the E_i are pairwise disjoint Borel sets. In other words, each x_A is a probability measure on B.

2. If $x_A(E) = x_{A'}(E) \; \forall x \in S$ and $\forall E \in B$, then $A = A'$; and if $x_A(E) = x'_A(E) \; \forall A \in 0$ and $\forall E \in B$, then $x = x'$.

3. For any real-valued function f on R and any $A \in 0$, $\exists A' \in 0$ such that $x_{A'}(E) = x_A(f^{-1}(E))$ $x \in S$ and $E \in B$. A' is denoted by $f(A)$ in the literature.

4. For $x_1, x_2, \ldots, \in S$ and $\lambda_1, \lambda_2, \ldots, \in [0,1]$ such that $\sum^\infty \lambda_i = 1$, $\exists y \in S$ such that

$$y_A(E) = \sum_1^\infty \lambda_i (x_i)_A(E),$$

$\forall E \in B$ and $\forall A \in 0$; y shall be denoted by $\sum_1^\infty \lambda_i x_i$. A *question* is an observable A such that $\forall x \in S$,

$$x_A(\{0,1\}) = 1.$$

If we define a quantity $m_x(q) = x_q(\{1\})$ for each question q and each $x \in S$, then $m_x(q)$ can be used to define a natural partial ordering in Q, the set of all questions, by $q_1 \leq q_2$ iff $m_x(q_1) \leq m_x(q_2) \; \forall x \in S$. If $m_x(q_1) + m_x(q_2) \leq 1 \; \forall x \in S$, then the questions q_1 and q_2 are *disjoint*. We next assume the axiom:

5. If $\{q_i\}$ is any pairwise disjoint sequence of ques-

tions, then $\sum_1^\infty q_i$ exists. Let q_E^A be the question represented by the function $\phi_E : A \to \{0,1\}$ which is 1 for $x \in E$ and is 0 for $x \notin E$. Any function $B \to O$, $E \to q_E$ is a *question-valued measure* if

(a) $m_x(q_\phi) = 0$ and $m_x(q_R) = 1$ $\forall x \in S$;

(b) $E \cap F = \emptyset$ implies q_E and q_F are disjoint;

(c) $E_i \cap E_j = \emptyset$ for $i \neq j$ implies $q(E_1 \cup E_2 \cup \cdots) = q_{E_1} + q_{E_2} + \cdots$.

We also assume the axioms:

6. If $q : E \to q_E$ is any question-valued measure, then $\exists A \in 0$ such that $q_E^A = q_E$ $\forall E \in B$. The assumptions listed below are also given in Mackey.[5]

7. The poset of all questions in quantum mechanics under the natural partial ordering is isomorphic to the partially ordered set of all closed subspaces of a separable, infinite-dimensional Hilbert space.

8. If q is any question that is represented by a Borel function that has nonzero values, then $\exists x \in S$ such that $m_x(q) = 1$.

There is an additional axiom regarding quantum dynamics which will not be discussed here.

A possible alternative for Axiom 2 above is the following assumption:

2′. If $m_x(A) = m_x(C)$ $\forall x \in S$, then $A = C$.

For a physical motivation of these axioms and their development, the reader is referred to Mackey.[5]

Using the concept of convex prestructures to revise Mackey's axioms, we shall show some additional physical motivation for them. Axiom 1 and 3 will be left as stated above. Mackey admits that Axiom 7 "seems entirely *ad hoc*" (Ref. 5, p. 71). A modification of Axiom 7, equally well suited to physical phenomena, will be indicated in this paper.

To replace Axioms 2, 4, and 7 we shall use certain axioms discussed below.

A. There exists a map

$$T : [0,1] \times S \times S \to S,$$

$$(\lambda, x, y) \to \langle \lambda, x, y \rangle$$

on the set of states such that

$$\langle \lambda, x, y \rangle_A(E) = (1 - \lambda) x_A(E) + \lambda y_A(E)$$

and (S, T) is a convex prestructure such that $p(A, \cdot, E)$ is an affine functional of S for each $A \in 0$ and each $E \in B$.

B. There exists a topology on S such that the sequence

$$\langle \lambda_1, x_2, x_1 \rangle,\ \langle \lambda_2, x_3, \langle \lambda_1, x_2, x_1 \rangle \rangle,\ \langle \lambda_3, x_4, \langle \lambda_2, x_3, \langle \lambda_1, x_2, x_1 \rangle \rangle \rangle,\ \cdots$$

converges whenever $\lambda_i \in [0,1]$ and $x_i \in S$ for $i = 1, 2, \ldots$.

Obviously, these two axioms are a rewording of Mackey's Axiom 4 in a language compatible with our development of convex prestructures. Axiom 2 will be replaced by Axioms C and D below.

C. If $x_A(E) = x_{A'}(E)$ $\forall x \in S$ and $\forall E \in B$, then $A = A'$.

D. There exists an equivalence relation on (S, T) such that x is equivalent to x' iff $x'_A(E) = x_A(E)$ $\forall A \in 0$ and $\forall E \in B$. Therefore, if $\langle \lambda, x, y \rangle = \langle \lambda, x, z \rangle$ for some $\lambda \in (0,1]$ and for some $x \in S$, then y should be equivalent to z; this will be denoted by $y * z$. Clearly $y * y$. Also $y * z$ implies $z * y$ for any convex prestructure. Next consider transitivity: if $x * y$ and $y * z$, then $\exists p \in S$ and $\exists \lambda \in (0,1]$ such that $\langle \lambda, p, x \rangle = \langle \lambda, p, z \rangle$. From this it seems natural to assume the axiom:

E. (S, T) is a P-convex structure.

One of the authors has shown that it is possible to define an equivalence relation, called *indistinguishability (mod S)* between two points x and y in a P-convex structure iff there exists a point $z \in S$ and $\lambda \in (0,1)$ such that $\langle \lambda, x, z \rangle = \langle \lambda, y, z \rangle$.[22] The set of indistinguishable (mod S) equivalence classes $S_{m,s}$ can have a map $T_{m,s}$ defined in terms of T such that $(S_{m,s}, T_{m,s})$ forms a full P-convex structure.

Theorem 7.1: Let (S, T) be a P-convex structure. Then (a) S^* separates $S_{m,s}$ equivalence classes, and (b) $\rho(x, y) = 0$ implies x and y are indistinguishable (mod S) if $S_{m,s}$ is isomorphic to a bounded convex set of a real linear space. Indistinguishability (mod S) gives an equivalence relation in which y is equivalent to z iff

$$\langle \lambda, x, y \rangle_A(E) = \langle \lambda, x, z \rangle_A(E)$$

for some $\lambda \in [0,1]$, some $x \in S$, and all $E \in B$. $(S_{M,S}, T_{M,S})$ with Axiom B above is therefore equivalent to (O, S, B, p) with Mackey's Axiom 2 and 4.

Unfortunately, physical measurements cannot indicate whether (S, T) is full, as measurements on indistinguishable states give identical results as can be seen from Theorem 7.1. The major difficulty is that a state x is simply a member of an undefined set S. Dirac defines a state as "an undisturbed motion that is restricted by as many conditions or data as are theoretically possible without mutual interference or contradiction (Ref. 26, p. 11)." Use of this definition seems to require fullness of (S, T). Until the concept of a quantum state is defined in a manner useful to axiomatic quantum mechanics, however, it will be necessary to consider the possibility of indistinguishable quantum states. Even if (S, T) is a full P-convex structure, the isomorphism theorems in Sec. 2 do not force (S, T) to be a subset of Euclidean space, as (S, T) may be part of a complete Riemannian manifold that has a large curvature.

Given a full P-convex structure (S, T) there are five associated classes of objects: the class $F(S)$ of all flats, the class $EF(S)$ of all extremal flats, the class $R(S)$ of all reduction maps, the class $L(S)$ of all filters, and the class $Q(S)$ of all questions. They can be ordered by the one-to-one morphism between them; this ordering is shown by the diagram in Fig. 1.

Mielnik requires that $L(S)$ replace $Q(S)$,[14] so that one is not concerned with questions, but with filters. Mackey's Axiom 7 (Ref. 5) and the requirement that states be represented by unit vectors in a Hilbert space together imply that $Q(S) \cong R(S) \cong F(S)$. Mielnik and Mackey are obviously in disagreement over the concept of questions vs filters. It seems reasonable, however, to place the following restriction on the "shape" of the set of states S:

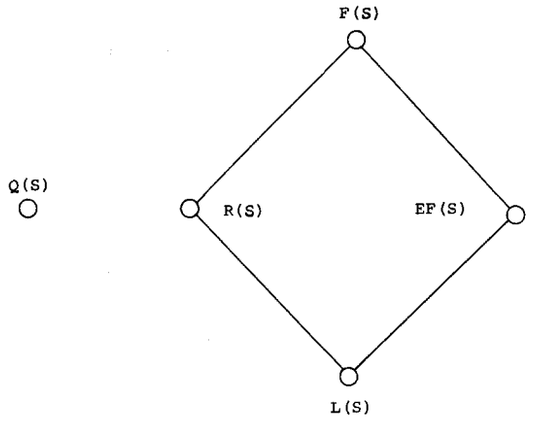


FIG. 1.

F. For the completion $(\overline{S_{M,S}}, \overline{T_{M,S}})$ of $(S_{M,S}, T_{M,S})$, $EF(\overline{S_{M,S}}) \cong L(\overline{S_{M,S}})$. The shape of the set of states will be explored by one of the authors in a future paper. An additional axiom is also required:

G. The completion $(\overline{S_{M,S}}, \overline{T_{M,S}})$ of $(S_{M,S}, T_{M,S})$ is homeomorphic to a compact convex subset of a topological linear space with a locally convex Hausdorff topology.

From the Krein–Milman Theorem, $EF(S_{M,S})$ must then be atomic, since there exist extreme points in S and every extremal flat has an extreme point. Therefore, $L(S_{M,S})$ is atomic. In the required topology, all M (mod S) equivalence classes are closed and there do not exist any proper nonempty open subsets of an equivalence class. The sequences mentioned in Axiom B therefore converge to a set of states which are all indistinguishable.

It should be noted that no reference has been made to the dimensionality of $S_{M,S}$. Mackey requires that the poset of questions be isomorphic to the lattice of subspaces of a separable, infinite-dimensional Hilbert space; however, Piron has shown that the Hilbert space H can be decomposed into mutually orthogonal subspaces $\{H_a\}_{a\in A}$ by some superselection rule.[27] The indexing set A may be continuous, as in the case of time, so that H need not be separable, even if each H_a is. For this reason, the dimensionality of S hasn't been discussed.

We conclude this section with a proposed set of axioms for a quantum mechanical system.

Axiom 7.1: x_A is a probability measure on the Borel sets of the real line R for each $x \in S$ and each $A \in 0$.

Axiom 7.2: If $x_A(E) = x_{A'}(E)\ \forall\, x \in S$ and $\forall E \in B$, then $A = A'$.

Axiom 7.3: For any real-valued Borel function f on R and any $A \in 0$, $\exists A' \in 0$ such that $x_{A'}(E) = x_A(f^{-1}(E))$ $x \in S$ and $\forall\, E \in B$.

Axiom 7.4: $\exists\, T : [0,1] \times S \times S \to S$ such that (S, T) is a P-convex structure and the $p(A, \cdot, E)$ are affine maps on $S\ \forall A \in 0$ and $\forall\, E \in B$.

Axiom 7.5: There exists a topology on S such that

(a) the sequence $\langle \lambda_1, x_2, x_1\rangle$, $\langle \lambda_2, x_3, \langle \lambda_1, x_2, x_1\rangle\rangle$, $\langle \lambda_3, x_4, \langle \lambda_2, x_3, \langle \lambda_1, x_2, x_1\rangle\rangle\rangle, \ldots$ converges whenever $\lambda_i \in [0,1]$ and $x_i \in S$ for $i = 1, 2, \ldots$; and

(b) the completion $\overline{S_{M,S}}$ of $S_{M,S}$ is homeomorphic to a compact, convex subset of a locally convex Hausdorff topological linear space.

Axiom 7.6: For the completion $(\overline{S_{M,S}}, \overline{T_{M,S}})$ of $(S_{M,S}, T_{M,S})$, $EF(\overline{S_{M,S}}) \cong L(\overline{S_{M,S}})$.

Axiom 7.7 If $q : E \to q_E$ is any question-valued measure, then there exists a filter q such that $q_E^A = q_E\ \forall\, E \in B$.

Axiom 7.8: If q is any filter represented by a Borel function with nonzero values, then $\exists\, x \in S$ such that $m_x(q) = 1$.

8. CONCLUSION

As can be seen from the above results, the framework of convex prestructures gives a generalizing and unifying formalism for Mielnik's convex set and Davies' and Lewis' operational methods in quantum mechanics. Also this framework lends itself naturally to a reformulation of Mackey's axioms. The authors hope that this approach will lead to greater understanding of the mathematical formalism used in axiomatic quantum theory.

[1]R. Haag, D. Kastler, J. Math. Phys. 5, 848 (1964).
[2]I. Segal, *Mathematical Problems of Relativistic Physics*, American Mathematical Society Lectures in Applied Mathematics (Providence, R.I., 1963).
[3]S. Gudder, "Axiomatic Quantum Mechanics and Generalized Probability" in *Probabilistic Methods in Applied Mathematics*, edited by A. Bharucha-Reid (Academic, New York, 1970), Vol. II.
[4]J.M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).
[5]G.W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[6]V. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, N.J., 1968), Vol. I.
[7]G. Dahn, Commun. Math. Phys. 9, 192 (1968).
[8]G. Ludwig, Commun. Math. Phys. 4, 331 (1967).
[9]G. Ludwig, Commun. Math. Phys. 9, 1 (1968).
[10]P. Stolz, Commun. Math. Phys. 11, 303 (1969).
[11]P. Stolz, Commun. Math. Phys. 23, 117 (1971).
[12]J. Gunson, Commun. Math. Phys. 6, 262 (1967).
[13]B. Mielnik, Commun. Math. Phys. 9, 55 (1968).
[14]B. Mielnik, Commun. Math. Phys. 15, 1 (1969).
[15]E.B. Davis, Commun. Math. Phys. 15, 277 (1969).
[16]E.B. Davies and J.T. Lewis, Commun. Math. Phys. 17, 239 (1970).
[17]C.M. Edwards, Commun. Math. Phys. 20, 26 (1971).
[18]C.M. Edwards, Commun. Math. Phys. 16, 207 (1970).
[19]N. Kronfli, Int. J. Theor. Phys. 3, 199 (1970).
[20]H. Neumann, Commun. Math. Phys. 23, 100 (1971).
[21]A. Peressini, *Ordered Topological Vector Spaces* (Harper and Row, New York, 1967).
[22]W.M. Cornette, Ph.D. dissertation, University of Denver (1973).
[23]S. Gudder, Commun. Math. Phys. 29, 249 (1973).
[24]J.M. Jauch and C. Piron, Helv. Phys. Acta 42, 842 (1969).
[25]B. Mielnik, private correspondence (May 1973).
[26]P.A.M. Dirac, *Quantum Mechanics* (Oxford U.P., Oxford, 1958).
[27]C. Piron, Helv. Phys. Acta 42, 330 (1969).

Representations of the coordinate transformation group

Martin S. Altschul*

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 23 October 1972; final revised manuscript received 26 December 1973)

In this paper we deal with the Lie (pseudo-) group of coordinate transformations in four dimensions. We discuss the detailed structure of the corresponding Lie algebra, especially including Theorem II, which states a limited set of structure relations that uniquely determine the structure of the algebra. Using Theorem II, we attack the problem of finding unitary representations of the pseudogroup. Finally we discuss some physical implications of these representations.

INTRODUCTION

In this article we study the structure and representation theory of the (pseudo) group which consists of all invertible, local, analytic transformations on some neighborhood in R^n. This set is closed under the operation of composition of transformations, but is not a true group because that operation is not always defined.[1] We call our (pseudo) group the complete coordinate transformation group in n dimensions (abbreviated CCTGn). Sets of transformations of this general type have been discussed by Cartan,[2] and more recently be Singer and Sternberg.[1] For physical reasons we are primarily interested in the CCTG4. We study the commutation relations of its infinitesimal generators, proving two theorems. We then apply our second theorem to the problem of describing representations of the CCTG4.

The CCTGn belongs to a class of objects known as "infinite Lie groups" or "Lie pseudogroups". These pesudogroups bear a strong similarity to ordinary Lie groups, but there is no completely satisfactory correspondence between the Lie pseudogroup and its Lie algebra. This is because "One-parameter subgroups do not fill a neighborhood of the identity in a. . . Lie (pseudo-) group."[3] The neighborhood of the identity of the "group" thus contains more information than does the algebra. However, we will deal primarily with the algebra of the CCTG4 (abbreviated ACCTG4) and will assume that in doing so we have not neglected anything of physical importance. We will find that this algebra looks just like an ordinary Lie algebra except that its dimension is infinite.

Let us now derive a Hermitian representation of the Lie pseudoalgebra of the CCTG4. There is one ready-made, faithful representation—the representation on the space of functions $f(x^\mu)$, $\mu=0,1,2,3$, where

$$x^0=t,\quad x^1=-x,\quad x^2=-y,\quad x^3=-z,$$

and (1)

$$x_0=t,\quad x_1=x,\quad x_2=y,\quad x_3=z\,.$$

We interpret $f(x^\mu)$ as a spatio-temporal probability amplitude (e.g., $f=[1/(2\pi)^2]\exp[i(p^0t+p^1x+p^2y+p^3z)]$ is the amplitude for a spinless particle with 4-momentum p^μ). If the functions f have the transformation property

$$f'(x'^\mu)=[J(x'/x)]^{-1/2}f(x^\mu)\,, \tag{2}$$

where $J(x'/x)$ is the Jacobian of the coordinate transformation from x to x', then the $f(x^\mu)$ will indeed transform under a Hermitian representation of the algebra. Since the elements of the CCTG4 are analytic, we may write

$$x'^\mu=a^\mu+\sum_\nu a^\mu{}_\nu x^\nu+\sum_{\nu\gamma}a^\mu{}_{\nu\gamma}x^\nu x^\gamma+\sum_{\nu\gamma\delta}a^\mu{}_{\nu\gamma\delta}x^\nu x^\gamma x^\delta+\cdots$$

$$(\mu,\nu,\gamma,\delta=0,1,2,3)\ \ (a^\mu,a^\mu{}_\nu,\text{ etc. are real numbers}) \tag{3}$$

in some neighborhood of the origin of the pseudogroup. If, for example, all the a's except the a^μ are zero, then the transformations are just translations, and of course the corresponding infinitesimal generators are

$$i\frac{\partial}{\partial x^\nu}=-i\partial_\nu=\frac{1}{\hbar}p_\nu\,. \tag{4}$$

In general, if all the a's except $a_\gamma{}^{\nu_1\cdots\nu_n}$ are zero, then the corresponding infinitesimal generator is

$$B_\gamma{}^{\nu_1\cdots\nu_n}=ix^{\nu_1}\cdots x^{\nu_n}\partial_\gamma+\tfrac{1}{2}i(\textstyle\sum\delta^\gamma{}_{\nu_l}x^{\nu_1}\cdots\widehat{x^{\nu_l}}\cdots x^{\nu_n}) \tag{5}$$

$$(\gamma,\nu_1,\ldots,\nu_n=0,1,2,3;\ l\text{ and }n\text{ are positive integers},\ l\leq n)\,.$$

Special cases are the familiar

$$J_i=\sum_{jk}\frac{1}{i}\epsilon_{ijk}(x_j\partial_k)\ \ (i,j,k=1,2,3)\,, \tag{6}$$

$$J_i=(1/\hbar)L_i\,,$$

$$K_i=\frac{1}{i}\left(x_0\frac{\partial}{\partial x_i}+x_i\frac{\partial}{\partial x_0}\right)\ \ (i=1,2,3)\,, \tag{7}$$

$$D^\nu=i\left(\sum_{\mu=0}^{3}x_\mu x^\mu\partial^\nu-2\sum_{\mu=0}^{3}x^\nu x_\mu\partial^\mu-4x^\nu\right)\ \ (\nu=0,1,2,3)\,, \tag{8}$$

$J_i=$ rotation generator,

$K_i=$ Lorentz generator,

$D^\nu=$ conformal generator.

We also define the operators

$$A_\gamma{}^{\nu_1\cdots\nu_n}=ix^{\nu_1}\cdots x^{\nu_n}\partial_\gamma\,, \tag{9}$$

$$C_\gamma{}^{\nu_1\cdots\nu_n}=(i)^n p^\gamma\partial_{p_{\nu_1}}\cdots\partial_{p_{\nu_n}}+\tfrac{1}{2}(i)^n(\textstyle\sum\delta^\gamma_i\partial_{p_{\nu_1}}\cdots\hat{\partial}_{p_{\nu_i}}\cdots\partial_{p_{\nu_n}}) \tag{10}$$

$$(\gamma,\nu_1,\ldots,\nu_n=0,1,2,3)\,.$$

The C's are equivalent to the B's, but they are given in terms of conjugate momenta coordinates. The A's are non-Hermitian operators, but they satisfy the same commutation relations as the B's and C's.

Let us have a graphic look at the ACCTG4. We group the elements of the algebra as follows:

$i\partial_\mu$ 0th row

$ix^\mu\partial_\nu$ 1st row

$ix^\mu x^\nu\partial_\gamma$ 2nd row

$ix^\mu x^\nu x^\gamma\partial_\epsilon$ 3rd row

etc., where $\mu,\nu,\gamma,\epsilon=0,1,2,3$. When we say that an operator is in the nth row, we mean that it is a linear combination of operators shown explicitly as in the nth row above. Each row as a vector space is finite-dimensional. Note that the ACCTG4 includes as finite subalgebras the rotation group, the Lorentz group, the conformal group $[SO(4,2)]$, and also $SL(4,r)$ and $GL(4,r)$.

FIRST THEOREM

Theorem I: The ACCTG4 has a finite generating subset of two elements. These elements are [in the faithful "A" representation given by Eq. (9)]

$$G_1=i\partial_x,$$

$$G_2=i[x^7(x^2\partial_x)+x^5(x\partial_y)+x^4(y\partial_z)+x^2(z\partial_t)+(t\partial_x)].$$

This theorem can be generalized to the ACCTG in n dimensions, with generators

$$G_1=i\partial_{x_1},$$

$$G_2=i(x_1^{2n+1}\partial_{x_1}+x_1^{2n-2}\partial_{x_2}+x_1^{2n-4}x_2\partial_{x_3}+x_1^{2n-6}x_3\partial_{x_4}+\cdots+x_n\partial_{x_1}).$$

Proof of Theorem I: The most general element of the ACCTG4 is a linear combination of elements of the form

$$ix_0^{n_0}x_1^{n_1}x_2^{n_2}x_3^{n_3}\partial_j \quad (j=0,1,2,3)$$

$(n_j\in Z^*=$the set of nonnegative integers$)$.

Starting from $ix_0^{\alpha}\partial_0$ $(\alpha\in Z^*)$ and commuting with

$ix_1\partial_0$ n_1 times,

$ix_2\partial_0$ n_2 times,

$ix_3\partial_0$ n_3 times,

we generate the operator

$$i(-1)^{(n_1+n_2+n_3)}\frac{\alpha!}{n_{0!}}x_0^{n_0}x_1^{n_1}x_2^{n_2}x_3^{n_3}\partial_0$$

if $n_0=\alpha-n_1-n_2-n_3\geq 0$.

(Please note that α and the n_j are exponents, not superscripts.) Hence given $ix_0^{\alpha}\partial_0$ $(\forall\alpha\in Z^*)$ and $ix_j\partial_0$ $(j=1,2,3)$, we can generate all the

$$ix_0^{n_0}x_1^{n_1}x_2^{n_2}x_3^{n_3}\partial_0.$$

Also, given $i\partial_0$ and $ix_0^3\partial_0$ it is easy to generate all the other $ix_0\partial_0$. By similar development for $ix_0^{n_0}x_1^{n_1}x_2^{n_2}x_3^{n_3}\partial_j$ $(j=1,2,3)$, we see that the entire algebra is generated by

$$i\partial_j, ix_j^3\partial_j, ix_j\partial_k \quad (j,k=0,1,2,3;\ j\neq k). \tag{11}$$

Noting the equation

$$i(ix_j^3\partial_j)=[ix_j^2\partial_k, i(x_kx_j\partial_j+x_k^2\partial_k)] \quad (j\neq k), \tag{12}$$

it is readily seen that the operators (11) are generated by

$$\partial_x, x\partial_y, y\partial_z, z\partial_t, t\partial_x, x^2\partial_x$$

and these in turn are generated by repeated commutation of G_1 with G_2, and some elementary algebraic manipulations.

SECOND THEOREM

A bracket expression is any commutator of commutators of commutators, etc. A bracker expression of order n contains n (not necessarily distinct) elements of a Lie algebra; e.g.,

$$[A_1,[[A_2,[A_3,A_4]],A_5],[A_6,A_7]]$$

is of order 7. A proper commutator of order n is an object of the form

$$[A_1,[A_2,[A_3,[\ldots,[A_{n-1},A_n]\ldots]$$

where $A_1,\ldots,A_n$ are elements of a Lie algebra.

From Theorem I we know that the ACCTG4 is generated by a finite set of elements; which is to say that each element of the algebra can be given in terms of bracket expressions of the elements of the generating set. Moreover, any nth order bracket expression (in elements $A_1,\ldots,A_n$) can be shown (by repeated use of the Jacobi identity) to be a sum of nth order proper commutators of $A_1,\ldots,A_n$. Hence, any commutator equation

$$[A,B]=iC \quad A,B,C\in\text{ACCTG4} \tag{13}$$

can be reduced to an equation between sums of proper commutators of the members of any generating set.

Suppose we have a Lie algebra V whose elements are matrix or differential operators, or both, and we wish to demonstrate that it constitutes a representation Γ of the ACCTG4. Then we must show that the operators satisfy the same commutation relations as the elements that they are supposed to represent. Suppose that we want a_j (where $a_j\in V$; $j=1,m$) to represent the elements of a generating set. From the previous paragraph it is clear that if all the appropriate equations

$$[a_j,X]=iY \quad (Y\in V,\ \forall X\in V) \tag{14}$$

required by the structure of algebra are satisfied, then all the appropriate relations

$$[A,B]=iC \quad (A,B,C\in V)$$

will be satisfied, and V will indeed represent the ACCTG4.

Equations (13) are the structure relations of the algebra. We have shown that they are satisfied if (14) are satisfied. Theorem II states an even more restricted set of commutation relations, which if satisfied insure that (14) and in turn (13) are satisfied. Thus, Theorem II gives a criterion which can be used to established that a set of operators is a representation of the ACCTG4.

Theorem II: Let V be a Lie algebra as above, and let "$ix^{\nu_1}\cdots x^{\nu_n}\partial_\gamma$" signify the element of V which we want to correspond to $B_\gamma^{\nu_1\cdots\nu_n}$. (We use the form "$ix^{\nu_1}\cdots x^{\nu_n}\partial_\gamma$" in order to enhance the reader's insight into the formulas that follow—for computational simplicity we will also drop the ubiquitous factors of "i" and use just "$x^{\nu_1}\cdots x^{\nu_n}\partial_\gamma$.") Then V will represent the ACCTG4 if the

following equations are satisfied:

$$[x_j\partial_k, \partial_l]=0, \quad j\neq k,\ k\neq l,\ j=l, \tag{15a}$$

$$[x_j\partial_k, x_k\partial_l]=x_j\partial_l, \quad j\neq k,\ k\neq l,\ j\neq l, \tag{15b}$$

$$[x_j\partial_k, x_l\partial_q]=0, \quad j\neq l \text{ or } q,\ k\neq l, \tag{15c}$$

$$[x_j\partial_k, x^2\partial_l]=0, \quad j\neq k,\ k\neq l,\ j\neq l, \tag{15d}$$

$$\begin{aligned}&[x_j{}^m x_k{}^n\partial_l, x_j{}^r x_k{}^s\partial_q]\\ &=\delta_l^j r x_j{}^{m+r-1}x_k{}^{n+s}\partial_q+\delta_l^k s x_j{}^{m+r}x_k{}^{m+s-1}\partial_q\\ &\quad-\delta_q^j m x_j{}^{m+r-1}x_k{}^{n+s}\partial_l-\delta_q^k n x_j{}^{m+r}x_k{}^{m+s-1}\partial_l\end{aligned} \tag{15e}$$

(where $m, n, r, s\in Z^+$ are exponents; and for (15e) $l=j$ or k, $q=j$ or k). Equations (15e) are just the structure relations of the ACCTG in two dimensions. Obviously, as a further refinement of Theorem II, (15e) can be replaced by the commutator relations involving any generating set of the ACCTG2 [cf. Eqs. (14)].

Proof of Theorem II: Our strategy in proving this theorem is to begin with (14) and eliminate equations as redundant if they can be derived from other structure relations by use of the Jacobi identity. We must not make use of properties which are specific to the form "$x^{\nu_1}\cdots x^{\nu_n}\partial_\gamma$."

We begin by selecting a generating subset of the ACCTG4. We do not use G_1 and G_2 but rather

$$\partial_j,\ x_j^2\partial_j,\ x_j\partial_k \quad (j,k=0,1,2,3;\ j\neq k). \tag{16}$$

We then select a set of canonical definitions of the operators $x^{\nu_1\cdots\nu_n}\partial_\gamma$, in terms of (16). If a set of operators is shown to represent the ACCTG4, then in retrospect the particular choice of definitions is mute. Recalling Eq. (12) we can define $x_j^3\partial_j$ easily in terms of (16). We define

$$\begin{aligned}x_j^4\partial_j&=[x_j^2\partial_j, x_j^3\partial_j],\\ x_j^n\partial_j&=\frac{1}{n-3}[x_j^2\partial_j, x_j^{n-1}\partial_j]\\ &=\frac{1}{(n-3)!}[x_j^2\partial_j,[x_j^2\partial_j,[\ldots[x_j^2\partial_j, x_j^3\partial_j]\ldots].\end{aligned} \tag{17}$$

Finally, we define

$$\begin{aligned}&x^n y^m z^r t^s\partial_x\\ &=\left(\frac{(n+m+r+s)!}{n!}\right)^{-1}\\ &\times\underbrace{[y\partial_x,[\ldots[y\partial_x}_{m},\underbrace{[z\partial_x,[\ldots[z\partial_x}_{r},\underbrace{[t\partial_x,[\ldots[t\partial_x}_{s},\\ &[x^{n+m+r+s}\partial_x]\ldots]\end{aligned} \tag{18}$$

and similarly for the operators $x^n y^m z^r t^s\partial_j$ $(j=0, 2, 3)$. Our remaining arguments utilize only these definitions and the Jacobi identity.

Equations (14) in the present context consist of the commutator relations of (16) with $x^n y^m z^r t^s\partial_j$. We wish to reduce (14) to (15). Let us consider commutator relations involving first row operators; for example,

$$[y\partial_x, x^n y^m z^r t^s\partial_x]. \tag{19}$$

The value of this commutator is fixed by definition provided $n>0$. Now consider $[y\partial_z, x^n y^m z^r t^s\partial_x]$. By repeated use of the Jacobi identity we find that the value of this commutator is fixed, provided that we are given all commutators of first row elements with each other, and

$$[y\partial_z, x^{n+m+r+s}\partial_x]=0. \tag{20}$$

Consider $[x\partial_y, x^n y^m z^r t^s\partial_x]$; once again using only the Jacobi identity and internal commutators of the first row, we fine that this commutator is fixed if we fix

$$\begin{aligned}&[x\partial_x, x^{n+m+r+s}\partial_x],\\ &[y\partial_y, x^{n+m+r+s}\partial_x],\\ &[x\partial_y, x^{n+m+r+s}\partial_x].\end{aligned} \tag{21}$$

But (21) involve just two coordinates and so are covered by (15e).

Now let us return to (20). By the definition of $x^{n+m+r+s}\partial_x$, we see that (20) is satisfied if

$$[y\partial_z, x^2\partial_x]=0, \tag{20'}$$

$$[y\partial_z, x^3\partial_x]=0. \tag{20''}$$

But (20′) is a special case of (15d), and using the definition of $x^3\partial_x$, (20″) can easily be reduced to several equations included in (15). Finally, consider (19) with $n=0$:

$$[y\partial_x, y^m z^r t^s\partial_x]=0. \tag{19'}$$

Using Eqs. (15b) and (15c), we can write

$$\begin{aligned}&[y\partial_x, y^m z^r t^s\partial_x]\\ &=\frac{m!}{(m+r+s)!}\\ &\times\underbrace{[z\partial_y,[\ldots[z\partial_y}_{r},\underbrace{[t\partial_y,[\ldots[t\partial_y}_{s},[y\partial_x, y^{m+r+s}\partial_x]\ldots].\end{aligned}$$

But $[y\partial_x, y^{m+r+s}\partial_x]=0$ by (15e). So (19′) is satisfied. Commutators like $[y\partial_y, x^n y^m z^r t^s\partial_x]$ quickly reduce to applications of the cases discussed above. Thus we have shown that commutators involving elements of the first row can be reduced to (15).

In order to reduce those cases of (14) which involve commutators of zeroth row elements (viz., $[\partial_j, x^n y^m z^r t^s\partial_k]$), we need only the commutators involving first row elements, plus the internal commutation relations of rows zero and one, and

$$[\partial_j, x_j^2\partial_j]=2x_j\partial_j, \tag{22'}$$

$$[\partial_j, x_k^2\partial_j]=0 \quad k\neq j. \tag{22''}$$

Of course (22) are special cases of (15d).

The only cases of (14) left to be considered are those involving commutators like

$$[x^2\partial_x, x^n y^m z^r t^s\partial_x], \tag{23'}$$

$$[y^2\partial_y, x^n y^m z^r t^s\partial_x]. \tag{23''}$$

Using (15d) we see that $y^2\partial_y$ commutes with $z\partial_x$ and $t\partial_x$. Consequently, we need only be concerned with $[y^2\partial_y, x^{m+r+s}y^m\partial_x]$, which is covered by (15e). This takes care of (23′) and a similar line of reasoning is used for (23″). Thus the theorem is proved.

Remark: With trivial modifications, such as

$$j, k, l, q = 1, 2, 3, \ldots, n;\ n \in Z^+,$$

Theorem II and its proof can be generalized to apply to the ACCTG in n dimensions, $n > 4$.

REPRESENTATIONS OF THE CCTG4

Formanek,[4] Havlicek, and Votruba[5] studied unitary representations of Infinite Lie Groups which contain the inhomogeneous Lorentz group and $SU(n)$. But we know of no research on the unitary representations of the Coordinate Transformation Group. We will briefly describe a few of our results.

First, let us discuss the importance of unitarity. The representation (call it Γ_B) whose infinitesimal generators are given by $B_\gamma^{\nu_1\cdots\nu_n}$, is unitary acting on the functions $f(x^\mu)$. We think of $f(x^\mu)$ as a 4-dimensional probability amplitude, and of $|f|^2$ as a probability density (a density of events), so that

$$f'(x'^\mu) = [J(x'/x)]^{-1/2} f(x^\mu), \tag{2}$$

$$|f'(x^\mu{}')|^2 d^4x' = |f(x^\mu)|^2 d^4x, \tag{2'}$$

$$\int |f'|^2 d^4x' = \int |f|^2 d^4x. \tag{2''}$$

We shall deal with representations which preserve (2) when acting on functions of the coordinates. Such representations are clearly unitary for the (pseudo) group and Hermitian for the algebra.

In Theorem II we have accomplished a substantial simplification of the structure of the ACCTG4. We shall now study some substructures of the algebra with the object of using Theorem II to build up representations.

Consider the one-dimensional ACCTG. Its representatives in the $C_\gamma^{\nu_1\cdots\nu_n}$ representation (call it Γ_0) are

$$p,$$
$$ip\partial_p + i/2,$$
$$p\partial_p\partial_p + \partial_p,$$
$$ip\partial_p\partial_p\partial_p + \tfrac{3}{2}i\partial_p\partial_p, \text{ etc.}$$

Let us say that $p, ip\partial_p + \frac{1}{2}, p\partial_p\partial_p + \partial_p$ represent, respectively, M, N, and Q. These three elements are isomorphic to the algebra of $SO(2, 1)$. Consider the $SO(2, 1)$ representations in which $M \sim p$ (we use "~" to indicate correspondence between an element and its representative); then by satisfying the structure relations of $SO(2, 1)$ we get

$$M \sim p,$$
$$N \sim ip\partial_p + i/2,$$
$$Q \sim p\partial_p\partial_p + \partial_p + cp^{-1};\ c \text{ is a real scalar.}$$

We call this Γ_c. We also find that

$$\text{Casimir operator} = N^2 - iM + MQ = \tfrac{1}{4} + c, \tag{24}$$

so the representations Γ_c are irreducible. Note that as $c \to 0$, we get back Γ_0, so our notation is justified.

We define S_n as the elements of the ACCTG1 which satisfy (in the Γ_0 representation)

$$S_n \sim \begin{cases} p(\partial_p)^n + (n/2)(\partial_p)^{n-1}, & n \text{ even}, \\ ip(\partial_p)^n + i(n/2)(\partial_p)^{n-1}, & n \text{ odd}. \end{cases} \tag{25}$$

Clearly,

$$S_n = [(-1)^n/i(n-3)][Q, S_{n-1}]. \tag{26}$$

By iteration we can define all the S_n in terms of Q and S_3:

$$S_n = \frac{1}{(n-3)!}(-1)^{n(n+1)/2}(i)^{n-3}[\underbrace{Q, [Q, [Q, [\ldots [Q}_{n-3}, S_3]]]]. \tag{27}$$

We wish to enlarge Γ_c to represent all the S_n. Using the method of Theorem II we find that the representatives need only satisfy (27) and

$$[M, S_3] = -3iQ,$$
$$[N, S_3] = -2iS_3, \tag{28}$$
$$[S_n, S_3] = (n-3)iS_{n+2},$$

where M, N, Q are given by Γ_0. Using these equations we can compute an expression for the representative of S_3 in the expanded representation Γ_M

$$S_3 \sim ip(\partial_p)^3 + i\tfrac{3}{2}(\partial_p)^2 + (3ic/p)\partial_p + (d - \tfrac{3}{2}ic)/p^2, \tag{29}$$

where c and d are now Hermitian operators on an expanded representation space. Further calculations with c and d are extremely difficult.

We now consider the two-dimensional ACCTG, which plays a special role in Theorem II. Applying a treatment similar to that used for the ACCTG1, we find two classes of representatives Γ_u and Γ_v for zeroth and first row elements:

Γ_0	Γ_u	Γ_v
p_1	p_1	p_1
p_2	p_2	p_2
$ip_1\partial_{p_1} + i/2$	$ip_1\partial_{p_1} + i/2$	$ip_1\partial_{p_1} + i/2$
$ip_2\partial_{p_2} + i/2$	$ip_2\partial_{p_2} + i\alpha\partial_\alpha + i$	$ip_2\partial_{p_2} + i/2$
$ip_1\partial_{p_2}$	$ip_1\partial_{p_2}$	$ip_1\partial_{p_2} - \beta p_1/p_2$
$ip_2\partial_{p_1}$	$ip_2\partial_{p_1} + \frac{\alpha}{p_1} - \frac{p_2}{p_1} \times (i\alpha\partial_\alpha + i/2)$	$ip_2\partial_{p_1} - \beta p_2/p_1$

where α and β are operators.

We have also obtained Hermitian representations of the zeroth and first rows of the ACCTG4, but have not been able to combine them with Γ_M.

DISCUSSION

Theorem II gives us a method of computing representations of the infinitesimal generators of the coordinate transformations. But why are we interested in these transformations? Most of them do not even ap-

proximately conserve familiar quantum numbers like mass and spin. Recently Jackiw, Fubini, and Hanson[6] have used the dilatation operator $\sum_\mu ix^\mu\partial_\mu$ in a quantization scheme. In their approach "Dilatations replace time translations as dynamical equations of motion." Field operators are quantized on a surface

$$\sum_\mu x_\mu x^\mu = \text{positive const} = \tau^2$$

and then transformed or propagated via the dilatation operator to other surfaces with different values of τ^2. A similar scheme may be envisioned using the conformal generators. With the use of such noninvariance transformations to propagate quantization schemes, one would like to know more about the representations of the pseudogroup of general coordinate transformations. We also believe that we can shed some light on Jackiw's observation that "2-dimensional models are in some way relevant to particle theory of the 4-dimensional physical world."[6] A glance at Theorem II reveals that nearly all the structure of the CCTG4 is implicit in the CCTG2. Hence, it is tempting to conclude that the essential transformation properties of spin zero particles can be encompassed in a two-dimensional theory.

We believe that the elements of the CCTG4 can be implemented physically in such a way that measurable quantities must transform under the action of these elements. For example, a device to measure spin in rectangular coordinates could, in principle, be redesigned along curvilinear lines to measure a transformed "curvilinear" spin. Different representations of the CCTG4 could be distinguished on the basis of the transformation behavior that they prescribe for the spin and other "intrinsic" measurables, as is the case with the conformal group.[7] Then by performing appropriate measurements it ought to be possible to determine which representation corresponds to an observed physical process.

*Present address: Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, Maryland 21205.

[1]I.M. Singer and S. Sternberg, J. d'Analyse Mathematique 15, 1 (1965).
[2]E. Cartan, *Oeuvres Completes*, Partie II (Gauthier-Villars, Paris, 1953), Vol. 2, pp. 567—857, pp. 1335—1384.
[3]C. Freifeld, "One Parameter Subgroups Do Not Fill a Neighborhood of the Identity in an Infinite-Dimensional Lie (Pseudo-) Group," in *Batelle Rencontres* (Benjamin, New York, 1968), pp. 538—543.
[4]J. Formanek, Czech. J. Phys. B 16, 1 (1966).
[5]M. Havlicek and J. Votruba, Czech. J. Phys. B 17, 809 (1967).
[6]S. Fubini, A.J. Hanson, and R. Jackiw, Phys. Rev. D 1, 1732 (1973).
[7]T. Yao, J. Math. Phys. 12, 315 (1971).

Tachyons and gravitation

A. K. Raychaudhuri

Physics Department, Presidency College, Calcutta 12, India
(Received 14 December 1972)

This note presents the following results: (1) In the weak field approximation, the tachyon apparently experiences a repulsion due to ordinary matter; (2) a tachyon turns back from a black hole after penetrating the Schwarzschild singularity; (3) in ever expanding isotropic universes the tachyon velocity becomes arbitrarily large at a finite time; (4) between tachyons themselves, there is apparently an attraction.

There have been some investigations on tachyons in general relativity. While Vaidya[1] has sought to identify an axially symmetric static metric as due to a tachyon, Foster and Ray[2] have considered a continuous distribution having the energy momentum tensor $T_{\mu\nu}=v_\mu v_\nu\rho$ where v^μ is a spacelike vector and ρ is positive. Such a distribution they have called a tachyon dust distribution.

In the present note we seek to investigate the behavior of spacelike geodesics in different gravitational fields so as to obtain an insight into the nature of gravitational interaction between tachyons and ordinary matter as well as between tachyons themselves.

In the weak field approximation one obtains[3]

$$g_{ik}=-(1+2\phi)\,\delta_{ik},\quad g_{00}=1-2\phi \tag{1}$$

so that the geodesic equation gives

$$\frac{dv^i}{ds}+2v^kv^i\phi_{,k}-\phi_{,i}\left[\left(\frac{dt}{ds}\right)^2+\left(\frac{dr}{ds}\right)^2\right]=0. \tag{2}$$

For ordinary material test particles, the condition of low spatial velocities leads to the Newtonian equation with ϕ as the Newtonian potential, but such a condition is not realizable for tachyons and one has a basically non-Newtonian situation with velocity dependent forces.

However, for the case of simple radial motion (i.e., where the world lines are orthogonal to the 2-spaces defined by $\phi=$ constant) one obtains

$$\frac{d^2x^i}{ds^2}=\phi_{,i}\left[\left(\frac{dt}{ds}\right)^2-\left(\frac{dr}{ds}\right)^2\right]$$
$$=\pm\,\phi_{,i}, \tag{3}$$

where the upper and lower signs hold respectively for timelike and spacelike lines and thus the tachyons apparently experience an inverse force of repulsion.

The situation seems interesting for the Schwarzschild field

$$ds^2=(1-2mr^{-1})\,dt^2-(1-2mr^{-1})^{-1}\,dr^2-r^2\,d\Omega^2, \tag{4}$$

and one obtains for spacelike geodesics

$$\left(\frac{dr}{ds}\right)^2=k^2-(1-2mr^{-1})(-1+h^2r^{-2}), \tag{5}$$

where the constants h and k are defined by

$$h=r^2\,\frac{d\phi}{ds},\quad k=\frac{dt}{ds}\,(1-2mr^{-1}). \tag{6}$$

Equation (5) shows that for radial motion (i.e., $h=0$) there exists a minimum of r at $r=2m(k^2+1)^{-1}$. Thus a tachyon moving radially inwards turns back after penetrating inside the Schwarzschild singularity. There also exist circular orbits for tachyon with $r<3m$.

We next consider the case of the isotropic expanding universe with the line element

$$ds^2=dt^2-G^2(t)\,d\sigma^2. \tag{7}$$

The geodesic equation gives, on integration,

$$\left(\frac{dt}{ds}\right)^2\pm1=KG^{-2} \tag{8}$$

where K is a constant and the positive or negative sign is to be used according as the geodesic is space like or time like. Thus if the model be open so that G increases indefinitely, dt/ds would vanish at a finite value of G or in other words the velocity of the tachyons $-g_{ik}(dx^i/dt)(dx^k/dt)$ [as measured by an observer at rest in the coordinate system of (7)] would become arbitrarily large at a finite time. It thus seems difficult to posit the existence of tachyons in open isotropic universes.

Lastly to have an idea about the interaction between the tachyons themselves, let us consider a continuous distribution similar to that of Foster and Ray[2]. Using the identity

$$(v^\mu{}_{;\nu\mu}-v^\mu{}_{;\mu\nu})\,v^\nu=R_{\nu\alpha}v^\nu v^\alpha \tag{9}$$

and the field equations

$$R_{\mu\nu}-\tfrac{1}{2}Rg_{\mu\nu}=-8\pi\rho v_\mu v_\nu, \tag{10}$$

one obtains for the simple case in which the v^μ congruence is hypersurface orthogonal and "shear-free,"[4]

$$\theta_{,\lambda}v^\lambda=-\tfrac{1}{3}\theta^2-\tfrac{4}{3}\pi\rho; \tag{11}$$

irrespective of whether v^μ is timelike or spacelike, θ in the above equation indicates the "expansion" $v^\mu{}_{;\mu}$. We thus see that the expansion of the congruence will be a monotone decreasing function as one proceeds along the world lines (for the tachyons, these are spacelike) indicating a basically attractive interaction.

[1] P. C. Vaidya, Current Science (India) 40, 651 (1971).
[2] J. C. Foster, Jr. and J. R. Ray, J. Math. Phys. 13, 979 (1972).
[3] A. Einstein, *The Meaning of Relativity* (Methuen, London, 1950).
[4] The shear defined as $\sigma_{\mu\nu}=v_{(\mu,\nu)}-\dot{v}_{(\mu}v_{\nu)}-\frac{1}{3}\theta(g_{\mu\nu}-v_\mu v_\nu)$ is orthogonal to v^μ but in case v^μ is spacelike, it is not necessarily a space tenosr. One may define a shear tensor which is in the 2-space orthogonal to v^μ, but this requires an additional timelike vector field. See, in this connection P. J. Greenberg, J. Math. Anal. Appl. 30, 128 (1970).

Regge lattice and Lie-algebra representations

Moshé Flato and Håkan Snellman

Physique mathématique, Collège de France, Paris 5^e, France
(Received 20 September 1972)

A possible description of Regge trajectories in terms of an irreducible representation of the Lie algebra of the 11-parameter Weyl group is proposed.

1. INTRODUCTION

It has been known for many years that the study of representations of Lie algebras is a tough subject, since the family of irreducible representations (irreps.) of a given Lie algebra is *a priori* a much richer object than the irreps. of a corresponding Lie group in a topological vector space.

For the case of the simplest complex simple Lie algebra—which is $sl(2)$—one can find in Ref. 1 a class of algebraically irreps. where one of the elements, the regular element J_3 of the Cartan—Weyl basis of $sl(2)$, has pure denumerable point spectrum Σ with multiplicity one. These are the representations $D(u, m_0)$ with m_0 and $u \in \mathbf{C}$ (u and $-u-1$ giving equivalent reps.), $0 \leq \mathrm{Re}\, m_0 < 1$, $m_0 \pm u \notin \mathbf{Z}$ for which $\Sigma = \{m_0 + n;\ n \in \mathbf{Z}\}$, the reps. $u^{\pm}$ with $u \in \mathbf{C}$, $2u \notin \mathbf{N}$, and $\Sigma = \{\pm(n-u);\ n \in \mathbf{N}\}$, and the usual finite-dimensional reps. $D(j)$, $2j \in \mathbf{N}$, where $\Sigma = \{j, j-1, \ldots, -j\}$.

Miller[1] and recently Arnal and Pinczon[2] have shown that this class of irreps. are up to equivalence all the irreps. where a regular element has an eigenvalue. The latter, however, also found many other algebraically irreps. among which one finds reps. with a nilpotent element in the Cartan—Weyl basis having the eigenvalue 1. Among these reps. we shall use one which is especially interesting since it admits a continuous spectrum for J_3 if realized in a suitable Hilbert space. We propose here to use the irreps. of the Lie algebra $\mathscr{W}$ of the 11-parameter Weyl group W (W being the semidirect product of the real positives $\mathbf{R}_+$ with the Poincaré group) for a description of Regge trajectories. The spin content of these reps. will be given by those of the irreps. of the inducing Lie algebras $su(2)$ or $su(1,1)$ in a manner to be described below.

The Weyl group is known to be the maximal space—time group that preserves causality globally. In contrast to the irreps. of the Weyl group W, which have continuous mass spectrum (with exception of the case $m=0$) we'll construct irreps. of $\mathscr{W}$, which for $m^2 > 0$ have discrete and for $m^2 < 0$ have continuous mass spectrum. When restricted to the Poincaré subalgebra, these reps. decompose into irreps. which for $m^2 > 0$ have the spin content given by one of the irreps. of Miller, with discrete spectrum for J_3, and for $m^2 < 0$ by one of those of Arnal and Pinczon, with continuous spectrum for J_3.

For $m^2 > 0$ any such irrep. of the Poincaré subalgebra will be interpreted as a collection of particles, the spin eigenvalues of which are given by the eigenvalues of J_3. This is in accordance with a description of a collection of particles with the same mass, with spin j and with representatives $|j, m_j\rangle$ ($-j \leq m_j \leq j$) glued together so that the space (e.g., in the semibounded case) contains only the representatives $|j, j\rangle$ for all $j \in \Sigma$ with Σ simple. Correspondingly, for $m^2 < 0$, the states will have continuous spin values. As an example of how such a state space could be realized one can envisage the particles carrying electric charge and magnetic moments. By applying a magnetic field to the particles and subsequent energy analysis (or application of optical pumping techniques) beams can be produced where the particle states have $m_j = j$ relative to a given direction. The ensuing pattern of the states in an irrep. of $\mathscr{W}$ is in accordance with the gross features of a family of Regge trajectories embraced (for $m^2 > 0$ or $m^2 < 0$) in *one* irrep. of $\mathscr{W}$, demonstrating the richness of this type of representation.

The technique for constructing the irreps. is that of Ref. 3 where an example is given of a relativistic symmetry algebra which admits of an irrep. with discrete mass spectrum. Our irrep. of $\mathscr{W}$ gives a mass spectrum linear in the mass. As an alternative to this type of irrep. of $\mathscr{W}$ one can utilize the technique of Ref. 4 [where an irrep. is given of the conformal Lie algebra $su(2,2)$ with discrete mass spectrum linear in $(\text{mass})^2$] to construct irreps. of $\mathscr{W}$ with this type of mass spectrum.

Finally, we mention that a more orthodox spin spectrum (where the states have really $2j+1$ components for every value j) can also be obtained, utilizing a 15-dimensional Lie algebra which is a Wigner—Inonü contraction of the conformal Lie algebra. However, this type of example is rather arbitrary, and is not minimal for our purpose as far as its dimension is concerned.

2. IRREDUCIBLE REPRESENTATIONS OF $\mathscr{W}$

The Lie algebra $\mathscr{W}$ of the 11-parameter Weyl group W is defined by the Poincaré generators P^μ and $M^{\mu\nu}$ which together with the dilatation generator D satisfy the commutation relations of the Poincaré algebra and, in addition,

$$[D, P^\mu] = P^\mu, \tag{2.1}$$

$$[D, M^{\mu\nu}] = 0. \tag{2.2}$$

$\mathscr{W}$ can be considered as a spectral unification of the Poincaré algebra with the two-dimensional solvable Lie algebra. This means that the technique of Ref. 3 is applicable for the construction of some irreps. of $\mathscr{W}$. We distinguish now among the following cases.

A. Case $m^2 > 0$

Let $\mathcal{H}_u$ be the Hilbert space obtained by completing the carrier space V_u of one of the irreps. of $sl(2)$ given by Miller with respect to the topology of l^2 defined by the algebraic basis in V_u.

Consider now a positive mass rep. of the Poincaré algebra in the Hilbert space $H_1 = L^2(\mathbf{R}^3, d^3p/2p_0) \hat{\otimes} \mathcal{H}_u$.

The corresponding generators are denoted by P'^{μ} and $M'^{\mu\nu}$ and the generators of the Lie algebra $su(2)$ of the inducing little group are J_1, J_2, and J_3 acting in $\mathcal{H}_u$. Let $\partial = d/dq$ be the differential operator in $L^2(T, dq)$, where T is the interval $[0, 2\pi a]$. The generators of $\mathcal{W}$ are then defined in $H = H_1 \bar{\otimes} L^2(T, dq)$ by

$$P^{\mu} = P'^{\mu} \bar{\otimes} i\partial, \tag{2.3a}$$

$$M^{\mu\nu} = M'^{\mu\nu} \bar{\otimes} \mathbf{1}_q, \tag{2.3b}$$

$$D = \mathbf{1} \bar{\otimes} [-\tfrac{1}{2}(q\partial + \partial q)], \tag{2.3c}$$

where $\mathbf{1}$ is the identity operator on H_1 and $\mathbf{1}_q$ is the identity operator on $L^2(T, dq)$. The bar here and in the following means operator closure. We now distinguish between two subspaces S_π and S_0 of H. S_π is the domain in H consisting of functions obtained by taking in $L^2(\mathbb{R}^3; d^3p/2p_0)$ functions belonging to $\mathcal{S}(\mathbb{R}^3)$, in $\mathcal{H}_u$ functions consisting of finite sequences in l^2, and in $L^2(T, dq)$ functions that are C^∞ periodic on T. S_0 is the subspace of S_π obtained by taking in $L^2(T, dq)$ functions that are C^∞ and vanish together with all their derivatives at $q = 0$ and $q = 2\pi a$. The commutation relations of $\mathcal{W}$ are satisfied on S_0. P^μ and D are skew-symmetric on S_0. The spectral domain for P^μ is, however, S_π where the eigenvalues of the (mass)^2 operator $-P^2 = -P'^2 \bar{\otimes}(-\partial^2)$ are given by $m^2 = m_0^2 n^2$, $n = 0, \pm 1, \pm 2$, etc., which gives the mass formula $m = m_0 n$, where $m_0 = a^{-1} m_0'$ and where $m_0'^2$ is the eigenvalue of $-P'^2$. For fixed n, the spin content of the corresponding Poincaré algebra rep. is given by that of $\mathcal{H}_u$ (i.e., of V_u) for which we take one of the irreps. of Miller.[1] Several possibilities, of which we mention two, can at this stage be envisaged.

(1) One can take the unbounded irreps. $D(u, m_0)$, in which case we shall interpret the absolute value of the eigenvalue of J_3 as the spin. The doubling of states will be interpreted as a parity doubling coming from the preparation of the space of states, e.g., as described in Ref. 1, by taking into account that both $|j, j\rangle$ and $|j, -j\rangle$ can equally easily be produced by changing the sign of the magnetic field. Thus the irreps. $D(u, 0)$ giving the spins $j = 0$, 1, 2 etc. and $D(u, \frac{1}{2})$ giving the spins $j = \frac{1}{2}$, $\frac{3}{2}$, etc. (with, e.g., $u = \frac{1}{4}$) are suitable for mesons and baryons, respectively.

(2) If we insist upon having semibounded irreps. (to avoid doubling) one interesting difference occurs between baryons and mesons. While for *baryons* the semibounded irrep. $(-\frac{1}{2})^+$ gives the states of spins (in our interpretation) $j = +\frac{1}{2}$, $+\frac{3}{2}$, etc., for *mesons* we have either the irrep. $(-1)^+$ with $j = 1$, 2, 3, etc. in which the fundamental pseudoscalar octet is missing, or, if we want to include the pseudoscalar particles in *one irrep.*, we are obliged to take the irrep. $(0 - \alpha)^+$ with $j = 0 + \alpha$, $1 + \alpha$, etc., where $\alpha \in \mathbb{C}$, $|\alpha| \neq 0$, which, however, can be chosen as small as we want. This difference between baryons and mesons might have to do with the fact that while the baryons contain one stable particle, the proton, no meson is known to be stable. Since none of the mesons is stable the difference explained above might have to do with the yet unknown definition of the spin of an unstable particle.

The negative energy states can be interpreted as antiparticles after applying the usual formalism of second quantization. The unfaithful rep. of the Poincaré algebra that occurs for $m = 0$ ($n = 0$) is interpreted as a type of infinitely degenerate vacuum state or as zero 4-momentum scattering states. The above representation is what we call for Lie algebra representations *topologically irreducible*. This means that it is the closure in the above-defined topology of an *algebraically irreducible* representation in the sense of Ref. 2.

B. Case $m^2 < 0$

For the case $m^2 < 0$ the Lie algebra of the inducing little group is $su(1,1)$. An irrep. of this algebra, which has continuous spectrum for J_3, can be obtained from an irrep. of $sl(2)$ given by Arnal and Pinczon,[2] where the nilpotent element has the eigenvalue $+1$.

The $sl(2)$ Lie algebra in the Cartan–Weyl basis is defined by the commutation relations $[Y, F] = F$, $[Y, G] = -G$, and $[F, G] = 2Y$. The Casimir invariant in this basis is given by $R = FG - Y + Y^2$. Following Ref. 2 an irrep. Π of this algebra in the algebraic space V_u spanned by the basis $\{f_n\}_0^\infty$ with $\Pi(F)f_0 = f_0$ exists such that $\Pi(Y)f_n = f_{n+1}$, $\Pi(F)f_n = [\Pi(Y) - 1]^n f_0$, $\Pi(G)f_n = [\Pi(Y) + 1]^n (rf_0 - f_1 - f_2)$, where $\Pi(R)f_n = rf_n$, $r \in \mathbb{C}$. A simple realization of such an irrep. in $L^2(\mathbb{R}^1, dt)$ with $f_0(t) = \exp(-t^2)$ is given as follows with $f_n(t) = t^n \exp(-t^2)$:

$$\Pi(Y) f_n(t) = t f_n(t),$$

$$\Pi(F)[t^n \exp(-t^2)] = (t-1)^n \exp(-t^2),$$

$$\Pi(G)[t^n \exp(-t^2)] = (t+1)^n (r - t - t^2) \exp(-t^2).$$

In this realization only the operator $\Pi(Y) = J_3$ is formally symmetric. Other realizations [in $L^2(0, \infty)$] exist which are symmetric, but since only $J_3 = \Pi(Y)$ has a direct physical interpretation we shall stick to the above one. The dense invariant domain on which the Lie algebra rep. of $su(1,1)$ is defined is $\mathcal{S}(\mathbb{R}^1)$ on which J_3 is essentially self-adjoint. J_3 admits the continuous spectrum $\Sigma = (-\infty, \infty)$. Of course in this case we shall start with

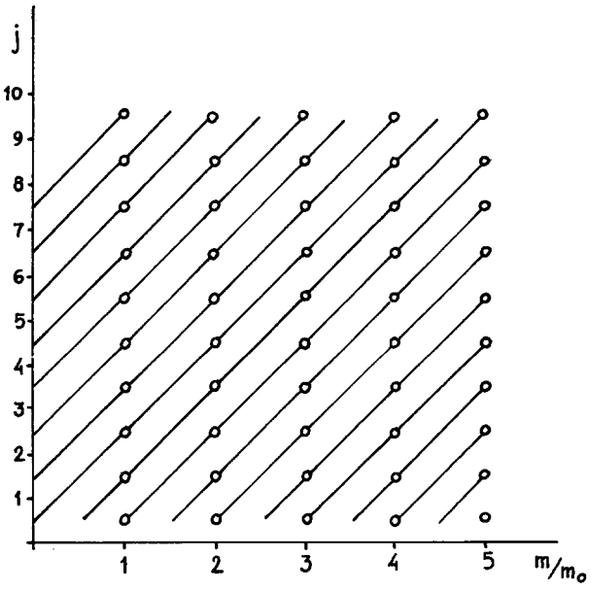


FIG. 1. Regge lattice with trajectories for baryons in an irreducible representation of w for $m^2 > 0$.

a local imaginary mass rep. of Poincaré ($-P'^2<0$) on a space $L^2(\Omega)\hat{\otimes}L^2(\mathbb{R},dt)$, where Ω is a one-sheeted hyperboloid and $L^2(\mathbb{R},dt)$ carries a local irrep. of the Lie algebra $su(1,1)$ of the inducing little group as indicated above. The mass-spectrum generating space in this case will be $L^2(\mathbb{R},dx)$ with the invariant subdomain $\mathcal{S}(\mathbb{R}^1)$ so that the dense invariant subdomain S_0' of the total Hilbert space $H=L^2(\Omega)\hat{\otimes}L^2(\mathbb{R}^1,dt)\hat{\otimes}L^2(\mathbb{R}^1,dx)$ consists of functions in H which for every x and t belong to a Gårding domain in $L^2(\Omega)$ and which are C^∞ and of fast decrease in t and x. The spectrum of $-P^2=-P'^2\bar{\otimes}(-\partial_x^2)$ is continuous and consists of $(-\infty,0)$. We thus have a topologically *irrep.* of $\mathcal{W}$, which has a continuum of negative (mass)^2 states with continuous spin spectrum.

The pattern for case $m^2>0$ is depicted in Fig. 1, which exhibits a Regge lattice with an infinite family of Regge trajectories drawn through the recurrences by respecting the $\Delta j=2$ rule coming from the signature phase factor $\tau=(-1)^{j+B/2}$, where B is the baryon number.

C. Other possibilities

Our example presented above gives a pattern of mass-spin states that is compatible with Regge trajectories that are naturally linear in the mass. Other types of reps. of $\mathcal{W}$ admit other mass dependencies if the two-dimensional solvable Lie algebra is represented, e.g., by means of pseudodifferential operators. Some unsolved mathematical problems in this connection hinder us yet from exhibiting their structure.

Another possibility is to consider instead of $\mathcal{W}$ the conformal Lie algebra $su(2,2)$. An irrep. of $su(2,2)$ that gives rise to a mass-spectrum linear in (mass)^2 has been considered earlier by one of us.[4] If the generators of that example, which are differential operators on an invariant domain $C_0^\infty(Q)$ (where Q is a "cube" in $\mathbb{R}^4$), are supplied with spin generators of the inducing little groups, then a pattern similar to the above one for $\mathcal{W}$ emerges which has a mass-spectrum linear in (mass)^2 in this case. The momenta are discrete, and are to be interpreted as the momenta of the components of a system "quasibound" in the interaction region Q.

This construction can of course be also directly applied to $\mathcal{W}$, giving rise to irreps. of $\mathcal{W}$ that have a mass-spectrum linear in (mass)^2 and a spin spectrum as in cases 2A and 2B above.

Finally, we'll discuss briefly an example of a 15-dimensional Lie algebra g [which is a contraction of the conformal Lie algebra $su(2,2)$] that gives rise to irreps. with discrete mass spectrum and with a spin spectrum where the spin is the conventional one, namely giving rise to $2j+1$ spin states for every irrep. of the Poincaré algebra in the decomposition. The structure of g is given by the Lorentz generators $M^{\mu\nu}$, the two 4-vector generators P^μ and X^μ, and the dilatation operator D, together with the commutation relations

$$[M^{\mu\nu},P^\lambda]=g^{\nu\lambda}P^\mu-g^{\mu\lambda}P^\nu, \tag{2.4a}$$

$$[M^{\mu\nu},X^\lambda]=g^{\nu\lambda}X^\mu-g^{\mu\lambda}X^\nu, \tag{2.4b}$$

$$[X^\mu,P^\lambda]=0, \tag{2.4c}$$

$$[X^\mu,X^\nu]=[P^\mu,P^\nu]=0, \tag{2.4d}$$

$$[D,P^\mu]=P^\mu, \tag{2.4e}$$

$$[D,X^\mu]=X^\mu, \tag{2.4f}$$

$$[D,M^{\mu\nu}]=0, \tag{2.4g}$$

the $M^{\mu\nu}$'s closing to the $so(3,1)$ algebra.

The 14-dimensional subalgebra g' of g, generated by $M^{\mu\nu}$, P^μ, and X^μ, having the Casimir invariants $P_\mu P^\mu$, $X_\mu X^\mu$, and $P_\mu X^\mu$, can be represented irreducibly on the Gårding domain of the representation space H_1 of a unitary irrep. $U(\mathcal{G}')$ of the Lie group $\mathcal{G}'$ given in Ref. 5. If we take $M^{\mu\nu}$ and P^μ to be the generators of the (physical) Poincaré subgroup $\mathcal{P}$, the reduction of a unitary irrep. of $\mathcal{G}'$ on the universal covering $\tilde{\mathcal{P}}$ of $\mathcal{P}$ for $-P^2>0$ as in Ref. 5 gives the decompositions $\bigoplus_{j=0}^{\infty}D(j,m^2)$ or $\bigoplus_{n=0}^{\infty}D(n+\frac{1}{2},m^2)$, where $D(j,m^2)$ are the usual Wigner unitary irreps. of the Poincaré group.

The Lie algebra g can now be represented in the Hilbert space $H=H_1\hat{\otimes}L^2(T,dq)$, where $L^2(T,dq)$ is the same space as in Sec. 2A. The generators in H are given by

$$M^{\mu\nu}=M'^{\mu\nu}\bar{\otimes}\mathbf{1}_q, \tag{2.5a}$$

$$P^\mu=P'^\mu\bar{\otimes}i\partial, \tag{2.5b}$$

$$X^\mu=X'^\mu\bar{\otimes}i\partial, \tag{2.5c}$$

$$D=\mathbf{1}\bar{\otimes}[-\tfrac{1}{2}(q\cdot\partial+\partial\cdot q)], \tag{2.5d}$$

where $M'^{\mu\nu}$, P'^μ, and X'^μ are the generators of g' in H_1.

The commutation relations (2.4a)–(2.4g) as well as the missing ones can be checked on the invariant dense domain S_0'' of H consisting of functions that belong to the Gårding domain of $U(\mathcal{G}')$ in H_1 and are C^∞ on $T=[0,2\pi a]$ and vanish together with all their derivatives at $q=0$ and $q=2\pi a$. The resulting rep. of g is *weakly Schur-irreducible* on S_0''. By this we mean that every bounded operator commuting with the spectral resolution of every generator on a domain where this is skew-adjoint is a multiple of the identity operator.

The mass-spectrum domain is again S_τ'' ($\supset S_0''$) where the functions with respect to q are simply C^∞ periodic on T. The spectrum of $-P^2=-P'^2\bar{\otimes}(-\partial^2)$ is again given by $m^2=n^2m_0^2$, $n=0,\pm1,\pm2$, etc. with $m_0=a^{-1}m_0'$. For each value of m, the spin content of this irrep. of g is the one indicated above.

A similar construction can be carried out for $m^2<0$.

3. DISCUSSION AND CONCLUSIONS

In our opinion one of the most interesting aspects of the above examples is that *one* single irrep. of $\mathcal{W}$ embraces a whole family of Regge trajectories (for $m^2>0$). The importance of this has to do with the fact that on one hand *one irrep.* can be considered as a quantum-mechanical elementary system and on the other hand that the structure of the pattern becomes very rigid. As can easily be seen from Fig. 1, the pattern, for $m^2>0$, has a lattice structure rather than giving emphasis on individual trajectories. Some hints to such a situation from

experimental observations have also been given.[6] For $m^2 < 0$ there is a continuum of spin-mass states. The two regions are mutually independent (since they belong to different irreps. of $\mathscr{W}$), something that also occurs in potential scattering models.

The introduction of representations of Lie algebras into physics asks in principle for an extension of the present rudimentary theory of measurements in quantum mechanics. The scheme of, e.g., von Neumann is at best applicable to a situation of stationary states, and since our Lie-algebra representations are intended to describe unstable particles (except for the proton) there is certainly no problem here with the fact that, e.g., all spin generators are not simultaneously essentially self-adjoint on the domains in question. Furthermore, to avoid misunderstanding, we emphasize that the irreps. presented here (apart from the orthodox example in Sec. 2C) are on much the same footing as the spectrum generating reps. of $so(4.1)$ relative to the H atom.

In confronting the irreps. of $\mathscr{W}$ above with the phenomenological situation as displayed, e.g., by Barger and Cline,[7] the analysis of data (in lack of any theory!) seems to suggest that the trajectories are linear in (mass)^2 with universal slopes, although deviations exist. Some years ago the data as well as the minds of many physicists were in favor of trajectories linear in the mass for baryons. Thus it may be too early yet to draw any definite conclusion. Anyhow, variations of the theme presented here are (as indicated in Sec. 2C) capable of producing mass dependencies of different types, e.g., linear in (mass)^2.

We conclude with three suggestions:

(1) Try to study, by even more examples, the real physical meaning (observables, states, measurement theory, etc.) of Lie-algebra representations.

(2) Try to study pseudodifferential representations of Lie algebras, since, as was explained in Sec. 2C, these might have some advantages with respect to our problem.

(3) Try to find out which additional *dynamical* restrictions really can impose the form of the trajectories on our group theoretical Regge lattice.

ACKNOWLEDGMENTS

The authors wish to thank Jacques Simon and Daniel Sternheimer for clarifying remarks. One of us (M.F.) also thanks Max Dresden for an enlightening discussion. The other one (H.S.) wishes to thank the French Centre National de la Recherche Scientifique for financial support.

[1]W. Miller, Mem. Amer. Math. Soc. No. 50 (1964); and *Lie Theory and Special Functions* (Academic, New York, 1968).
[2]D. Arnal and G. Pinczon, Thèse, Université de Dijon, France (1972); and J. Math. Phys. 15, 350 (1974).
[3]M. Flato and D. Sternheimer, Commun. Math. Phys. 12, 296 (1969).
[4]M. Flato and D. Sternheimer, Phys. Rev. Lett. 16, 1185 (1966).
[5]E. Angelopoulos, Ann. Inst. H. Poincaré 18, 39 (1973).
[6]V. Barger and D. Cline, Phys. Rev. Lett. 20, 298 (1968).
[7]V. Barger and D. Cline, *Phenomenological Theories of High Energy Scattering* (Benjamin, New York, 1969).

Properties of the $(:\varphi^4:)_{1+1}$ interaction Hamiltonian

Basilis Gidas *

Department of Physics, Harrison M. Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan 48104

(Received 1 November 1972)

Using a convergent expansion of the resolvent of the Hamiltonian $H = H_0 + \lambda V$, $V = \int dx \times (x) : \varphi^4 : (x)$, $g(x) \in C_0^\infty$, $g(x) \geq 0$, we give a simple proof of (a) the self-adjointness of the Hamiltonian and (b) the volume independent lower bound of the vacuum energy per unit volume. Also, we obtain some coupling constant analyticity properties of the Hamiltonian, and the limit $(H_0 + \lambda v - z)^{-1} \to (H_0 - z)^{-1}$, $z \in \rho(H_0)$ in norm as $|\lambda| \to 0$ uniformly in $\{\lambda : |\arg\lambda| < \pi\}$.

1. INTRODUCTION

The past few years have seen the birth of a new branch of quantum field theory whose purpose is to prove rigorously the existence of model field theories satisfying certain physical and mathematical requirements (axioms). Glimm and Jaffe[1] have pushed the two space–time dimensional φ^4-interaction [abbreviated $(:\varphi^4:)_{1+1}$] to almost a theory which is known to satisfy all the Haag–Kastler axioms and many of the Wightman axioms. A basic step in the construction of the field theory is to prove the boundedness below and the self-adjointness of the Hamiltonian operator

$$H(\lambda) = H_0 + \lambda V(g) = H_0 + \lambda \int dx\, g(x) : \varphi^4 : (x), \tag{1.1}$$

where H_0 is the free Hamiltonian, λ the coupling constant, and $g(x) \in C_0^\infty(R)$ is real value. The first proof of the semiboundedness in a finite volume with periodic boundary conditions was given in the pioneer work of Nelson.[2] His method was extended by Glimm[3] who replaced the periodic box by a fixed $g(.)$ space cut-off. Later, Glimm and Jaffe[4] obtained a lower bound of the Hamiltonian proportional to the volume (i.e., a volume independent bound of the vacuum energy per unit volume). The first proof of the self-adjointness of the Hamiltonian (1.1) was given by Glimm and Jaffe.[5] Segal[6] simplified the proof of self-adjointness and developed powerful techniques which were elaborated further and systematized by Simon and Hoegh-Krohn.[7] Recently, Federbush[8] considered a convergent expansion of the resolvent for the Hamiltonian (1.1) and obtained easily the semiboundedness of the Hamiltonian.

On the other hand, there has been another trend in the rigorous study of model field theories. The second trend involves the study of coupling constant analyticity of various objects associated with the theory. It also, examines the asymptoticity of the perturbation series of quantities such as the ground state, the ground state energy, and equal time vacuum expectation values. Often, the exact objects can be recovered from the asymptotic series by proper summability methods such as Padé[12,13] or Borel.[14,15]

In this paper we use Federbush's expansion of the resolvent to give a simple proof of the self-adjointness of the Hamiltonian, and of the volume independent lower bound of the vacuum energy per unit volume. Also, we study the Hamiltonian (1.1) for complex values of the coupling constant λ. We prove that Federbush's expansion of the resolvent is uniformly convergent for λ in $\{\lambda : |\arg\lambda| < \pi\}$. This yields full cut plane analyticity of the resolvent, and analyticity of the ground state energy and vacuum vector in a neighborhood of the positive real axis. In addition, it implies that if $z \in \rho(H_0)$, $\rho(H_0)$ the resolvent set of H_0, then

$$(H_0 + \lambda V(g) - Z^{-1} \to (H_0 - z)^{-1} \tag{1.2}$$

in norm as $|\lambda| \to 0$ uniformly in $|\arg\lambda| < \pi$. (1.2) is important in the study of the asymptotic nature and the Borel summability of the Rayleigh–Schrödinger perturbation series for the ground state and the ground state energy. A corollary of (1.2 is the uniqueness of the ground state for small values of the coupling constant (the uniqueness of the ground state for any values of the coupling constant is a more difficult problem[4,7]). In a forthcoming paper,[16] using the methods of this paper, we give a simple proof of the self-adjointness of a local Lorentz generator formally given by

$$M = M_0 + M_1(g) = \tfrac{1}{2} \int :\{\pi^2(x) + \nabla\varphi(x)^2 + \mu_0^2\varphi^2(x)\} : x\,dx + V(xg). \tag{1.3}$$

The organization of this paper is as follows: In Sec. II, we summarize the most important ingredients of Federbush's expansion and prove the self-adjointness of the Hamiltonian. In Sec. III, we prove the volume independent lower bound of the vacuum energy per unit volume, and in Sec. IV, we prove the uniform convergence of Federbush's expansion for values of the coupling constant in a certain complex domain.

2. A CONVERGENT EXPANSION FOR THE RESOLVENT AND SELF-ADJOINTNESS OF THE HAMILTONIAN

In this section we summarize the main points of Federbush's paper[8] and prove the self-adjointness of the Hamiltonian. Following Ref. 8, we consider

$$H_\kappa = H_0 + V_\kappa(g) = H_0 + \lambda \int dx g(x) : \varphi_\kappa^4 : (x), \tag{2.1}$$

where $g \in C_0^\infty(R)$, $0 \leq g(x) \leq 1$, H_0 is the free Hamiltonian, and φ_κ is the boson field with a momentum cut-off κ. Let P_i be the projection operator onto states with number of particles in the interval $(2^i, 2^{i+2})$, $i = -1, 0, 1, 2, \ldots$, and P_e and P_d the projection operators onto states with number of particles in the ranges

$$\bigcup_{i=\text{even}} (2^i - 4 \leq N \leq 2^i + 4), \tag{2.2a}$$

$$\bigcup_{i=\text{odd}} (2^i - 4 \leq N \leq 2^i + 4), \tag{2.2b}$$

respectively. We define

$$H_i = P_i H_\kappa P_i = P_i H_0 P_i + P_i V_\kappa P_i = H_{0,i} + V_{\kappa,i}, \tag{2.3}$$

$$H_e = \sum_{i=\text{even}} H_i = H_0 + \sum_{i=\text{even}} V_{\kappa,i}, \tag{2.4}$$

$$H_d=\sum_{i=\text{odd}} H_i=H_0+\sum_{i=\text{odd}} V_{\kappa,i}, \tag{2.5}$$

$$W_e=V_\kappa-\sum_{i=\text{even}} V_{\kappa,i}, \tag{2.6}$$

$$W_d=V_\kappa-\sum_{i=\text{odd}} V_{\kappa,i}. \tag{2.7}$$

Then

$$H=H_e+W_e=H_d+W_d. \tag{2.8}$$

Federbush considers the expansion

$$\begin{aligned}R_\kappa(b;H_\kappa)&=R_\kappa(b;H_e)\\&\quad-R_\kappa(b;H_d)W_eR_\kappa(b;H_e)\\&\quad+R_\kappa(b;H_e)W_dR_\kappa(b;H_d)W_eR_\kappa(b;H_e)\\&\quad-\cdots\cdots,\end{aligned} \tag{2.9a}$$

$$\begin{aligned}&=R_\kappa(b;H_e)\\&\quad-R_\kappa(b;H_d)P_eW_eP_eR_\kappa(b;H_e)\\&\quad+R_\kappa(b;H_e)P_dW_dP_dR_\kappa(b;H_d)P_eW_eP_eR_\kappa(b;H_e)\\&\quad-\cdots,\end{aligned} \tag{2.9b}$$

where b is a large positive number, and $R(Z)=(Z+A)^{-1}$ denotes the resolvent of operator A. The results of Ref. 8 can be summarized in Theorem (2.1).

Theorem 2.1: There exists a finite constant α, independent of κ, such that for $b>\alpha$, expansion (2.9) converges in the uniform operator topology and is continuous in κ for $0\leq\kappa\leq+\infty$. $R_\infty(b)$ is the resolvent of an operator $H=H_\infty(g)$ such that $H>-\alpha$.

The basic estimates which yield theorem (2.1) are

Estimate 1:

$$H_i\geq 2^{i-1}P_i, \quad \text{for large } i; \tag{2.10}$$

Estimate 2:

$$\|P_eP_iR(b;H_i)P_d\|\leq c_1\exp(-c_2 2^{i/2}) \tag{2.11}$$

for large i, and for some $c_1, c_2>0$.

To obtain Estimate 1, we choose an increasing sequence of momentum cut-offs and write

$$H_i=P_iH_0P_i+P_iV_\kappa P_i+P_i(V_\kappa-V_{\kappa_i})P_i. \tag{2.12}$$

Clearly, one has

$$P_iH_0P_i\geq m2^iP_i. \tag{2.13}$$

By undoing the Wick ordering of V_{κ_i}, we obtain the momentum cut-off dependent bound

$$V_{\kappa_i}\geq-\text{const}(\ln\kappa_i)^2. \tag{2.14}$$

By a standard N_τ estimate[1,3]

$$\|(N+I)^{-1}(V_\kappa-V_{\kappa_i})(N+I)^{-1}\|\leq 0(\kappa_i^{-1/2}). \tag{2.15}$$

Estimate 1 is obtained from (2.13–2.15) by choosing

$$\kappa_i=\exp[(M/C)^{1/2}2^{(i-1)/2}],$$

where c is the constant in (2.13).

The proof of Estimate 2 is based on the following theorem:

Theorem 2.2: Let A be a positive self-adjoint operator of norm less or equal to M, and $|\alpha\rangle$ and $|\beta\rangle$ two vectors of unit length. Suppose

$$\langle\alpha|A^k|\beta\rangle=0, \text{ for } 0\leq k\leq N. \tag{2.16}$$

Then, for any $\mu>0$ a real number,

$$|\langle\alpha|R(+\mu;A)|\beta\rangle|\leq\frac{4}{\mu\sqrt{2\mu/M}}(1+\sqrt{2\mu/M})^{-N}. \tag{2.17}$$

Federbush applies Theorem 2.2 to the operator $A_i=b+P_i(N+V_\kappa(g))P_i-2^{i-1}$ which satisfies

$$A_i\geq 0, \tag{2.18}$$

$$\|b+P_i(N+V_\kappa(g))P_i\|\leq d2^{2i}. \tag{2.19}$$

Estimate 2 is obtained by using $N\leq$ const H_0, and choosing $|\alpha_i\rangle=P_iP_d|\alpha\rangle$, $|\beta\rangle=P_iP_e|b\rangle$ ($|a\rangle$ and $|b\rangle$ normalized vectors), $\mu_i=2^{i-1}$, $M_i=d2^{2i}$, and $N_i<[(2^{i+1}-4)-(2^i+4)]/4$, and i large enough.

Our main result in this section is Theorem 2.3.

Theorem 2.3: The operator H defined in Theorem 2.1 is self-adjoint. The proof of Theorem 2.3 follows from the following two lemmas.

Lemma 2.4: Let $\epsilon>0$ be sufficiently small. Then there exists a constant $c(\epsilon)$ such that

$$\|R(b;H_e)\|, \ \|R(b;H_d)\|\leq 1/b, \tag{2.20}$$

$$\|W_eP_dR(b;H_d)P_e\|, \ \|W_eP_eR(b;H_e)P_d\|<\tfrac{1}{2}, \tag{2.21}$$

$$\|R(b;H_d)W_eR(b;H_e)\|\leq c(\epsilon)/b^{1+\epsilon}, \tag{2.22}$$

$$\|W_dR(b;H_d)W_eR(b;H_e)\|\leq c(\epsilon)/b^{1+\epsilon}. \tag{2.23}$$

Proof: Inequality (2.20) is easily obtained from Estimate 1. Let $|a\rangle$ and $|b\rangle$ be two normalized states in the Fock space. Then, to obtain Estimate (2.21), we consider

$$\begin{aligned}&\langle a|W_dP_dR(b;H_d)P_e+b\rangle\\&\quad=\sum_{i=\text{odd}}\langle a|W_dP_dR(b;H_d)P_eP_i+b\rangle\\&\quad=\sum_{i=\text{odd}}\langle a|W_dP_dR(b;H_i)P_e|b\rangle\\&\quad=\sum_{i=\text{odd}}\langle a|(I-P_i)V_\kappa P_iP_dP_iR(b;H_i)P_e|b\rangle,\\&\quad\leq\sum_{i=\text{odd}}c_1 2^{2i}\exp(-c_2 2^{i/2})<\tfrac{1}{2}.\end{aligned} \tag{2.24}$$

In the last step above we have used Estimate 2 and a standard N_t estimate. Similar arguments establish inequalities (2.22) and (2.23).

Remark: Estimates (2.20) through (2.23), without the b-independence of the bounds, were also used in Ref. 8.

Lemma 2.5: For b large enough, the series (2.9) converges in the uniform operator topology to an operator $R_\kappa(b)$ which is a continuous in κ pseudoresolvent and satisfies

$$\text{norm-}\lim_{b\to+\infty} bR_\kappa(b)=I, \tag{2.25}$$

$$R_\kappa(b):\mathcal{H}\to\mathcal{H} \text{ is injective}, \tag{2.26}$$

where $\mathcal{H}$ is the Fock-Hilbert space.

Proof: Estimates (2.20)–(2.23) imply that the nth term in the expansion (2.9) is bounded by $c(\epsilon)^nb^{-1-n\epsilon}$. Therefore, the series (2.9) converges, in the uniform operator topology, for $b>(\epsilon)^{1/\epsilon}$, for $0\leq\kappa\leq+\infty$. The continuity in κ is obtained from the uniform in κ esti-

matès (2.20)—(2.23). The bound $c(\epsilon)^n/b^{1+n\epsilon}$ of the nth order term yields

$$\lim_{b\to+\infty} b\{\|R(b;H_e)\| + \|R(b;H_d)W_eR(b;H_e)\| + \cdots\}$$
$$= \lim_{b\to+\infty} b\|R(b;H_e)\| + \lim_{b\to+\infty} b\{\|R(b;H_d)W_eR(b;H_e)\| + \cdots\}. \tag{2.27}$$

The second term above goes to zero as $b\to+\infty$ while the first term goes to one. It is easily shown that $R_\kappa(b)$ is a pseudoresolvent, and satisfies (2.26).

Proof of Theorem 2.3 *and of Theorem* 2.1: Let $\mathcal{N}(b)$ be the null space of $R_\infty(b)$, i.e.,

$$\mathcal{N}(b) = \{\Phi \in \mathcal{H} : R_\infty(b)\Phi = 0\}. \tag{2.28}$$

Since $R_\infty(b)$ is a pseudoresolvent, $\mathcal{N}(b)$ is independent of b, and, by (2.25), $\mathcal{N}(b) = \{0\}$. We define

$$H(b) = -b + R_\infty(b)^{-1} \tag{2.29}$$

and domain

$$D(H(b)) = R_\infty(b)\mathcal{H}. \tag{2.30}$$

Let $\Phi \in \mathcal{H}$ be orthogonal to $R_\infty(b)\mathcal{H}$. Then,

$$(\Phi, R_\infty(b)\Psi) = (R_\infty(b)\Phi, \Psi) \tag{2.31}$$

for all $\Psi \in \mathcal{H}$. Since $\mathcal{N}(b) = \{0\}$, we get $\Phi = 0$, and, therefore, $D(H(b))$ is dense in $\mathcal{H}$. The pseudoresolvent property of $R_\infty(b)$ implies that $D(H(b))$ is independent of b. Therefore, for large b, $H = H(b)$ is bounded below and independent of b. The self-adjointness of this operator follows from the following lemma.[17]

Lemma 2.6: If T is an operator on the Hilbert space $\mathcal{H}$, and if T^{-1} exists and has dense domain, then $(T^*)^{-1} = (T^{-1})^*$.

3. VACUUM ENERGY PER UNIT VOLUME

The Hamiltonian $H(g) = H_0 + V(g)$ has a unique ground state $\Omega(g)$ with eigenvalue $E(g)$. According to the perturbation theory, $E(g)$ is proportional to the volume of space in which the particles interact in each order of the Rayleigh—Schrödinger (≡ Feynman!) perturbation series. However, the perturbation expansion for $E(g)$ diverges.[18] Thus, we cannot conclude from perturbation theory that $E(g)$ is proportional to the volume. In this section we prove rigorously that the prediction of the perturbation theory is correct. Our main tool in the proof is the localization indices introduced by Glimm and Jaffe in a similar context.[19] We consider the Hamiltonian $H(g) = N + V(g)$.

Theorem 3.*d*: Let $g(x) \in C_0^\infty(R)$ have the following properties:

(i) $0 \le g(x) \le 1$; (3.1)

(ii) for some constant $\alpha > 0$,

$$\left|\frac{dg(x)}{dx}\right| \le \alpha. \tag{3.2}$$

Set $m(g) =$ measure (supp. g). There exist constants $a > 0$ and $c > 0$, independent of g, such that

$$0 \le c(H(g) + am(g)). \tag{3.3}$$

Remark 1: Inequality (3.3) implies that there exists $c_1 > 0$ independent of g, such that

$$-c_1 m(g) \le E(g) \le 0, \tag{3.4}$$

that $E(g) \le 0$ is trivial.

Remark 2: Estimate (3.4) is the main technical step in proving the *locally Fock* property[19] of the representation (of the algebra of local observables) associated with the Hilbert space of the physical states obtained in the limit $g \to 1$.

Proof of Theorem 3.1: Let $\eta(x) \in C_0^\infty(R)$, $0 \le \eta(x) \le 1$, be such that the translates

$$\eta_j(x) = \eta(x - j), \quad j \in Z \tag{3.5}$$

define a partition of unity:

$$\sum_{j\in Z} \eta_j(x) = 1 \quad \text{for all } X \in R. \tag{3.6}$$

Let $I(g) = \{j \in Z : \text{supp} g \cap \text{supp}\eta_j(x) \neq \}$ and $|I(g)|$ denote the number of elements in $I(g)$. We decompose g and $V(g)$ into a sum of local parts

$$g = \sum_{j\in I(g)} \eta_j g, \tag{3.7}$$

$$V(g) = \sum_{j\in I(g)} V(\eta_j g), \tag{3.8a}$$

$$= \sum_{j\in I(g)} \int dk_1 \cdots dk_4 \sum_{\alpha=0}^{4} \binom{4}{\alpha} \hat{b}_j(k_1 \cdots k_4) a^*_{k_1} \cdots a^*_{k_\alpha} \times a_{-k_{\alpha+1}} \cdots a_{-k_\alpha} \tag{3.8b}$$

where

$$\hat{b}_j(k_1 \cdots k_4) = [1/(2(2\pi))^2]\widehat{(\eta_j g)}(k_1 + \cdots + k_4) \prod_{l=1}^{4} \mu_{k_l}^{-1/2}, \tag{3.9}$$

$$\mu_{k_l}^2 = k_l^2 + \mu_0^2. \tag{3.10}$$

Instead of using g, we use a simplified space cutoff g_n defined by

$$g_n(x) = \sum_{|j|\le n} \eta_j(x), \quad n \in Z^+. \tag{3.11}$$

Instead of (3.3), we will prove

$$0 \le H(g_n) + 0(n) = N + V(g_n) + 0(1). \tag{3.12}$$

We prove (3.12) for each translate separately, i.e., we prove

$$0 \le N_{\text{loc}} + V(\eta_j) + 0(1). \tag{3.13}$$

N_{loc} is a local number operator to be defined below, and then summing over all translates we obtain (3.12) and hence (3.3). To prove (3.13) we introduce localization indices in configuration space (localization indices in momentum space[4] could be used as well). First, we introduce local N_τ operators.

Let $\mu_k = (k^2 + \mu_0^2)^{1/2}$ denote the one particle energy. Let a_k^* and a_k be creation and annihilation operators in momentum space, and $A^\#(x) = (2\pi)^{-1/2}\int dk\, e^{\pm ikx} A_k^\#$ annihilation and creation operators in configuration space. If $\mu_k^\tau \in O_m(R)$, $0 \le \tau \le 1$, is considered as a multiplication operator on $S(R^1)$, then the configuration space operation μ_x^τ corresponding to μ_k^τ is convolution by a kernel $k_\tau(x) \in O_m'(R)$ (for notation see Ref. 20). $k_\tau(.)$ decreases

exponentially at infinity. Explicitly,[21]

$$k_\tau(x)=\frac{2^{\tau/2+1}}{\Gamma(-\tau/2)}\left(\frac{\mu_0}{|x|}\right)^{\tau+1/2}\int_0^\infty \exp(-\mu_0|x|\cosh t)\times\cosh t\left(\frac{1+\tau}{2}\right)dt. \tag{3.14}$$

For $\tau\geq -1$

$$\left|\frac{d^n}{dx^n}k_\tau(x)\right|\leq O[\exp(-\mu_0|x|)],\quad \text{as } |x|\to+\infty,\quad n=0,1,2,\cdots. \tag{3.15}$$

For $\tau<-1$, (3.15) holds if μ_0 is replaced by $\mu_0-\epsilon$, for any $\epsilon>0$. For $\delta>0$

$$|k_\tau(x)|\leq C(\delta)|x|^{-1-\delta}\quad \text{for all } x. \tag{3.16}$$

For $\zeta\in O_M(R)$, nonnegative real, we define the local number operator

$$N_{\tau,\zeta}=\int dxdy\,A^*(x)\zeta(x)k_\tau(x-y)\zeta(y)A(y), \tag{3.17a}$$

$$=\int dk_1dk_2\,a^*_{k_1}\Big(\int dk\,\hat\zeta(k_1-k)\mu_k^\tau\hat\zeta(k-k_2)\Big)a_{k_2}. \tag{3.17b}$$

Strictly speaking, $N_{\tau,\zeta}$ is the Friedrichs extension of the positive operator defined by the right hand side of (3.17) on D_0, the set of vectors in Fock space with a finite number of particles and wavefunctions in ζ. If $\tau=1$, then $N_{1,\zeta}$ provides a local energy operator H_0^{loc}. For $0\leq\tau<\frac{1}{4}$, one can replace ζ by the characteristic function $E_B(x)$ of an interval B, to obtain a sharply localized operator $N_{\tau,B}$. N_{loc} in (3.13) is of this type. Using techniques of Fourier analysis, Glimm and Jaffe[19] proved that

$$E_B(x)\mu_x^\tau E_B(x)\leq \text{const}\,\zeta(x)\mu_x^{2\tau+\epsilon}\zeta(x), \tag{3.18}$$

if $\tau<\frac{1}{2}$, and if $\zeta(x)\equiv 1$ on a neighborhood of B^-. Furthermore, if $\zeta_j(x)=\zeta(x-j)$, then

$$\sum_{j\in Z}\zeta_j(x)\mu_x^\tau\zeta_j(x)\leq \text{const}\,\mu_x^\tau. \tag{3.19}$$

These inequalities on the single particle space sym $L_2(R)$ lead to the estimate

$$\sum_{j\in Z}N_{\tau,B+j}\leq \text{const}N_\tau\leq \text{const}H_0 \tag{3.20}$$

of the sum of local number operators by (global) N_τ operators. Estimate (3.20) holds as an inequality between positive self-adjoint operators in Fock space. It shows that in order to prove (3.12) [or (3.3)] it is enough to prove (3.13), or equivalently

$$0\leq N_{\text{loc}}+H_I(g)+O(1), \tag{3.21}$$

where N_{loc} is a localized number operator, and g is supported in a fixed interval B.

We prove (3.21) by the method of Sec. II. Our discussion in that section shows that we only need to prove (2.15), i.e.,

$$\|(N_{\text{loc}}+I)^{-1}(V_\kappa-V_{\kappa i})(N_{\text{loc}}+I)^{-1}\|\leq O(\kappa_i^{-1/2}). \tag{3.22}$$

Of course, one has to check also (2.25), but this is not hard. To prove (3.22), we note that $V'_{\kappa_i}=V_\kappa-V_{\kappa_i}$ is a sum of five monomials in creation and annihilation operators

$$V'_{\kappa_i}(g)=\int dk_1\cdots dk_4\hat b'_{\kappa_i}(k_i\cdots k_4)\sum_{\alpha=0}^4\binom{4}{\alpha}a^*_{k_1}\cdots a^*_{k_\alpha}\times a_{-k_{\alpha+1}}\cdots a_{-k_4}, \tag{3.23}$$

$$=\int dx_1\cdots dx_4 b'_{\kappa_i}(x_i\cdots x_4)\sum_{\alpha=0}^4\binom{4}{\alpha}A^*(x_1)\cdots A^*(x_\alpha)A(-x_{\alpha+1})\cdots A(-x_4), \tag{3.24}$$

where

$$\hat b'_{\kappa_i}(k_1,\ldots,k_4)=\hat b_\kappa(k_1\cdots k_4)-\hat b_{\kappa_i}(k_1\cdots k_4) \tag{3.25}$$

and $b'_{\kappa_i}(x_1,\ldots,x_4)$ is the Fourier transform of $\hat b'_{\kappa_i}(k_1,\ldots,k_4)$.

Without loss of generality, we consider the case where B is an open interval whose closure is contained in $(0,1)$. Let

$$X_j=(j,j+1) \tag{3.26}$$

and N_{X_j} the corresponding local number operators, $j=0,\pm1,\pm2,\ldots$, which are commuting self-adjoint operators. We define N_{loc} by

$$N_{\text{loc}}=\sum_{j=-\infty}^{+\infty}N_{X_j}\exp[-\mu_0(|j|/2)] \tag{3.27}$$

and localize the operator $V'_{\kappa_i}(g)$ as follows. Define

$$b'_{\kappa_i j_1\cdots j_4}(x_1,\ldots,x_4)=b'_{\kappa_i}(x_1\cdots x_4)\prod_{l=1}^4E_{j_l}(x_l), \tag{3.28}$$

where $E_{j_l}(x)$ is the characteristic function of X_j. Equation (3.28) localizes each coordinate X_l in the interval X_j. Using (3.28) we will prove that $b'_{\kappa_i}(x_1,\ldots,x_4)$ is exponentially small at infinity in each of the variables X_l. We prove

Lemma 3.1: $b'_{\kappa_i;j_1\ldots j_4}(x_1,\ldots x_4)$ is a bounded operator on $L_2(R^4)$ satisfying

$$\|b'_{\kappa_i;j_1\cdots j_4}\|_2\leq\|b'_{\kappa_i}\|_2\exp(-\mu_0|j_1|\cdots-\mu_0|j_4|), \tag{3.29a}$$

$$\leq O(\kappa_i^{-1/2})\exp(-\mu_0|j_1|\cdots-\mu_0|j_4|). \tag{3.29b}$$

Proof: The transition from (3.29a) to (3.29b) has been established by Glimm[3] (see also Ref. 1). Thus, we need only prove (3.29a). The crux of the proof is a representation of $b'_{\kappa_i;j_1\cdots j_4}(x_1,\ldots,x_4)$ developed by Glimm and Jaffe (Ref. 19, pp. 84–95). Let $\zeta(x)$ be a $C_0^\infty(R)$ with support in χ_0 and which equals one on a neighborhood of B. We define operators

$$K_l=\mu_{x_l}^{-1/2}\zeta(x_l)\mu_{x_l}^{1/2}\quad l=1,\ldots,4, \tag{3.30}$$

where kernels $k(x_l z_l)$ are tempered distributions. Estimate (3.15) of k_τ implies that $k(x_l,z_l)$ is a C^∞ function of x_l and z_l such that

$$|k(x_l,z_l)|\leq O[\exp(-\mu_0|x_l|-\mu_0|Z_l|)]. \tag{3.31}$$

We now define localized operators K_{l,i_l} by

$$K_{l,jl}=E_{jl}(x_l)\mu_{x_l}^{-1/2}\zeta(x_l)\mu_{x_l}^{1/2},\quad l=1,\ldots,4, \tag{3.32}$$

whose kernels $k_{i_l}(x_l,z_l)$ satisfy

$$k_{j_l}(x_l,z_l)=\begin{cases}k(x_l,z_l), x_l\in X_{j_l}\\ 0, \text{ otherwise}\end{cases}. \tag{3.33}$$

Using (3.31), one can prove that for $j_l\neq 0$ the operator

norm of K_{l,j_l} is bounded by

$$\|K_{l,j_l}\| \leq O[\exp(-\mu_0|j_l|)]. \tag{3.34}$$

It is not hard to show that (3.34) remains true for j_l, even for $j_l=0$. The representation of $b'_{\kappa_i;j_1\cdots j_4}$ that we are after is

$$b'_{\kappa_i;j_1\cdots j_4}=(\hat{\prod_{l=1}}K_{l,j_l})b'_{\kappa_i}. \tag{3.35}$$

Representation (3.35) together with estimate (3.34) proves *Lemma* 3.1.

We now return to the proof of (3.22). From (3.24) we get for $\Phi \in D_0$.

$$\|V'_{\kappa_i}(g)\| \leq \sum_{j_1\cdots j_4} \|V'_{\kappa_i;j_1\cdots j_4}\Phi\| \tag{3.36a}$$

$$\leq \text{const} \sum_{j_1\cdots j_4} \|b'_{\kappa_i;j_1\cdots j_4}\|_2 \,\|\hat{\prod_{l=1}}(N_{X_{j_l}}+I)^{1/2}\Phi\| \tag{3.36b}$$

$$\leq 0(\kappa_i^{-1/2}) \sum_{j_1\cdots j_4} \exp(-\mu_0|j_1|\cdots\mu_0|j_4|) \times \|\prod_{l=1}^{4}(N_{X_{j_l}}+I)^{1/2}\Phi\| \tag{3.36c}$$

$$\leq 0(\kappa_i^{-1/2}) \sum_{j_1\cdots j_4} \exp(-\mu_0|j_1|\cdots\mu_0|j_4|) \times \exp\left(\mu_0\frac{j_1}{2}+\cdots+\mu_0\left|\frac{j_4}{2}\right|\right)\times\|(N_{\text{loc}}+I)^2\Phi\| \tag{3.36d}$$

$$\leq 0(\kappa_i^{-1/2})\|(N_{\text{loc}}+I)^2\Phi\|. \tag{3.36e}$$

Inequality (3.36b) is an elementary local N_τ estimate. In (3.36c) we used (3.29b). In (3.36d–e) we used the estimate $\|(N_{X_j}+I)^{1/2}(N_{\text{loc}}+I)^{-1/2}\| \leq \exp(\mu_0|j/2|)$.

Estimate (3.36) implies (3.22). QED

4. COUPLING CONSTANT ANALYTICITY

In this section we study $H_\lambda = H_0 + \lambda V(g)$ for λ in the complex λ plane cut along the negative real axis. We prove that Federbush's expansion for the resolvent is valid for values of λ in the domain

$$\{\lambda: |\arg\lambda| \leq \pi-\epsilon\}, \tag{4.1}$$

where $\epsilon>0$. The basic theorem used in the proof is a generalization of Theorem 2.2 for bounded sectorial operators[22].

Theorem 4.1: Let A be a bounded sectorial operator of norm less or equal to M, i.e., the numerical range $\theta(A)$ of A is the subset of the sector

$$|\arg(z-\gamma)| \leq \theta, \quad 0 \leq \theta < \pi/2, \quad \gamma \text{real} \tag{4.2}$$

lying in a circle with center γ and radius M. Let $|\alpha\rangle$ and $|\beta\rangle$ be two vectors of unit length. Suppose

$$\langle\alpha|A^k|\beta\rangle = 0, \quad \text{for } 0 \leq k \leq N, \tag{4.3}$$

then for any $\mu>0$ a real number:

$$|\langle\alpha|R(+\mu;A)|\beta\rangle| \leq 0[\exp(-N\sqrt{2\mu/M})], \quad N \text{ and } M \text{ large}. \tag{4.4}$$

The proof of this theorem is similar to the proof of Theorem 2.2, see Ref. 8. It is based on a theorem by Bernstein (Ref. 23, p. 84, and Ref. 24, pp. 130–31 and pp. 280–81) in the theory of the best approximation of analytic functions, stating that if $f(z)$ is defined in $\{Z: |\arg z| < \pi/2, |z| \leq 1\} = D$ and it is analytic in an ellipse with foci at -1 and 1 and with sum of its semiaxes equal to $r>1$, then $f(z)$ may be approximated on D by a Fourier–Chebyshev polynomial series of degree n within

$$(2f_{\max}/r-1)(1/r^n) \tag{4.5}$$

in the uniform norm. $f_{\max}$ is the supremum of the absolute value of $f(z)$ in the ellipse.

Let

$$\begin{aligned} A_\lambda &= b + H_i(\lambda) - 2^{i-1} \\ &= b + P_i H_0 P_i + \lambda P_i V P_i - 2^{i-1} \\ &= b + H_{0,i} + \lambda V_i - 2^{i-1}. \end{aligned} \tag{4.6}$$

By realizing V as a multiplication operator on the probability space $L_2(Q,dq)$, we can write $V = V^+ - V^-$ with $V^+V^- = 0$ (Refs. 1,7). Using (2.24) and

$$V^- \leq \text{const}\, N + \text{const} \tag{4.7}$$

implied by Theorem (2.1), we see that A_λ is a sectorial operator, i.e., there exists a γ such that

$$|\text{Im}(\Phi, A_\lambda\Phi)| \leq \tan(\arg\lambda/2)\{\text{Re}(\Phi,(A_\lambda-\gamma)\Phi\} \tag{4.8}$$

for λ such that $|\text{avg}\lambda| \leq \pi-\epsilon, \epsilon>0$. Furthermore, if $A'_\lambda = b + P_i N P_i + \lambda P_i V P_i - 2^{i-1}$, then $\|A'_\lambda\| \leq e2^{2i}$. Therefore, Theorem 4.1 is applicable to the operator A'_λ. We make the transition from A'_λ to A_λ by using $N \leq \text{const}\, H_0$. Thus, as in the case of real λ, we obtain the convergence of Federbush's expansion (2.9) for λ in (4.1). This convergence implies that $H(\lambda)$ is a family of analytic operators of type (B) in the sense of Kato (Ref. 22, pp. 345–397). From the general theory of analytic perturbations, there follows

Theorem 4.2: (1) Let H_λ be the self-adjoint operator defined in Sec. II for $\lambda>0$. Then H_λ has a resolvent analytic continuation to the cut λ-plane.

(2) The ground state Ω_λ, normalized by $\|\Omega_\lambda\|=1$, $\Omega_\lambda \geq 0$ and the ground state energy E_λ have an analytic continuation to a neighborhood of the real axis.

Another property implied by the uniform convergence of (2.9) in the domain (4.1) is

Theorem 4.3: Let $\rho(H_0)$ the resolvent set of H_0. For $-b \in \rho(H_0)$

$$R_\lambda(b) = (H_0+\lambda V+b)^{-1} \to R_0(b) + (H_0+b)^{-1} \tag{4.9}$$

in norm as $|\lambda| \to 0$ in $|\arg\lambda| < \pi$.

Proof: Let $R^{(n)}(b;H_\lambda)$ be the nth term in the expansion (2.9). Then, one easily sees that for $n \leq 1$

$$R^{(n)}(b;H_\lambda) \xrightarrow{\text{norm}} 0 \tag{4.10}$$

as $|\lambda| \to 0$, $\lambda \in \{\lambda: |\arg\lambda| \leq \pi-\epsilon\}$; and for $n=0$

$$R^{(0)}(b;H_\lambda) \xrightarrow{\text{norm}} R_0(b) = (b+H_0)^{-1} \tag{4.11}$$

limits (4.10) and (4.11) imply (4.9)

Using standard "stability theorems" (Ref. 22, pp. 206–07, and Theorem 1.7, p. 368), we obtain

Corollary 4.4: For $\epsilon>0$, there is a $\Lambda>0$ such that if $z \in \rho(H_0)$, then $z \in \rho(H_\lambda)$ for $\lambda \in \{\lambda: |\text{avg}\lambda| \leq \pi-\epsilon, |\lambda| \leq \Lambda\}$.

Therefore, H_λ has only one eigenvalue near zero. This eigenvalue is analytic in $\{\lambda: |\arg\lambda| \leq \pi - \epsilon,\ |\lambda| \leq \Lambda\}$.

Remark: Theorem (4.3) is useful in proving[7,15,25] that the perturbation series for $E(\lambda)$, $\Omega(\lambda)$, and $W(\lambda)$, the equal time vacuum expectation valves, are asymptotic series which are Borel summable to the exact solutions.

*Present address: Department of Mathematics, University of Washington, Seattle, Washington 98195.

[1]J. Glimm and A. Jaffe, "Field Theory Models", in the *1970 Les Houches Lectures*, edited by C. DeWitt and R. Stora (Gordon and Breach, New York, 1971).
[2]E. Nelson, "A Quartic Intervation in Two Dimensions", in *Mathematical Theory of Elementary Particles*, edited by R. Goodman and J. Segal (M.I.T. Press, Cambridge, 1966).
[3]J. Glimm, Commun. Math. Phys. **8,** 12 (1968).
[4]J. Glimm, A. Jaffe, Ann. Math. **91,** 362 (1970).
[5]J. Glimm, A. Jaffe, Phys. Rev. **176,** 1945 (1968).
[6]I. Segal, Proc. Natl. Acad. Sci. USA **57,** 1178 (1967); J. Funct. Anal. **4,** 404 (1969); Bull. Am. Math. Soc. **75,** 1390 (1969); Ann. Math. **92,** 463 (1970).
[7]B. Simon and R. Höegh-Krohn, "Hypercontractive Semigroups and Two Dimensional Self-Coupled Bose Fields", J. Funct. Analysis, to appear.
[8]P. Federbush, "A Convergent Expansion for the Resolvent of $:\varphi^4:_{1+1}$," preprint.
[9]C. Bender and T. T. Wu, Phys. Rev. **184,** 1231 (1969).
[10]J. J. Loeffel and a. Martin, to appear in Proc. May, 1970, Programme 25 Conference, Strasburg.
[11]B. Simon, Ann. Phys. (N.Y.) **58,** 76 (1970).
[12]A. S. Wightman, "Analyticity in the Coupling Constant" in Proc. CNRS Conference, May, 1969.
[13]J. L. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B **30,** 656 (1969).
[14]S. Graffi, V. Greechi, and B. Simon, Phys. Lett. B **32,** 631 (1970).
[15]B. Simon, Phys. Rev. Lett. **25,** 1583 (1970).
[16]B. Gidas, J. Math. Phys. **15**, 867 (1973).
[17]N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963).
[18]A. Jaffe, Commun. Math. Phys. **1,** 127 (1965).
[19]J. Glimm and A. Jaffe, "The $\lambda(\varphi^4)_2$ Quantum Field Theory without Cutoffs. III", to appear.
[20]L. Schwartz, *La théorie des distributions, I, II* (Hermann, Paris, 1966).
[21]G. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1944).
[22]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[23]G. Meinardus, *Approximation von funktionen und ihre numerische Behandlung* (Springer-Verlag, Berlin, 1964).
[24]A. F. Timan, *Theory of Approximation of Functions of a Real Variable* (MacMillan, New York, 1963).
[25]L. Rosen and B. Simon, "The $(\varphi^{2n})_2$ Field Hamiltonian for Complex Complex Coupling Constant," preprint.

JMAPA 00NOV373 5 9 R 9: 3

On the self-adjointness of the Lorentz generator for $(:\varphi^4:)_{1+1}$

Basilis Gidas*

Department of Physics, Harrison M. Randall Laboratory of Physics, The University of Michigan, Ann Arbor, Michigan 48104

(Received 26 September 1972)

An alternative proof to that provided by Jaffe and Cannon of the self-adjointness of the local Lorentz generator for the $(:\varphi^4:)_{1+1}$ quantum field theory is given. The proof avoids the use of second-order estimates and a singular perturbation theory.

In this brief note, we establish the self-adjointness of the local Lorentz generator for the two-dimensional $:\varphi^4:$ interaction by the method of Ref. 1. This result has been previously obtained by Cannon and Jaffe[2] using first- and second-order estimates, and a singular perturbation theory. Here we avoid the use of second-order estimate and the Glimm–Jaffee singular perturbation theory.[3] It is hoped that a new proof may lead to some new results and insights.

The $(:\varphi^4:)_{1+1}$ quantum field theory has been brought to a very satisfactory stage mainly by the work of Glimm and Jaffe.[4] On the Fock space, they constructed a densely defined bilinear form $\varphi(x,t)$, continuous in x and t, which gives rise to a unique self-adjoint operator

$$\varphi(f)=\int dxdt\,\varphi(x,t)f(x,t) \tag{1}$$

for a real function $f\in C_0^\infty(R^2)$. The C^*-algebra of local observables is defined as the norm closure

$$=(\bigcup_B \quad (B))^-. \tag{2}$$

Here the union is taken over bounded regions B of space–time and (B) is the weakly closed (von Neumann) algebra generated by

$$\{\exp[i\varphi(f)]: f=\bar{f}\in C_0^\infty(R)\}. \tag{3}$$

The Poincaré group $P=\{a,\Lambda\}$ is the semidirect product of R^2 with R^1,

$$\{a,\Lambda\}\{a',\Lambda'\}=\{a,\Lambda a',\Lambda\Lambda'\}, \tag{4}$$

where $a\in R^2$ is a space–time translation, $a=(\alpha,\tau)$, and Λ is the one-parameter Lorentz rotation

$$\Lambda_\beta:(x,t)\to(x\cosh\beta+t\sinh\beta, x\sinh\beta+t\cosh\beta). \tag{5}$$

Poincaré covariance means that there exists a representation

$$\sigma_{\{a,\Lambda\}}(\quad (B))=\quad (\{a,\Lambda\}B) \tag{6}$$

for all bounded open sets B and all $\{a,\Lambda\}\in P$. The covariance of the local algebras ensures the covariance of the field operators, namely

$$\sigma_{\{a,\Lambda\}}(\varphi(f))=\varphi(f_{\{a,\Lambda\}}) \tag{7}$$

with

$$f_{\{a,\Lambda\}}(x,t)=f(\{a,\Lambda\}^{-2}(x,t)). \tag{8}$$

Space–time covariance was proven by Glimm and Jaffe.[5] The time translation is implemented locally by a unitary operator $U(t;B)$, i.e.,

$$\sigma_t(\quad (B))=U(t;B)\quad (B)U^{-1}_{(t;B)} \tag{9}$$

with

$$U(t;B)=\exp[itH(g)], \tag{10}$$

where $H(g)$ is the Hamiltonian with a space cutoff $g(x)\in C_0^\infty(R)$, $g(x)\equiv 1$ on a sufficiently large set depending on B. The space translation is implemented by $\exp(-ixP$, where P is the free field momentum operator.

The pure Lorentz transformation is locally implemented by a unitary operator $U(\Lambda_\beta;B)$, i.e.,

$$\sigma_{\Lambda_\beta}(\quad (B))=U(\Lambda_\beta;B)\quad (B)U^{-1}(\Lambda_\beta;B). \tag{11}$$

The formal infinitesimal generator of Lorentz transformations in a region B is

$$\begin{aligned}M(g)&=M_0+M_I(g)\\&=\int xH_0(x)dx+\int xH_I(x)g(x)dx,\end{aligned} \tag{12}$$

where the space cutoff function $g=1$ on a sufficiently large interval. Here, $H(x)=H_0(x)+H_I(x)$ is the energy density. Using space–time covariance, Cannon and Jaffe showed that it suffices to consider region B of space–time in the domain $x>0$. Also, it is technically convenient to use different spatial cutoffs in the free and the interaction part of M. Thus, for a region B in $x>0$, we take

$$M=M(g_0,g)=M_0+M_1, \tag{13a}$$

$$M_0=\alpha H_0, \tag{13b}$$

$$M_1=H_0(xg_1)+H_I(xg_2), \tag{13c}$$

where $\alpha>0$, $xg_i(x)$, $xg_2(x)\geq 0$, $g_0(x),g(x)\in C_0^\infty(R^+)$, and

$$\alpha+xg_1(x)=x=xg_2(x) \tag{14}$$

for x in a sufficiently large interval of the positive x axis. Here we have defined $g_0(x)=xg_1(x)$, and $g(x=xg_2(x)$. The first step toward proving that $M=M(g_0,g)$ is the infinitesimal generator for local Lorentz rotations, is to prove the self-adjointness of M.

We write

$$\begin{aligned}M=\alpha H_0&+H_{0,\kappa}(g_0)+H_{I,\kappa}(g)+[H_0(g_0)-H_{0,\kappa}(g_0)]\\&+[H_I(g)-H_{I,\kappa}(g)],\end{aligned} \tag{15}$$

where as usual κ is an upper momentum cutoff. We first estimate each term in (15). By undoing the Wick ordering we obtain

$$H_{0,\kappa}(g_0)\geq -c_1\kappa^2, \tag{16}$$

$$H_{I,\kappa}(g)\geq -c_2(\ln\kappa)^2, \tag{17}$$

where c_1,c_2 are positive constants independent of κ. By a standard N_τ estimate[6]

$$\begin{aligned}&\|(N+1)^{-1}(H_1(g)-H_{I,\kappa}(g))(N+I)^{-1}\|\leq c_3\kappa^{-1/2},\\&\times c_3>0.\end{aligned} \tag{18}$$

To estimate the difference $H_0(g_0)-H_{0,\kappa}(g_0)$, we write

$$H_0(g_0)=H_0^{(1)}(g_0)+H_0^{(2)}(g_0), \tag{19}$$

with

$$H_0^{(1)}(g_0)=\frac{1}{2(2\pi)}\int dk_1dk_2\,\hat{g}_0(k_1-k_2)\,\frac{\mu(k_1)\mu(k_2)+K_1k_2+\mu_0^2}{[\mu(k_1)\mu(k_2)]^{1/2}}\times a(k_1^*)a(k_2) \tag{20}$$

$$H_0^{(2)}(g_0)=\frac{2}{2(2\pi)}\frac{1}{2}\int dk_1dk_2\,\hat{g}_0(k_1-k_2)\,\frac{-\mu(k_1)\mu(k_2)+K_1k_2+\mu_0^2}{[\mu(k_1)\mu(k_2)]^{1/2}}\times(a^*(k_1)a^*(-ka)+a(-k_1)a(k_2)). \tag{21}$$

$H_0^{(1)}(g_0)$ is a sum of three terms having the form $A^*K(g_0)A$ in configuration space, where $K(g_0)$ is a multiplication operator with a nonnegative kernel. Therefore, $H_0^{(1)}(g_0)$, and, similarly, $H_0^{(1)}(g_0)=H_{0,\kappa}^{(1)}(g_0)$ are nonnegative operators. Jaffe and Cannon proved that $H_0^{(2)}(g_0)$ has an L_2 kernel and

$$\|(N+I)^{-1/2}(H_I^{(2)}(g_0)-H_{0,\kappa}^{(2)}(g_0))(N+I)^{-1/2}\|\le c_4^{\kappa-1/4},\quad c_4>0. \tag{22}$$

Finally, we estimate the free term αH_0 by

$$\alpha H_0\ge\alpha\mu_0 N. \tag{23}$$

Let P_n be the projection onto states with numbers of particles lying in the range

$$n^\beta\le N<(n+2)^\beta,\quad \beta\ge 4. \tag{24}$$

We note

$$\sum_{n=\text{even}}P_n=\sum_{n=\text{odd}}P_n=I. \tag{24'}$$

Picking $\kappa_n=\exp[(1/c_2)n^{\beta/2}]$, and using (16), (17), (18), (22) and (23), we quickly obtain

$$P_n\alpha H_0P_n\ge\alpha\mu_0P_nNP_n\ge\alpha\mu_0n^\beta, \tag{25}$$

$$P_nH_{0,\kappa_n}(g_0)P_n\ge -C,\quad \exp(2/c_2),\ n^{\beta/2}, \tag{26}$$

$$P_nH_{I,\kappa_n}(g)P_n\ge -n^\beta, \tag{27}$$

$$P_n(H_0^{(1)}(g_0)-H_{0,\kappa}^{(1)}(g_0))P_n\ge 0, \tag{28}$$

$$\|P_n(H_I(g)-H_{I,\kappa}(g))P_n\|\le d_1\exp(-d_1'N^{\beta/2},\quad d_1,d_1'>0,$$

$$\|P_n(H_0^{(2)}(g_0)-H_{0,\kappa}^{(2)}(g_0))P_n\|\le d_2\exp(-d_2'N^{\beta/2}), \tag{29}$$

$$d_2,d_2'>0. \tag{30}$$

Using (25) through (30) and choosing an appropriate α, we get

$$M_n=P_nMP_n\ge dn^\beta P_n, \tag{31}$$

where d is a positive constant. For d large enough we get

$$b+M_n\ge dn^\beta. \tag{32}$$

Let M' be obtained from M by replacing αH_0 by αN, a multiple of the particle number operator. Then

$$b+M_n\ge b+M_n'=b+P_n(\alpha N+H_0(g_0)+H_1(g))Pn. \tag{33}$$

By a standard N_τ estimate

$$\|b+M_n'\|\le d'N^{2\beta} \tag{34}$$

for some constant d'.

Following Ref. 7, we define P_e and P_d as the projection operators onto states with number of particles in the ranges

$$\bigcup_{n=\text{even}}(n^\beta-4\le N\le n^\beta+4) \tag{35}$$

and

$$\bigcup_{n=\text{odd}}(n^\beta-4\le N\le n^\beta+4), \tag{36}$$

respectively. We define M_e and M_d by

$$M_e=\sum_{n=\text{even}}M_n=\alpha H_0+\sum_{n=\text{even}}P_nM_1P_n, \tag{37}$$

$$M_d=\sum_{n=\text{odd}}M_n=\alpha H_0+\sum_{n=\text{odd}}P_nM_1P_n. \tag{38}$$

We write M in two different forms

$$M=M_e+L_e=M_d+L_d, \tag{39a}$$

$$L_e=M_1-\sum_{n=\text{even}}P_nM_1P_n, \tag{39b}$$

$$L_d=M_1-\sum_{n=\text{odd}}P_nM_1P_n, \tag{39c}$$

where the ranges (35) and (36) have been chosen so that

$$P_eL_eP_e=L_eP_e=P_eL_e=L_e, \tag{40}$$

$$P_dL_dP_d=L_dP_d=P_dL_d=L_d. \tag{41}$$

Federbush's expansion of the resolvent is

$$\begin{aligned}R(-b;M)=&R(-b;M_e)\\&-R(-b;M_d)L_eR(-b;M_e)\\&+R(-b;M_e)L_dR(-b;M_d)L_eR(-b;M_e)\\&-\cdots,\end{aligned} \tag{42a}$$

$$\begin{aligned}=&R(-b;M_e)\\&-R(-b;M_d)P_eL_eP_eR(-b;M_e)\\&+R(-b;M_e)P_dL_dP_dR(-b;M_d)P_eL_eP_e\\&\times R(-b;M_d)\\&-\cdots.\end{aligned} \tag{42b}$$

Our main result is the following theorem.

Theorem 1: Let g_0,g satisfy (14), with $g_0,g\ge 0$, $g_0,g\in C_0^\infty(R^+)$. Then there is a finite constant δ such that (42) converges in the uniform operator topology for $b>\delta$. The limit $R(-b)$ is the resolvent of a self-adjoint operator M such that $M>\delta$.

The proof of this theorem follows from the following three lemmas:

Lemma 1: For n large enough, there exist positive constants c_1, c_2 independent of n such that

$$\|P_eP_nR(-b;M_n)P_d\|\le c_1\exp(-c_2n^{\beta/2}). \tag{43}$$

Proof: Since $\alpha H_0\ge\alpha\mu_0N$, it is enough to prove (43) with M_n' replacing M_n. Estimates (32) and (34) permit us to apply Theorem 2.4 of Ref. 1, with $|a_n\rangle=P_nP_d|a\rangle$, $|\beta_n\rangle P_nP_e|b\rangle$, $\mu_n=dn^\beta$, $A=b+M_n'-dn^\beta D_n=d'n^2\beta$ (this corresponds to M in the notation of Ref. 1), and $N<\{[(n+1)^\beta-4]-(n^\beta+4)\}/4$. Thus we obtain (43).

Lemma 2: Let $\epsilon>0$ be smal enough. Then there exist constant $c(\epsilon)$ such that

$$\| R(-b;M_e)\|,\ \| R(-b;M_d)\| \leq 1/b, \tag{44}$$

$$\| L_e P_d R(-b;M_d)P_e\|,\ \| L_e P_e R(-b;M_e)P_d\| < \tfrac{1}{2}, \tag{45}$$

$$\| R(-b;M_d)L_e R(-b;M_e)\| \leq c(\epsilon)b^{-1-\epsilon}, \tag{46}$$

$$\| L_d R(-b;M_d)LR(-b;M_e)\| \leq c(\epsilon)^2 b^{-1-\epsilon}. \tag{47}$$

Proof: Inequality (44) is an easy consequence of estimate (31). Let $|a\rangle$ and $|b\rangle$ be two normalized states in the Fock space. To prove (45), we consider

$$\begin{aligned}\langle a|\, L_d P_d R(-b;M_d)P_e\,|\,b\rangle &= \sum_{n=\text{odd}} \langle a|\, L_d P_d R(-b;M_c)P_e P_n\,|\,b\rangle \\ &= \sum_{n=\text{odd}} \langle a|\, L_d P_d P_n R(-b;M_n)P_e\,|\,b\rangle \\ &= \sum_{n=\text{odd}} \langle a|\,(1-P_n)M_1 P_n P_d P_n \\ &\qquad \times R(-b;M_n)P_e\,|\,b\rangle \\ &\leq \sum_{n=\text{odd}} c_1(n+2)^{2\beta}\exp(-c_2 n^{\beta/2}) < \tfrac{1}{2}.\end{aligned}$$

In the last step above, we have used estimate (43) and a standard N_τ estimate. Similar arguments establish estimated (46) and (47).

Lemma 3: For b large enough, the series (42) converges uniformily to an operator $R(-b)$ which is a pseudoresolvent and satisfies

$$\lim_{b\to+\infty}(-b)R(-b) = I \tag{48}$$

in the norm operator topology.

Proof: Estimates (44) through (47) imply that the nth term in (42) is bounded by $c(\epsilon)^n b^{-1-n\epsilon}$. Therefore, the series converges for $b > c(\epsilon)^{1/\epsilon}$ to an operator $R(-b)$. Clearly,

$$\lim_{b\to+\infty} b\,[\| R(-b;M_d)L_e R(-b;M_e)\| + \cdots] = 0. \tag{49}$$

This implies (48). It is not hard to prove that $R(-b)$ is a pseudoresolvent.

Proof of Theorem 1: We follow the proof of Theorem 2.2 in Ref. 1 Equality (48) implies that $R(-b)$ is an inversible operator. Then $-b-(R(-b)^{-1}$ defines M whose domain is independent of b because of the pseudoresolvent property of $R(-b)$. The self-adjointness follows from the next lemma.[8]

Lemma 4: If $T:\mathcal{H}\to\mathcal{H}$ is an operator with dense domain on the Hilbert space $\mathcal{H}$, and if T^{-1} exists and has a dense domain, then $(T^*)^{-1} = (T^{-1})^*$.

*Present address: Department of Mathematics, University of Washington, Seattle, Washington 98195

[1]B. Gidas, J. Math. Phys. 15, 861 (1973).
[2]J. Cannon and A. Jaffe, Commun. Math. Phys. **17**, 261 .
[3]While this paper was in preparation, a set of notes by J. Glimm and A. Jaffe, "Boson Quantum Fields" came out, in which they treat the self-adjointness of the local Lorentz generater without using second-order estimates.
[4]J. Glimm and A. Jaffe, "Quantum Field Theory Models", in *Les Houches Lectures, 1970*, edited by C. Dewitt and R. Storna (Gordon and Breach, New York, 1971).
[5]J. Glimm and A. Jaffe, Am. Math. **91**, 362 (1970).
[6]J. Glimm, Commun. Math. Phys. **8**, 12 (1968).
[7]P. Federbush, "A Convergent Expansion for the Resolvent of $:\varphi^4:_{1+1}$", preprint.
[8]N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963).

On the theory of plasma turbulence

N. Marcuvitz

Polytechnic Institute of New York, Department of Electrical Engineering and Electrophysics, Graduate Center, Farmingdale, New York 11735
(Received 5 December 1973)

Microscopic to macroscopic transitions in classical, nonequilibrium, particle-field systems pose a class of stochastic perturbation problems that can be solved via the formal theory of scattering. Such solutions permit the nonlinear microscopic equations descriptive of a turbulent plasma field to be ensemble averaged into macroscopic particle- and wave-kinetic equations. The method is essentially operator theoretical. The original microscopic operator is first conventionally decomposed into deterministic and stochastic parts and then via scattering theory its inverse is evaluated in terms of explicit, formally exact, operator expressions that admit rapidly convergent, nonsecular, series representations. These results may be obtained by either operator algebra or diagram methods, the former being preferred. The derivation appears to be more physically and analytically transparent than in most existing procedures and has the virtue of exhibiting explicitly higher order terms, some of which are novel. The theory is illustrated for the case of a simple isotropic electron plasma by the derivation of kinetic equations for particles and for waves.

I. INTRODUCTION

Theories of plasma turbulence, in common with other field theories, employ perturbation theoretic procedures to derive series representations for the physical quantities of interest. The utility of such representations is dependent on the rapidity of convergence of these series. Operator techniques, employed in the spirit of the formal theory of scattering, will be shown to provide alternative and formally exact representations of turbulent field quantities and their n-point ensemble averages. In appropriate parametric ranges these formally exact expressions are expandable into rapidly convergent series; improper expansions lead to series with poor convergence or even secular terms.

The formal theory is illustrated for an electron plasma in which the initial state is defined by specification of the ensemble average of the stochastic electron distribution function and the 2-point correlation of the electric field intensity at $t=0$. The subsequent temporal evolution of the electron distribution function is derivable from a stochastic Green's function indicative of the possible phase space trajectories of an electron in a stochastic electric field with arbitrary statistical properties. For self-consistency these electric field properties are then identified via Maxwell's equations with those excited by the stochastic electron distribution. The difficult aspect of this initial value problem is the determination of a suitable expression for the stochastic Green's function of the Klimontovitch equation defining the electron distribution function.

The stochastic field operator characterizing the Klimontovitch equation is variously decomposable into an "unperturbed" average part and a "perturbed" stochastic part. The corresponding inverse operator, or Green's function, is representable via the methodology of the formal theory of scattering by a formally exact operator expression whose expandability is dependent on the "smallness" of the choice of the perturbation part. Different choices (renormalizations) lead to different convergence rates of series expansions. Renormalization procedures, both implicit and explicit, have been utilized by many workers in plasma turbulence (Dupree, Weinstock, Kraichnan, Kadomtsev, Ichikawa, Nishikawa, *et al.*[1]), in random wave propagation (Tatarskii, Keller, Furutzu, Bremmer, *et al.*[2]), and of course in quantum field theory (Prigogine, Balescu, Kubo, *et al.*[3]). Most of the renormalization procedures used in plasma turbulence are based, in the spirit of singular perturbation theory, on one or another *approximation* to the average particle Green's function, whose inverse represents the choice for the unperturbed component of the Klimontovich field operator (the diffusively streaming component). These approximations, which complicate the analysis, may be deferred by using formally exact expressions to obtain a precise definition of the n-point ensemble averages of interest; the latter may then be approximated at the final stage of analysis.

The principal results of this paper can be phrased in terms of a generic stochastic perturbation problem (7) whose exact stochastic Green's function, defined in (8), is explicitly decomposed in (23) into an ensemble averaged Green's function defined in (20) and a fluctuation therefrom. The latter components are expressed in terms of two coupled formally exact operator expressions given in (24) and (25), typical series representations of which are given in (30). These results yield in (37a) the formal solution of the original stochastic problem from which can be derived the various averages and n-point correlations necessary for the kinetic equation descriptive of average behavior.

For the case of the electron plasma, whose microscopic field equations are summarized in (1) and (2), the analysis of the Klimontovich equation leads to a precise definition of the average electron distribution function (37b) or equivalently the kinetic equation for electrons (47b). Only at the final stage of analysis are approximations introduced in the form of explicit series expansions whose convergence properties depend on the particular choice of perturbation or renormalization. A special choice of renormalization leads to Dupree's "modified orbit" analysis of plasma turbulence, derived by him via a test wave formalism[1]; the renormalization procedure serves to simplify and clarify Dupree's derivation and to display explicitly higher order terms. The self consistent stochastic electric field is defined by a generic equation (40) similar to (7) and may be solved by the same scattering formalism. For simplicity, however, the analysis is carried

through only approximately albeit with some novel features and leads to a kinetic equation for waves (57).

II. ABSTRACT FORMULATION OF BASIC PROBLEM

Let us make specific the initial value problem to be considered. At each space–time point $\mathbf{r}, t$ in a nonrelativistic electron plasma with immobile neutralizing ions, the dependence of the stochastic electron distribution function $\hat{f}=\hat{f}(\mathbf{v}, \mathbf{r}, t)$ on the stochastic electric field $\hat{\mathbf{E}}=\hat{\mathbf{E}}(\mathbf{r}, t)$ is defined by the Klimontovitch equation[4]

$$(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla-\frac{e}{m}\hat{\mathbf{E}}\cdot\nabla_{\mathbf{v}})\hat{f}=0 \tag{1}$$

where e, m, and $\mathbf{v}$ are the electron charge, mass, and velocity. The self-consistent dependence of the electric field $\hat{\mathbf{E}}$ on $\hat{f}$ is given by the Maxwell equations which, on elimination of the magnetic field $\hat{H}$, may be written in operator form as

$$\mathcal{Y}_M(\nabla, \frac{\partial}{\partial t})\cdot\hat{\mathbf{E}}=e\int\mathbf{v}\hat{f}d\mathbf{v} \tag{2}$$

where the operator (admittance) dyadic $\mathcal{Y}_M$ characteristic of the Maxwell field is given by

$$\mathcal{Y}_M\rightarrow\epsilon_0\frac{\partial}{\partial t}\mathbf{1}+\frac{\nabla\times\nabla\times\mathbf{1}}{\mu_0\partial/\partial t} \tag{2a}$$

with $\mathbf{1}$ the unit dyadic and ϵ_0, μ_0 the vacuum dielectric constant and permeability in MKS units. For uniqueness of solution to (1) and (2) initial values of $\hat{f}$, $\hat{\mathbf{E}}$, $\hat{\mathbf{H}}$ have to be prescribed; this implies a complete statement of their average and n-point correlations at say $t=0$. However, we shall limit these statements solely to averages and 2-point correlations. Before elucidating this point, let us introduce some notational concepts.

A stochastic field $\hat{\Psi}(\mathbf{r}, t)$ may be defined at each space time point $\mathbf{r}, t$ of a physical system by adducing an ensemble of similarly prepared systems. At the given point $\mathbf{r}, t$ the random function $\hat{\Psi}(\mathbf{r}, t)$ assumes different values in the various systems that constitute the ensemble. An ensemble average of the field $\hat{\Psi}(\mathbf{r}, t)$ can be evaluated at each point $\mathbf{r}, t$ and will be denoted by $\langle\hat{\Psi}(\mathbf{r}, t)\rangle=\Psi(\mathbf{r}, t)$; a more explicit definition of the average field in terms of a particular ensemble probability distribution function will be unnecessary for the following considerations. Knowledge of the average field permits decomposition of the stochastic field as

$$\hat{\Psi}(\mathbf{r}, t)=\Psi(\mathbf{r}, t)+\tilde{\Psi}(\mathbf{r}, t), \tag{3}$$

where $\tilde{\Psi}(\mathbf{r}, t)$ is the "fluctuation field". The fluctuation field $\tilde{\Psi}(\mathbf{r}, t)$ displays stochastic properties similar to that of the field $\hat{\Psi}(\mathbf{r}, t)$ but with the convenient property

$$\langle\tilde{\Psi}(\mathbf{r}, t)\rangle=0.$$

In consequence, one infers from (1) that the 2-point ensemble averages or autocorrelations $\langle\hat{\Psi}(\mathbf{r}, t)\,\hat{\Psi}(\mathbf{r}', t')\rangle$ and $\langle\tilde{\Psi}(\mathbf{r}, t)\,\tilde{\Psi}(\mathbf{r}', t')\rangle$ are related by

$$\langle\hat{\Psi}(\mathbf{r}, t)\,\hat{\Psi}(\mathbf{r}', t')\rangle=\Psi(\mathbf{r}, t)\,\Psi(\mathbf{r}', t')+\langle\tilde{\Psi}(\mathbf{r}, t)\,\tilde{\Psi}(\mathbf{r}', t')\rangle. \tag{5}$$

In the stochastic systems of interest our concern is primarily with the space–time evolution of the average field and the 2-point correlation of the fluctuation field.

With the above notation the initial value statement appropriate to (1) and (2) prescribes

$$f(\mathbf{v}, \mathbf{r}, 0),\quad\langle\tilde{f}(\mathbf{v}, \mathbf{r}, 0)\tilde{f}(\mathbf{v}', \mathbf{r}', 0)\rangle, \tag{6a}$$

$$\mathbf{E}(\mathbf{r}, 0),\quad\langle\tilde{\mathbf{E}}(\mathbf{r}, 0)\,\tilde{\mathbf{E}}(\mathbf{r}', 0)\rangle, \tag{6b}$$

with a similar statement for the magnetic field. One should recall that only "transverse" components of electric and magnetic fields can be prescribed independently; longitudinal electric field components follow from $\hat{f}$. The initial value problem involves determination of the evolution of the plasma state from the prescribed initial condition. Stationary turbulent states form a special class wherein the (not arbitrarily prescribable) initial state is stationary.

The primary difficulty in the self-consistent solution of (1) and (2) is associated with the nonlinear stochastic nature of the Klimontovitch equation (1). In Sec. 4 we shall find explicit expandable representations for $\hat{f}$ in terms of $\hat{\mathbf{E}}$. In Sec. 5 there are obtained a kinetic equation for particles, i.e., for $f(\mathbf{v}, \mathbf{r}, t)$, and a kinetic equation for waves, i.e., for the correlation spectrum $\mathcal{S}(\mathbf{k}, \mathbf{r}, t)$ of the 2-point correlation $\langle\tilde{\mathbf{E}}(\mathbf{r}, t)\,\tilde{\mathbf{E}}(\mathbf{r}', t')\rangle$.

In abstract form the Klimontovitch equation (1) can be written as

$$(L_0-\tilde{V})\hat{f}=0 \tag{7}$$

where, in a $\mathbf{v}, \mathbf{r}, t$ space the unperturbed operator L_0 and the perturbation $\tilde{V}$, whose average $\langle\tilde{V}\rangle=0$, are represented by

$$L_0\rightarrow\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla-\frac{e}{m}\mathbf{E}\cdot\nabla_v,$$

$$\tilde{V}\rightarrow\frac{e}{m}\tilde{\mathbf{E}}\cdot\nabla_v.$$

Equation (7) is a generic operator equation for the field $\hat{f}$ wherein the initial condition (6a), which defines the domain of the operator $L_0-\tilde{V}$, acts equivalently as a (deterministic and stochastic) impulsive source, applied at $t=0$. It is convenient to consider first a unit deterministic source; this defines the inverse operator, or stochastic Green's function, problem:

$$(L_0-\tilde{V})\hat{G}=\mathbf{1}, \tag{8}$$

where in the $\mathbf{v}, \mathbf{r}, t$ space[5]

$$\hat{G}\rightarrow\hat{g}(\mathbf{v}, \mathbf{r}, t;\ \mathbf{v}', \mathbf{r}', t'),$$

$$\mathbf{1}\rightarrow\delta(\mathbf{v}-\mathbf{v}')\,\delta(\mathbf{r}-\mathbf{r}')\,\delta(t-t'),$$

and $\delta(u)$ is the Dirac delta function. For uniqueness, the domain of $\hat{G}$ will be defined by the requirement $\hat{g}(\mathbf{v}, \mathbf{r}, t;\ \mathbf{v}', \mathbf{r}', t')=0$ for $t<t'$; whence (8) defines the "creation" Green's function, representative of the creation of an electron at $\mathbf{v}', \mathbf{r}'$ at time t' and of its subsequent "propagation" to a point $\mathbf{v}, \mathbf{r}$ in phase space at time t. Explicit representations of the operator $\hat{G}$ shall be derived in Sec. 3 both by diagrammatic and operator algebra procedures.

III. REPRESENTATIONS OF G

Methods familiar from the formal theory of scattering provide means for solving (8) for $\hat{G}$ in a variety of ways.[6] A conventional procedure first introduces the coherent unperturbed operator G_0, inverse to the opera-

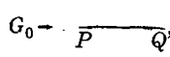
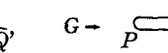
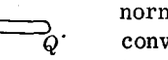
tor L_0, and defined by

$$L_0 G_0 = 1, \tag{9}$$

where the domain of L_0 is such that $G_0 \to g_0(\mathbf{v}, \mathbf{r}, t; \mathbf{v}', \mathbf{r}', t') = 0$ for $t < t'$, as is the case for $\hat{G}$. The Green's function G_0 is representative of the phase space trajectory for unperturbed "propagation" to $\mathbf{v}, \mathbf{r}, t$ of an electron created at $\mathbf{v}', \mathbf{r}', t'$. The perturbation operator $\tilde{V}$ in (8) is viewed as "scattering" the unperturbed propagation, G_0, the latter being assumed known. This scattering viewpoint and its relation to diagram techniques will now be considered.

Diagram method

A pictorial overview of the relevance of scattering concepts in the evaluation of $\hat{G}$ and its ensemble average $G = \langle \hat{G} \rangle$ is expressible in conventional diagrammatic terms. Elucidation of these terms will be omitted at this point since their derivation follows more simply via the operator method below. Creation of an electron at one point P and its propagation to another point Q, as represented either by $\hat{G}$, G_0, or G, will be pictured respectively by

$\hat{G} \to$  $G_0 \to$  $G \to$ 

Thus, in Fig. (1a), the stochastic propagation, $\hat{G}$, of an electron from a creation point P to an observation point Q will be pictured in terms of a direct trajectory, G_0, and a scattered contribution in two different but related ways. As shown in Fig. (1a), the upper right-hand diagram depicts the scattered contribution in terms of an electron starting from the creation point P, undergoing unperturbed "propagation" to the scatterer $\tilde{V}$, being "scattered" and thence stochastically "propagating" via $\hat{G}$ to the observation point Q. The alternative lower diagram in Fig. (1a) displays the scattered contribution in terms of unperturbed electron propagation via G_0 to and from a "multiple scatterer" $\hat{T}$. By iterative substitution for $\hat{G}$ in the first scattering diagram one can infer that the operator $\hat{T}$ is representable as an infinite series of conventional diagrams involving $\tilde{V}$ and G_0. Although the terms of this series display secularities, or divergences, the lower $\hat{T}$ diagram in Fig. (1a) can be ensemble averaged to obtain the diagrammatic representation for G, in terms of $T = \langle \hat{T} \rangle$, shown in the upper diagram of Fig. (1b). The lower diagram of Fig. (1b) arises from selective summation of the so-called (nonsecular) connected diagrams in the T representation, and provides an alternative picture of the scattered part of the averaged electron propagation G; the electron first undergoes unperturbed propagation to the "effective" average scatterer V_c and thence average propagation to the observation point. The hereby defined "effective field" or "mass operator" V_c is composed of only the connected diagrams in the series representation of T; it is thereby represented as a summable but sometimes poorly convergent series.

The above diagrammatic derivation is conventional[2]; it is reviewed primarily to exhibit the isomorphic quartets of operators $\hat{G}$, $\tilde{V}$, $\hat{T}$, G_0 and of G, V_c, T, G_0, which always appear in a scattering phrasing of perturbation theoretic analyses. For solution of say initial value problems, difficulties in the analytical use of these formally exact operators arise from the need to represent them as series, which may be secular or poorly convergent; if so, the use of renormalization techniques permit more convergent series representations.

Assuming the operator $\hat{T}$ in Fig. (1a) is known, and hence also the average operator T in Fig. (1b), one can illustrate a renormalization procedure by subtraction of the $\hat{G}$ and G diagrams in these figures. One thereby infers the alternative (renormalized) diagrams shown in Fig. (1c). The upper right-hand diagram, obtained by subtraction of the lower and upper diagrams in Figs. (1a) and (1b), resp., now depicts the unperturbed electron propagation in terms of G, and the scattering contribution in terms of $\tilde{T} = \hat{T} - T$ and G_0. Since G_0 and G are related as shown in Fig. (1b), one is led successively to the middle and lower right-hand diagrams of Fig. (1c). In terms of the new quartet of operators $\hat{G}$, $\tilde{V} - V_c$, $\tilde{T}_c$, G the stochastic electron propagation $\hat{G}$ is now represented as taking place in an averaged medium, wherein direct unperturbed propagation is characterized by G and scattering by $\tilde{V} - V_c$ or $\tilde{T}_c$. Scattering in this medium is given by the "small" renormalized scattering operator $\tilde{V} - V_c$ with $\tilde{T}_c$ as a convergentally representable multiple scattering operator. A noteworthy feature of this renormalized representation of $\hat{G}$, depicted in the middle right hand diagram of Fig. (1c), is the explicit decomposition of $\hat{G}$ into an average part G and a fluctuation $G\tilde{T}_c G$.

Operator method

Analytical equivalents of the $\hat{G}$ diagram representations in Fig. (1) are derivable by operator algebraic manipulations of the defining equation (8) for $\hat{G}$. To do so, we first complete the operator definition of $\hat{G}$ by

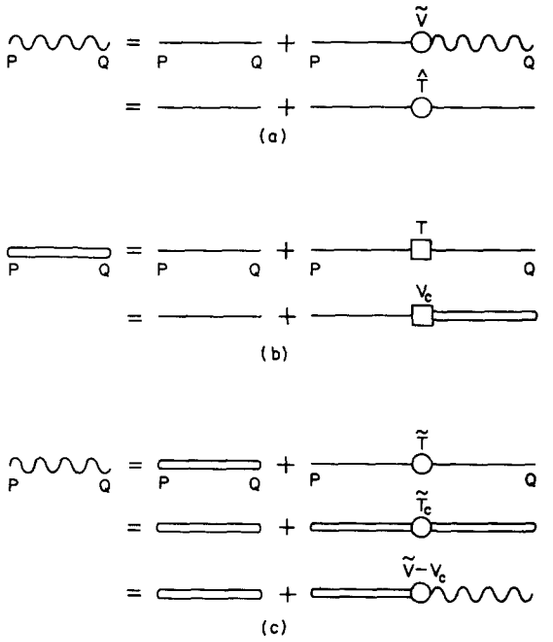


FIG. 1. (a) Diagram representation of $\hat{G}$ in terms of G_0; (b) diagram representation of G in terms of G_0; (c) diagram representation of $\hat{G}$ in terms of G.

appending a statement as to the domain and range of the operator $L_0 - \tilde{V}$. This is done by defining, as above, $\hat{G}$ as the "creation" Green's function and also defining an adjoint to $\hat{G}$ by a more complete rewriting of (8) as[7]

$$(L_0 - \tilde{V})\hat{G} = 1 = \hat{G}(L_0 - \tilde{V}). \tag{10}$$

A similar completeness of definition of the unperturbed creation Green's function G_0 leads to the rewriting of (9) as

$$L_0 G_0 = 1 = G_0 L_0, \tag{11}$$

whose solution G_0 is presumed known. On left multiplication of (10) by G_0 and use of (11), one then obtains as the analytical equivalent of the diagrams in Fig. (1a)

$$\hat{G} = G_0 + G_0\tilde{V}\hat{G} = G_0 + G_0\hat{T}G_0, \tag{12}$$

where the right-hand expression results from the definition

$$\tilde{V}\hat{G} = \hat{T}G_0. \tag{13}$$

The $\hat{T}$ operator, representative of multiple scattering by the stochastic perturbation $\tilde{V}$ in the background G_0, is given, on using (12) to derive the relation $\hat{G} = (1 - G_0\tilde{V})^{-1}G_0$, by the formally exact abstract expressions

$$\hat{T} = \tilde{V}(1 - G_0\tilde{V})^{-1} = \tilde{V} + \tilde{V}\hat{G}\tilde{V}. \tag{14}$$

In passing, it is of interest to note that the inverse operator $(1 - G_0\tilde{V})^{-1}$ may be represented by the conventional Neumann series

$$(1 - G_0\tilde{V})^{-1} = 1 + G_0 V + G_0\tilde{V}G_0\tilde{V} + \cdots. \tag{15}$$

However, the singular ("curvilinear orbit") nature of G_0 implies that the terms of the series (15) display secular divergences[2] that limit its range of applicability for representation of $\hat{T}$.

The formally exact expressions (12) and (14), although not readily expandable, may be ensemble averaged to yield, as the equivalent of the Fig. (1b) diagrams, the ensemble averaged G in the form

$$G = G_0 + G_0 V_c G = G_0 + G_0 T G_0, \tag{16}$$

where

$$T = \langle\hat{T}\rangle = \langle\tilde{V}G\tilde{V}\rangle \tag{17}$$

is the ensemble averaged multiple scattering operator and where, in analogy to the stochastic definition (13), one has introduced the coherent definition

$$V_c G = TG_0. \tag{18}$$

From (18) and (16), one observes, isomorphic to (14), that

$$T = V_c(1 - G_0 V_c)^{-1} = V_c + V_c G V_c. \tag{19}$$

The average operator G, defined in (16) depicts a coherent scattering process characterized by the "smoothed" scattering operator V_c in a background G_0. An alternative (differential equation type) definition, analogous to (10), follows from multiplication of (16) by L_0 and use of (11) as

$$(L_0 - V_c)G = 1 = G(L_0 - V_c), \tag{20}$$

which provides a nonlinear defining equation for G, the nonlinearity arising from the dependence of V_c on G.

Although the average Green's function G is of interest, *per se*, as a conditional probability for finding an electron created at a prescribed point in phase space, it can also be utilized as the unperturbed background in a renormalized scattering representation of $\hat{G}$. Thus, on subtraction of the right-hand expressions in (12) and (16), one obtains

$$\hat{G} = G + G_0\tilde{T}G_0, \tag{21a}$$

as pictured in the upper right-hand diagram of Fig. (1c), where from (14)

$$\tilde{T} = \hat{T} - T = \tilde{V} + \tilde{V}\hat{G}\tilde{V} - \langle\tilde{V}\hat{G}\tilde{V}\rangle. \tag{21b}$$

Noting from (16) that

$$G_0 = G(1 + V_c G)^{-1} = (1 + GV_c)^{-1}G, \tag{22}$$

one can reexpress (21a) as

$$\hat{G} = G + G\tilde{T}_c G, \tag{23}$$

where the stochastic operator $\tilde{T}_c$, representative of multiple scattering in the smoothed background, G, is given by

$$\tilde{T}_c = 1 + V_c G)^{-1}\tilde{T}(1 + GV_c)^{-1}. \tag{24}$$

The operator V_c is derivable from $\tilde{T}_c$. On averaging (10) and comparing with (20), one infers $V_c G = \langle\tilde{V}\hat{G}\rangle$ whence on utilizing (23)

$$V_c = \langle\tilde{V}G\tilde{T}_c\rangle. \tag{25}$$

Substitution of (25) and (21b) into (24) provides a formally exact operator equation for $\tilde{T}_c$ that can be iteratively expanded because of the nonsingular (diffusively spread orbit) nature of G.

The representation (23) of $\hat{G}$ is a desirable form in that the direct and scattered components of $\hat{G}$ are explicitly identifiable as ensemble average and fluctuation components, the latter having a zero average. A revealing insight into the physical significance of (23) is obtained by first rewriting the defining equation (10) for $\hat{G}$ in the renormalized form

$$(L_0 - V_c - \hat{V}_1)\hat{G} = 1, \tag{26}$$

whence the scattering is now caused by a stochastic perturbation $\hat{V}_1 \equiv \tilde{V} - V_c$ of the coherent renormalized background operator $L_0 - V_c$. On left multiplication of (26) by G and use of (20), one obtains

$$\hat{G} = G + G\hat{V}_1\hat{G} = G + G\tilde{T}_c G, \tag{27}$$

where the right-hand expression follows from the defining identity

$$\hat{V}_1\hat{G} = \tilde{T}_c G,$$

whence one infers on use of the left part of (27) that

$$\tilde{T}_c = \hat{V}_1(1 - G\hat{V}_1)^{-1} = \hat{V}_1 + \hat{V}_1\hat{G}\hat{V}_1 \tag{28a}$$

and by (25)

$$V_c = \langle\hat{V}_1 G\tilde{T}_c\rangle. \tag{28b}$$

The alternative expressions (24) and (28a) for $\tilde{T}_c$ can be shown to be equivalent, as may be anticipated from the identification of (23) and (27). Incidentally, (27) is the analytical equivalent of the lower diagrams in Fig. (1c).

The utility of the representation (27) for the stochastic Green's function $\hat{G}$ is dependent on one's ability both to evaluate $G=\langle\hat{G}\rangle$ and to obtain rapidly convergent representations of the formally known operators $\tilde{T}_c$ and V_c. By series expansion of the inverse operators in (24) one obtains for $\tilde{T}_c$

$$\tilde{T}_c=(1-V_cG+V_cGV_c\ldots)\tilde{T}(1-GV_c+GV_cGV_c\ldots) \quad (29a)$$

where on use of (23) in (21b)

$$\tilde{T}=\tilde{V}+\tilde{V}G\tilde{V}-\langle\tilde{V}G\tilde{V}\rangle+\tilde{V}G\tilde{T}_cG\tilde{V}-\langle\tilde{V}G\tilde{T}_cG\tilde{V}\rangle. \quad (29b)$$

On substitution of (25) into (29a), one derives by iterative use of (29b) the following series expansions (to fourth order in $\tilde{V}$) for $\tilde{T}_c$:

$$\begin{aligned}\tilde{T}_c=\tilde{V}&+\tilde{V}G\tilde{V}-\langle\tilde{V}G\tilde{V}\rangle+\tilde{V}G\tilde{V}G\tilde{V}-\langle\tilde{V}G\tilde{V}G\tilde{V}\rangle\\&-\langle\tilde{V}G\tilde{V}\rangle G\tilde{V}-\tilde{V}G\langle\tilde{V}G\tilde{V}\rangle\\&+\tilde{V}G\tilde{V}G\tilde{V}G\tilde{V}-\langle\tilde{V}G\tilde{V}G\tilde{V}G\tilde{V}\rangle-\tilde{V}G\langle\tilde{V}G\tilde{V}\rangle G\tilde{V}\\&+\langle\tilde{V}G\langle\tilde{V}G\tilde{V}\rangle G\tilde{V}\rangle\\&-\langle\tilde{V}G\tilde{V}\rangle G\tilde{V}G\tilde{V}-\tilde{V}G\tilde{V}G\langle\tilde{V}G\tilde{V}\rangle+2\langle\tilde{V}G\tilde{V}\rangle G\langle\tilde{V}G\tilde{V}\rangle\\&-\langle\tilde{V}G\tilde{V}G\tilde{V}\rangle G\tilde{V}-\tilde{V}G\langle\tilde{V}G\tilde{V}G\tilde{V}\rangle+\ldots\end{aligned} \quad (30a)$$

whence by (25)

$$\begin{aligned}V_c=\langle\tilde{V}G\tilde{V}\rangle&+\langle\tilde{V}G\tilde{V}G\tilde{V}\rangle+\langle\tilde{V}G\tilde{V}G\tilde{V}G\tilde{V}\rangle-\langle\tilde{V}G\langle\tilde{V}G\tilde{V}\rangle G\tilde{V}\rangle\\&-\langle\tilde{V}G\tilde{V}\rangle G\langle\tilde{V}G\tilde{V}\rangle+\ldots\end{aligned} \quad (30b)$$

The range of validity of the series expansions (30) is difficult to ascertain because of the complex nonlinear nature of the Eq. (24) defining $\tilde{T}_c$. Since the expansions (30) are in terms of the average operator G, which in contrast to G_0 is nonsingular at least until trapping is well developed, they are more rapidly convergent than nonsecular perturbative expansions in terms of G_0. Their domain of convergence is limited by the requirement that the bound or norm of the ($\tilde{T}_c$ dependent) operator product GV_c is less than unity—i.e., that $\tilde{V}$ and V_c are "small."

Representations of $\tilde{T}_c$ and V_c with an ordering of terms different than that in (30) can be obtained by expansion of the expression (28a). Thus for $\hat{V}_1=\tilde{V}-V_c$ "small", but with $\tilde{V}$ and V_c permitted to be "large," one finds by expansion of (28) to second order in $\hat{V}_1$:

$$\tilde{T}_c=\tilde{V}_1+\langle\tilde{V}_1G\tilde{V}_1\rangle+\ldots, \quad (31a)$$

$$V_c=\langle\hat{V}_1G\hat{V}_1\rangle+\ldots. \quad (31b)$$

In terms of $\tilde{V}$ one has, to indicated order in $\hat{V}_1$,

$$\tilde{T}_c=\tilde{V}+\tilde{V}G\tilde{V}-\langle\tilde{V}G\tilde{V}\rangle-\tilde{V}G\langle\tilde{V}G\tilde{V}\rangle-\langle\tilde{V}G\tilde{V}\rangle G\tilde{V}+\ldots, \quad (32a)$$

$$V_c=\langle\tilde{V}G\tilde{V}\rangle+\langle\tilde{V}G\tilde{V}\rangle G\langle\tilde{V}G\tilde{V}\rangle+\ldots. \quad (32b)$$

The ordering of terms in the representations (32) is evidently different than in (30); inclusion of the remaining terms will of course lead to identical overall series.

It is of interest to contrast diagrammatically the V_c representation (30b) in terms of G with the conventional representation[2] in terms of G_0. The diagram equivalent of Eq. (30b) is displayed in Fig. (2a). The corresponding and less rapidly convergent G_0 representation is shown in Fig. (2b); the latter follows from that in Fig. (2a) on use of the diagram equivalent of the relation between G and G_0 given in (22).

The utility of the representations (30) is predicated upon an explicit knowledge of the average operator G. In a frequently employed procedure for evaluation of G one expresses the Fourier ($\mathbf{k}$, ω) spectrum of $\hat{G}$ in terms of the stochastic $\mathbf{v}$, $\mathbf{r}$, t phase space trajectory of an electron in an $\hat{\mathbf{E}}$ field and employs cumulant expansions[8] to find the ensemble average $G=\langle\hat{G}\rangle$ in terms of various n-point averages of $\hat{\mathbf{E}}$. In abstract, however, one has a difficult task of solving the nonlinear equation (20); the complexity of the task arises from the nonlinear nature of V_c. Since V_c is given in (30b) as a series of the form

$$V_c=A_1+A_2+A_3\ldots, \quad (33)$$

a standard perturbative procedure consists of rewriting (20) as

$$(L_0-A_1-V_{c1})G=1 \quad (34)$$

with $V_{c1}=V_c-A_1$, defining an "unperturbed" operator G_1 by

$$(L_0-A_1)G_1=1=G_1(L_0-A_1), \quad (35)$$

and deriving in the usual manner the solution

$$G=(1-G_1V_{c1})^{-1}G_1$$

which is expandable as

$$G=G_1+G_1V_{c1}G_1+G_1V_{c1}G_1V_{c1}G_1+\ldots. \quad (36)$$

A familiar choice (Kraichnan, Dupree, Tatarskii) for A_1 is suggested by the first term in (30b), viz,

$$A_1=\langle\tilde{V}G_1\tilde{V}\rangle \quad (36a)$$

which leads to a nonlinear equation (35) for G_1 that is solvable for restricted parameter ranges. A less optimum, but explicit, choice (Bourett,[2] "straight line orbit") is

$$A_1=\langle VG_0V\rangle, \quad (36b)$$

with G_0 defined by (11) leads to a linear equation for G_1. In practice one usually employs only the first term G_1 in the series (35) to approximate G, the higher order

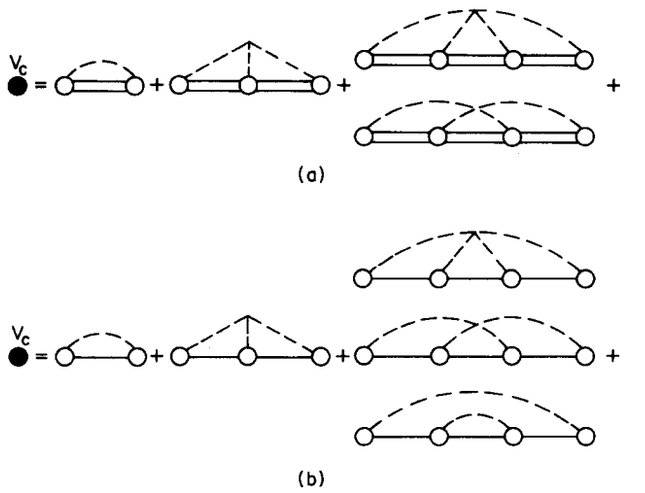


FIG. 2. Diagram representation of effective field operator V_c: (a) renormalized, (b) conventional; $\bullet\!\rightarrow V_c$, $\circ\!\rightarrow\tilde{V}$.

terms providing an error estimate. Better perturbative approximations for G are obtainable in a conventional manner by using A_1+A_2 instead of A_1, $V_{c2}=V_c-A_1-A_2$ instead of V_{c1}, G_2 instead of G_1, etc., in (34) and (35).

IV. REPRESENTATIONS OF $\hat{f}$ AND $\hat{\mathbf{E}}$

The derivation in Sec. III of an abstract expression for $\hat{G}$ permits a representation of the solution of Eq. (1) for $\hat{f}$ in terms of the statistical properties of $\hat{\mathbf{E}}$ and of the cross correlation of $\hat{\mathbf{E}}$ with a stochastic distribution function $\hat{f}^a$ indicative of the phase space evolution of the statistical initial condition $\hat{f}(\mathbf{v}, \mathbf{r}, 0)$ in an average background described by G. To determine the desired statistical properties of $\hat{\mathbf{E}}$ one utilizes the Maxwell equation (2), for $\hat{\mathbf{E}}$ in terms of the above $\hat{f}$ solution, to obtain a defining equation for $\hat{\mathbf{E}}$ in terms of the statistical properties of $\hat{f}^a$ and of their cross correlation with the initially prescribed statistics of $\hat{\mathbf{E}}(\mathbf{r}, 0)$. Such self consistent representations of $\hat{f}$ and $\hat{\mathbf{E}}$ permit one to derive the kinetic equation for particles (i.e., for f) and the kinetic equation for waves [i.e., for the correlation spectrum density of $\langle\tilde{\mathbf{E}}(\mathbf{r}, t)\tilde{\mathbf{E}}(\mathbf{r}', t')\rangle$]; these kinetic equations are coupled and must be solved self consistently.

The abstract expression for $\hat{G}$ derived in (27), viz,

$$\hat{G}=G+G\tilde{T}_cG \tag{37}$$

explicitly exhibits average and fluctuation components and permits a similar component representation of the abstract solution $\hat{f}$ of the Klimontovitch equation (1). Since $\hat{G}$ arises from a coherent point source, whereas $\hat{f}$ may be regarded as arising from a stochastically distributed impulsive source $\hat{s}\rightarrow\hat{f}(\mathbf{v}, \mathbf{r}, 0)\delta(t)$ (representative of the initial excitation), Eq. (1) yields the desired solution

$$\hat{f}=\hat{f}^a+G\tilde{T}_c\hat{f}^a, \tag{37a}$$

where $\hat{f}=\hat{G}\hat{s}$, and where $\hat{f}^a=G\hat{s}$ represents the "average" evolution of the stochastic initial state which for $t>0$ satisfies the homogenous equation

$$(L_0-V_c)\hat{f}^a=0.$$

The average f follows explicitly from (37a) as*[9]

$$f=f^a+G\langle\tilde{T}_c\tilde{f}^a\rangle, \tag{37b}$$

whence by subtraction of (37a) and (37b) the fluctuation $\tilde{f}$ is given by

$$\tilde{f}=\tilde{f}^a+G\tilde{T}_cf^a+G(\tilde{T}_c\tilde{f}^a-\langle\tilde{T}_c\tilde{f}^a\rangle). \tag{37c}$$

The statistical properties of $\hat{\mathbf{E}}$ follow from $\hat{f}$ in (37a) on use of the Maxwell equation (2). Since this determination is effected in physical $\mathbf{r}, t$ space, rather than $\mathbf{v}, \mathbf{r}, t$ phase space, it is convenient to employ in physical space the abstract notation $\hat{f}(\mathbf{v})$ instead of $\hat{f}$, and also $g(\mathbf{v}, \mathbf{v}')$ instead of G, $\tilde{T}_c(\mathbf{v}, \mathbf{v}')$ instead of $\tilde{T}_c$, etc. Thus, in physical space, the abstract solution (37a) is written

$$\hat{f}(\mathbf{v})=\hat{f}^a(\mathbf{v})+\int g(\mathbf{v}, \mathbf{v}_1)\tilde{T}_c(\mathbf{v}_1, \mathbf{v}_2)\hat{f}^a(\mathbf{v}_2)d\mathbf{v}_1d\mathbf{v}_2, \tag{38}$$

whence on substitution into the Maxwell equation (2) one has exactly

$$\mathcal{Y}_M\cdot\hat{\mathbf{E}}=e\int\mathbf{v}\hat{f}^a(\mathbf{v})d\mathbf{v}+e\int\mathbf{v}g(\mathbf{v}, \mathbf{v}_1)\tilde{T}_c(\mathbf{v}_1, \mathbf{v}_2)\hat{f}^a(\mathbf{v}_2)d\mathbf{v}d\mathbf{v}_1d\mathbf{v}_2. \tag{39}$$

Utilizing for simplicity only the first term in the $\tilde{T}_c$ expansion (30a), viz., $\tilde{T}_c\simeq\tilde{V}\rightarrow(e/m)\tilde{\mathbf{E}}\cdot\nabla_v$, one approximately reexpresses (39) in physical space as an abstract nonlinear equation for $\hat{\mathbf{E}}$ (neglecting *nonlinear* wave "parametric" and particle damping terms):

$$(Y_0-\tilde{V}_e)\hat{E}=-\hat{J}, \tag{40}$$

where

$$\begin{aligned}
&Y_0=Y_M+V_e\rightarrow\mathcal{Y}_M-\frac{e^2}{m}\int\mathbf{v}g(\mathbf{v}, \mathbf{v}_1)\nabla_{\mathbf{v}_1}f^a(\mathbf{v}_1)d\mathbf{v}d\mathbf{v}_1,\\
&V_e=V_e+\tilde{V}_e\rightarrow\frac{e^2}{m}\int\mathbf{v}g(\mathbf{v}, \mathbf{v}_1)\nabla_{\mathbf{v}_1}\hat{f}^a(\mathbf{v}_1)d\mathbf{v}d\mathbf{v}_1,\\
&\hat{J}=\hat{J}^a+\hat{V}_eE+\ldots,\\
&-\hat{J}^a-e\int\mathbf{v}\hat{f}^a(\mathbf{v})d\mathbf{v}.
\end{aligned} \tag{41}$$

Equation (40)has the same abstract form as the Klimontovitch equation (7) for $\hat{f}$, save for the minor addition of the right-hand source term. Thus, as in the case of (10), one introduces an abstract (in $\mathbf{r}, t$ space) stochastic Green's function $\hat{Z}$ by

$$(Y_0-\tilde{V}_e)\hat{Z}+1, \tag{42}$$

whose abstract solution is isomorphic to that given in (23), namely,

$$\hat{Z}=Z+Z\tilde{T}_{ce}Z, \tag{43}$$

where the ensemble averaged Green's function Z is defined as in (20) by

$$(Y_0-V_{ce})Z=1, \tag{44}$$

V_{ce}, the "modified" medium operator, being indicative on the average of wave (quasiparticle) scattering by electron density fluctuations. As in (30) the first few terms in the expandable representation of $\tilde{T}_{ce}$ and V_{ce} are

$$\tilde{T}_{ce}=\tilde{V}_e+\tilde{V}_eZ\tilde{V}_e-\langle\tilde{V}_eZ\tilde{V}_e\rangle+\ldots, \tag{45a}$$

$$V_{ce}=\langle\tilde{V}_eZ\tilde{V}_e\rangle+\langle\tilde{V}_eZ\tilde{V}_eZ\tilde{V}_e\rangle+\ldots. \tag{45b}$$

One then infers from (43) the desired abstract representation of the solution $\hat{E}$ of (41) as

$$\hat{E}=\hat{E}_\alpha+Z\tilde{T}_{ce}\hat{E}_\alpha, \tag{46a}$$

where $\hat{E}_\alpha$, the stochastic electric field excited by both the initial condition [impulse $\hat{J}^s\rightarrow\epsilon_0\hat{\mathbf{E}}(\mathbf{r}, 0)\delta(t)$] and the current source $\hat{J}$ acting in an average medium with wave propagation described by Z, is

$$\hat{E}_\alpha=\hat{E}^a-Z\hat{J}=-Z(\hat{J}^s+\hat{J});$$

the stochastic field $\hat{E}^a=-Z\hat{J}^s$ evidently represents the evolution of the stochastic initial field $\hat{\mathbf{E}}(\mathbf{r}, 0)$ in the average medium described by Z. From (46a) the ensemble average E follows as

$$E=E_\alpha+Z\langle\tilde{T}_{ce}\tilde{E}_\alpha\rangle \tag{46b}$$

and the fluctuation $\tilde{E}$ as

$$\tilde{E}=\tilde{E}_\alpha+Z\tilde{T}_{ce}E_\alpha+Z(\tilde{T}_{ce}\tilde{E}_\alpha-\langle\tilde{T}_{ce}\tilde{E}_\alpha\rangle). \tag{46c}$$

The coupled equations (37) and (46) comprise the desired self consistent representations of $\hat{f}$ and $\hat{E}$.

One can rewrite the "integral representations" (37 a, b, c) in "differential equation" form as

$$(L_0-V_c)\hat{f}=\tilde{T}_c\hat{f}^a \tag{47a}$$

whose average and fluctuation components are

$$(L_0 - V_c)f = \langle \tilde{T}_c \tilde{f}^a \rangle, \tag{47b}$$

$$(L_0 - V_c)\tilde{f} = \tilde{T}_c f^a + (\tilde{T}_c \tilde{f}^a - \langle \tilde{T}_c \tilde{f}^a \rangle). \tag{47c}$$

From the definition of $\hat{f}^a$ given in connection with (37a) and with the use of Eq. (20) for G, one observes that, subsequent to the excitation $\hat{s}$, Eq. (47a) rephrases the original Klimontovitch equation (7) in terms of a renormalized *coherent* operator $L_0 - V_c$ and a stochastic source $\tilde{T}_c \hat{f}^a$, both of which are formally expressed in terms of the statistical properties of $\tilde{E}$ and $\tilde{f}^a$. Equation (47b) leads to an abstract version of a generalized Lenard–Balescu or Fokker–Planck equation (a kinetic equation for particles) provided the right-hand source term, which is dependent on the correlation of $\tilde{\mathbf{E}}$ and $\tilde{f}^a$, is evaluated with the aid of the Maxwell equations. Equations (47b) and (47c) display the coupling and dependence of the background f and fluctuation $\tilde{f}$ on the statistical properties of the electric field E.

Correspondingly, subsequent to initial excitation, the integral representations (46) of $\hat{E}$ can be rewritten in "differential equation" form as

$$(Y_0 - V_{ce})\hat{E} = -\hat{J} + \tilde{T}_{ce}\hat{E}_\alpha \tag{48a}$$

whose average and fluctuation components are manifestly

$$(Y_0 - V_{ce})E = -J + \langle \tilde{T}_{ce}\tilde{E}_\alpha \rangle, \tag{48b}$$

$$(Y_0 - V_{ce})\tilde{E} = -\tilde{J} + \tilde{T}_{ce}E_\alpha + (\tilde{T}_{ce}\tilde{E}_\alpha - \langle \tilde{T}_{ce}\tilde{E}_\alpha \rangle). \tag{48c}$$

From the definition of E_α and E^a given under (46a), it is evident that for $t > 0$

$$(Y_0 - V_{ce})\hat{E}_\alpha = -\hat{J} \quad \text{and} \quad (Y_0 - V_{ce})\hat{E}^\alpha = 0. \tag{49}$$

Equation (48a) constitutes the desired rephrasing of the self consistent Maxwell equation (40)in terms both of a renormalized *coherent* admittance operator $Y_0 - V_{ce}$ and of stochastic source terms dependent on the statistical properties of $\hat{f}^a$ and on Z; the term V_{ce} is indicative of anomalous wave absorption. The average equation (48b) is used to determine the evolution of the background field E, whereas the $\tilde{E}$ fluctuation equation (48c) provides a means for deriving the kinetic equation for the wave fluctuations.

V. KINETIC EQUATIONS

Kinetic equations, which provide an average macroscopic description of a turbulent field, generally involve a coupling of certain measures of the background and fluctuation fields. In an electron plasma the background measures are provided by average fields $f(\mathbf{v}, \mathbf{r}, t)$, $\mathbf{E}(\mathbf{r}, t)$ which are normally weakly varying in space and time. On the other hand, the fluctuation fields $\tilde{f}(\mathbf{v}, \mathbf{r}, t)$, $\tilde{E}(\mathbf{r}, t)$ are rapidly varying stochastic fields that must be correlated to obtain the desired macroscopic measures. To this end it is first desirable to transform the fast $\mathbf{r}, t$ dependence of the fluctuation fields. Thus, for example, the electric field $\tilde{\mathbf{E}}$ is represented in terms of its Fourier–Laplace amplitude $\tilde{\mathbf{E}}_{k\omega}$ by[10]

$$\tilde{\mathbf{E}}(\mathbf{r}, t) = \sum_{k,\omega} \tilde{\mathbf{E}}_{k\omega} \exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)] \tag{50}$$

whence the "2-point autocorrelation" of the electric field may be spectrally represented as

$$\langle \tilde{\mathbf{E}}(\mathbf{r}, t)\tilde{\mathbf{E}}(\mathbf{r}', t') \rangle = \sum_{\mathbf{k},\omega} \mathcal{E}_{k\omega}(\mathbf{r}', t') \exp[i(\mathbf{k}\cdot\mathbf{r}'' - \omega t'')] \tag{51a}$$

with the (dyadic) "correlation spectrum" $\mathcal{E}_{k\omega}$ given by

$$\mathcal{E}_{\mathbf{k}\omega}(\mathbf{r}, t) = \sum_{\mathbf{k}'',\omega''} \langle \tilde{E}_{k\omega}\tilde{E}^*_{\mathbf{k}',\omega'} \rangle \exp[i(\mathbf{k}''\cdot\mathbf{r} - \omega'' t)] \tag{51b}$$

where $\mathbf{r}'' = \mathbf{r} - \mathbf{r}'$, $t'' = t - t'$ and $\mathbf{k}'' = \mathbf{k} - \mathbf{k}'$, $\omega'' = \omega - \omega'$ are fast space–time variables and slow "frequency" variables, respectively. The correlation spectrum $\mathcal{E}_{k\omega}(\mathbf{r}, t)$ generally varies slowly in space–time, and for given $\mathbf{k}$ possesses poles at complex $\omega = \omega_\beta(\mathbf{k})$; the latter reveal the possible collective oscillatory wave excitation and correlation damping of the plasma field. Accordingly, one integrates the trace $\mathcal{E}_{k\omega} = \mathcal{E}_{k\omega}$ over a suitable Laplace contour C and thereby obtains via residues

$$\int_C \mathcal{E}_{k\omega}(\mathbf{r}, t)\frac{d\omega}{2\pi} = \sum_\beta \mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t), \tag{52}$$

where the "kinetic" correlation spectrum $\mathcal{E}_\beta$ is characteristic of the βth collective wave type. The $\mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t)$ play the role of kinetic (phase space) distribution functions descriptive of β type quasiparticles of momentum $\mathbf{k}$ and position $\mathbf{r}$ and are analogous to the distribution functions $f_a(\mathbf{v}, \mathbf{r}, t)$ descriptive of a type particles of velocity $\mathbf{v}$ and position $\mathbf{r}$. It is to be noted from (51a) and (52) that the total wave (fluctuation) energy density at $\mathbf{r}, t$ is proportional to the scalar

$$\langle \tilde{\mathbf{E}}(\mathbf{r}, t)^2 \rangle = \sum_{\mathbf{k}}\sum_\beta \mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t). \tag{53}$$

A basic (initial value) problem in plasma turbulence specifies the particle distribution function $f(\mathbf{v}, \mathbf{r}, 0)$ and the correlation spectra $\mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, 0)$ at $t = 0$ and inquires as to their subsequent temporal evolution. At any time t the distribution $f(\mathbf{v}, \mathbf{r}, t)$ is determined by the "kinetic equation for particles" which is directly derivable from the formally exact abstract equation (37b). A corresponding "kinetic equation for waves," derivable after some manipulation from the abstract equation (48c) yields the evolution of $\mathcal{E}(\mathbf{k}, \mathbf{r}, t)$. To illustrate the derivation of these kinetic equations from the abstract solutions in Sec. IV we shall retain for an electron plasma only leading terms in the series representations of $\tilde{T}_c$, V_c and $\tilde{T}_{ce}$, V_{ce} in Eqs. (37) and (48). In obtaining (48) we have neglected terms in (40) involving higher than 2-point correlations; inclusion of the next order terms would require a self-consistent treatment of 3-point correlations including the initial conditions. This truncation approximation implies a limitation on the strong turbulence applicability of the results, but does include nonlinear and trapping effects beyond the quasilinear level.

To make explicit the derivation of the kinetic equation of waves from Eq. (48c), we first transform the fast r, t dependence of the latter into a $\mathbf{k}, \omega$ basis wherein it takes the form

$$\mathcal{Y}(\mathbf{k}, \omega)\cdot\tilde{\mathbf{E}}_{k\omega} = -\tilde{\mathbf{J}}^a_{k\omega} + \ldots \tag{54}$$

with the nonlinear plasma admittance $\mathcal{Y}$ given by

$$\mathcal{Y}(k, \omega) = \mathcal{G}(\mathbf{k}, \omega) - i\mathcal{B}(\mathbf{k}, \omega) = 1/\mathcal{Z}(\mathbf{k}, \omega)$$

$$\simeq -i\omega\epsilon_0\left[1 + \left(1 - \frac{k^2c^2}{\omega^2}\right)\mathbf{1}_T\right] - \frac{e^2}{m}\int \mathbf{v} g_{k\omega}(\mathbf{v}, \mathbf{v}')\nabla_{\mathbf{v}'} f^a(\mathbf{v}')d\mathbf{v}\,d\mathbf{v}' + \ldots$$

and

$$\mathcal{Z}(\mathbf{k}, \omega) = \mathcal{R}(\mathbf{k}, \omega) - i\mathcal{X}(\mathbf{k}, \omega),$$

$$-\tilde{J}^a_{\mathbf{k}\omega} = e\int \mathbf{v}\tilde{f}^a_{\mathbf{k}\omega}(\mathbf{v})\,d\mathbf{v}.$$

For simplicity, terms corresponding to $\tilde{V}_e\mathcal{Z}\langle\tilde{V}_e(\tilde{E}^a - Z\tilde{J}^a)\rangle$, etc., on the right side of (54), and the anomalous absorption term V_{ce} in $\mathcal{Y}(\mathbf{k}, \omega)$ have been omitted. For notational simplicity the weak $\mathbf{r}, t$ dependence of average quantities such as $f^a(\mathbf{v}, \mathbf{r}, t)$ and $g(\mathbf{v}, \mathbf{r}, t; \mathbf{v}', \mathbf{r}', t')$ has not been explicitly indicated. The quantity $\mathbf{1}_L = \mathbf{k}_0\mathbf{k}_0$ denotes the longitudinal unit dyadic, $\mathbf{1}_T = \mathbf{1} - \mathbf{1}_L$ is the transverse unit dyadic, and the unit vector $\mathbf{k}_0 = \mathbf{k}/|\mathbf{k}|$ denotes the propagation direction of the $\mathbf{k}, \omega$ wave.

From (54), one forms the autocorrelation relation

$$\langle\tilde{\mathbf{E}}_{\mathbf{k}\omega}\cdot\mathcal{Y}^*(\mathbf{k}', \omega')\cdot\tilde{\mathbf{E}}^*_{\mathbf{k}',\omega'}\rangle = -\langle\tilde{\mathbf{E}}_{\mathbf{k}\omega}\cdot\tilde{\mathbf{J}}^{a*}_{\mathbf{k}',\omega'}\rangle = \langle\tilde{\mathbf{J}}^{a*}_{\mathbf{k}',\omega'}\cdot\mathcal{Z}(\mathbf{k}, \omega)\cdot\tilde{\mathbf{J}}^a_{\mathbf{k}\omega}\rangle,$$

or, in an evident and more succinct dyadic notation,

$$\mathcal{Y}^*(\mathbf{k}'\omega'):\langle\tilde{\mathbf{E}}_{\mathbf{k}\omega}\tilde{\mathbf{E}}^*_{\mathbf{k}',\omega'}\rangle = \mathcal{Z}(\mathbf{k}, \omega):\langle\tilde{\mathbf{J}}^a_{k\omega}\tilde{J}^{a*}_{\mathbf{k}'\omega'}\rangle \tag{55}$$

where $\mathcal{Y}^*$ is the adjoint to $\mathcal{Y}$, i.e., the transposed conjugate dyadic. On inverse transformation of (55) with respect to the slow frequency variables $\mathbf{k}'', \omega''$, there is obtained via (51b)

$$\mathcal{Y}^*\left(\mathbf{k} - \frac{\nabla}{i}, \omega + \frac{1}{i}\frac{\partial}{\partial t}\right):\mathcal{E}_{k\omega}(\mathbf{r}, t) = \mathcal{Z}(\mathbf{k}, \omega):\mathcal{G}^a_{\mathbf{k}\omega}(\mathbf{r}, t), \tag{56}$$

where

$$\mathcal{G}^a_{\mathbf{k}\omega} = e^2\int \mathbf{v}\mathbf{v}'\mathcal{F}^a_{\mathbf{k}\omega}(\mathbf{v}, \mathbf{v}')d\mathbf{v}\,d\mathbf{v}',$$

$$\mathcal{F}^a_{\mathbf{k}\omega}(\mathbf{v}, \mathbf{v}') = \sum_{\mathbf{k}'',\omega''}\langle\tilde{f}^a_{\mathbf{k}\omega}(\mathbf{v})\tilde{f}^{a*}_{\mathbf{k}'\omega'}(\mathbf{v}')\rangle\exp[i(\mathbf{k}''\cdot\mathbf{r} - \omega''t)]. \tag{56a}$$

In (56a) the weak $\mathbf{r}, t$ dependence of the current correlation spectrum $\mathcal{G}^a_{\mathbf{k}\omega}$ and density correlation spectrum $\mathcal{F}^a_{\mathbf{k}\omega}$ is implicit.

For fixed $\mathbf{k}$ the plasma impedance $\mathcal{Z}(\mathbf{k}, \omega)$ exhibits both branch line and pole singularities in the complex ω-space. A two sheeted Riemann surface is required to display this ω dependence of the dyadic components of $\mathcal{Z}(\mathbf{k}, \omega)$; in the isotropic case, where $\mathcal{Z}$ is diagonal, only longitudinal and transverse components are relevant. Poles of $\mathcal{Z}(\mathbf{k}, \omega)$ occur at $\omega = \bar{\omega}_\beta(\mathbf{k}) = \omega_\beta - i\gamma_\beta$, corresponding to the zeros of $\det\mathcal{Y}(\mathbf{k}, \omega)$, and define the dispersion relations $\omega_\beta(\mathbf{k})$ and the damping (or growth) rates $\gamma_\beta(\mathbf{k})$ of the collective modes of the plasma. Poles (if any) lying on the so-called physical sheet [wherein $\mathcal{Z}(\mathbf{k}, \omega)$ vanishes suitably for infinite ω] yield the discrete (unstable) eigenmode spectrum. The poles on the unphysical sheet correspond to the so-called quasi-modes (the damped Landau modes of plasma theory) whose utility lies in the fact that they provide a rapidly convergent equivalent of the continuous eigenmode representation of the field.[11]

To determine $\mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t)$ defined in (52), it is first necessary to integrate $\mathcal{E}_{\mathbf{k}\omega}$ of (56) over a suitable ω-contour C. The singularities of $\mathcal{Z}(\mathbf{k}, \omega)$ and the disposition of the Laplace contour C on the relevant two sheeted Riemann surface are shown in Fig. 3. After bringing $\mathcal{Y}^*$ to the right side of (56), one integrates $\mathcal{E}_{\mathbf{k}\omega}$ over the contour C, which runs above the poles $\bar{\omega}_\beta$ in the physical sheet and is closed in the lower half of the unphysical ω-sheet by a semicircle C' at ∞. An implicit convergence factor $[\exp(-i\omega t'')$ with $t'' = 0+]$ assures vanishing of the contribution on the semicircle, whence, by residue theory each pole *within* the closed contour C yields via (52), the result

$$\mathcal{Y}^*\left(\mathbf{k} - \frac{\nabla}{i}, \bar{\omega}_\beta + \frac{1}{i}\frac{\partial}{\partial t}\right):\mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t) = \left(\frac{i\partial\mathcal{Y}(\mathbf{k}, \bar{\omega}_\beta)}{\partial\omega}\right)^{-1}:\mathcal{G}^a_{\mathbf{k}\omega_\beta}(\mathbf{r}, t). \tag{57}$$

In an isotropic electron plasma, wherein $\mathcal{Y} = Y_L\mathbf{1} + Y_T\mathbf{1}_T$ and the zeros of Y_L and Y_T are close to the real ω-axis, one finds, on noting the admittance decomposition in (54), for both Y_L and Y_T (and hence the subscript is omitted), that

$$Y(\mathbf{k}, \omega_\beta) \approx G(\mathbf{k}, \omega_\beta) \quad \text{since} \quad B(\mathbf{k}, \omega_\beta) \approx 0,$$

$$i\frac{\partial Y}{\partial\omega}(\mathbf{k}, \omega_\beta) \approx \frac{\partial B}{\partial\omega}(\mathbf{k}, \omega_\beta)$$

and

$$\gamma_\beta = G(\mathbf{k}, \omega_\beta)\Big/\frac{\partial B}{\partial\omega}(\mathbf{k}, \omega_\beta)$$

$$\nabla_{\mathbf{k}}\omega_\beta = -\nabla_{\mathbf{k}}B(\mathbf{k}, \omega_\beta)\Big/\frac{\partial B}{\partial\omega}(\mathbf{k}, \omega_\beta),$$

$$\nabla\omega_\beta = \nabla B(\mathbf{k}, \omega_\beta)\Big/\frac{\partial B}{\partial\omega}(\mathbf{k}, \omega_\beta), \tag{58}$$

where an implicit weak dependence of Y, and hence B, on $\mathbf{r}, t$ is to be remembered. In view of (58) one infers from (57) that, if $\mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t)$ (for either a longitudinal or transverse component) is weakly varying in space and time (neglecting nonlinear "parametric" terms),

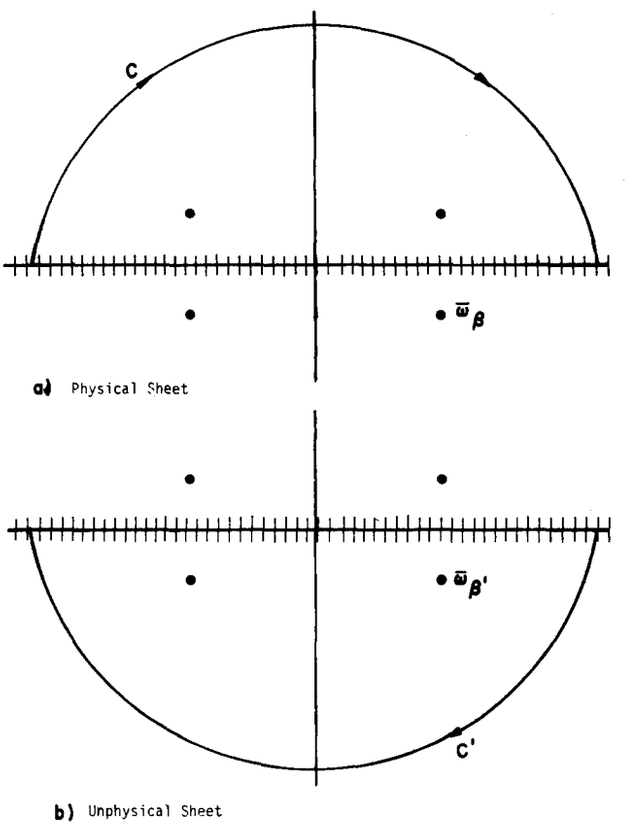


FIG. 3. Singularities of $\mathbf{Z}(\mathbf{k}, \omega)$: (a) physical sheet, (b) unphysical sheet.

$$\left(\frac{\partial}{\partial t}+\nabla_{\mathbf{k}}\omega_\beta\cdot\nabla-\nabla\omega_\beta\cdot\nabla_k\right)\mathcal{E}_\beta=2\gamma_\beta\mathcal{E}_\beta-\frac{\mathcal{J}^a_{\mathbf{k}\omega_\beta}}{[\partial B(\mathbf{k},\omega_\beta)/\partial\omega]^2} \tag{59}$$

which defines the desired kinetic equation of waves for the kinetic correlation spectrum $\mathcal{E}_\beta(\mathbf{k}, \mathbf{r}, t)$ of the βth collective wave type.

The kinetic equation for particles follows from the abstract Eq. (47b) on explicit representation of both the operator V_c and the expression $\langle\tilde{T}_c\tilde{f}^a\rangle$ in terms of autocorrelations of $\tilde{E}$ and $\tilde{f}^a$. In (47b) V_c operates on a weakly varying function, whence it is representable via (30b) in the form

$$V_c\to\nabla_{\mathbf{v}}\cdot\mathcal{D}(\mathbf{v},\mathbf{r},t)\cdot\nabla_{\mathbf{v}} \tag{60}$$

with, if only the leading term in (30b) is retained, the velocity diffusion coefficient $\mathcal{D}$ given by

$$\mathcal{D}(\mathbf{v},\mathbf{r},t)\simeq\frac{e^2}{m^2}\int g(\mathbf{v},\mathbf{r},t;\mathbf{v}',\mathbf{r}'t')\langle\tilde{\tilde{\mathbf{E}}}(\mathbf{r},t)\tilde{\tilde{\mathbf{E}}}(\mathbf{r}',t')\rangle d\mathbf{v}'d\mathbf{r}'dt'. \tag{60a}$$

In a $\mathbf{k}, \omega$ basis the integral in (60a) can be expressed as

$$\mathcal{D}(\mathbf{v},\mathbf{r},t)=\frac{e^2}{2m^2}\sum_{\mathbf{k},\omega}\int[g_{\mathbf{k},\omega}(\mathbf{v},\mathbf{v}')+g^*_{\mathbf{K},\Omega}(\mathbf{v},\mathbf{v}')]\mathcal{E}_{\mathbf{k}\omega}(\mathbf{r},t)d\mathbf{v}' \tag{60b}$$

where $\mathbf{K}=\mathbf{k}-\nabla/i$ and $\Omega=\omega+(1/i)(\partial/\partial t)$ denote operators, and where despite the weak $\mathbf{r}, t$ dependence of $g_{\mathbf{k}\omega}$, the space and time derivatives ∇, $\partial/\partial t$ are assumed to act only on $\mathcal{E}_{\mathbf{k}\omega}(\mathbf{r}, t)$. For the steady state, wherein $\mathcal{E}_{\mathbf{k}\omega}(\mathbf{r}, t)=\mathcal{E}_{\mathbf{k}\omega}$ is independent of $\mathbf{r}, t$, Eq. (60b) simplifies to the familiar form

$$\mathcal{D}(\mathbf{v})=\frac{e^2}{m^2}\sum_{\mathbf{k},\omega}\int\mathrm{Re}\,g_{\mathbf{k}\omega}(\mathbf{v},\mathbf{v}')\mathcal{E}_{\mathbf{k}\omega}d\mathbf{v}'. \tag{60c}$$

Similarly, from (30a) the abstract expression $\langle\tilde{T}_c\tilde{f}^a\rangle$ can be exactly represented as

$$\langle\tilde{T}_c\tilde{f}^a\rangle\to\nabla_v\cdot\mathbf{A}(\mathbf{v},\mathbf{r},t), \tag{61}$$

where, if only the leading term in (30a) is retained,

$$\mathbf{A}(\mathbf{v},\mathbf{r},t)\simeq\frac{e}{m}\langle\tilde{\tilde{\mathbf{E}}}(\mathbf{r},t)\tilde{f}^a(\mathbf{v},\mathbf{r},t)\rangle. \tag{61a}$$

In a $\mathbf{k}, \omega$ basis (61a) becomes

$$\mathbf{A}(\mathbf{v},\mathbf{r},t)=\frac{e}{2m}\sum_{\substack{\mathbf{k},\omega\\ \mathbf{k}'',\omega''}}\int\langle\tilde{\mathbf{E}}_{\mathbf{k}\omega}\tilde{f}^{a*}_{\mathbf{k}''\omega''}(\mathbf{v})+\tilde{\mathbf{E}}^*_{\mathbf{k}''\omega''}\tilde{f}^a_{\mathbf{k}\omega}(\mathbf{v})\rangle \times\exp[i(\mathbf{k}''\cdot\mathbf{r}-\omega''t)]$$

whence, using the approximation (54) and inferring therefrom that

$$\tilde{\tilde{\mathbf{E}}}_{\mathbf{k}\omega}=\tilde{\tilde{\mathbf{E}}}^a_{\mathbf{k}\omega}-\mathcal{Z}(\mathbf{k},\omega)\cdot\tilde{\mathbf{J}}^a_{\mathbf{k}\omega}$$

with $\mathcal{Y}(\mathbf{k},\omega)\cdot\tilde{\tilde{\mathbf{E}}}^a_{\mathbf{k}\omega}=0$ and $-\tilde{\mathbf{J}}^a_{\mathbf{k}\omega}=e\int\mathbf{v}'\tilde{f}^a_{\mathbf{k}\omega}(\mathbf{v}')d\mathbf{v}'$, one expresses $\mathbf{A}$ as

$$\mathbf{A}(\mathbf{v},\mathbf{r},t)=\frac{e^2}{2m}\sum_{\mathbf{k},\omega}\left[\mathcal{Z}(\mathbf{k},\omega)+\mathcal{Z}^*\left(\mathbf{k}-\frac{\nabla}{i},\omega+\frac{1}{i}\frac{\partial}{\partial t}\right)\right]\cdot\cdot\int\mathbf{v}'\mathcal{J}^a_{\mathbf{k}\omega}(\mathbf{v},\mathbf{v}')d\mathbf{v}', \tag{61b}$$

provided the "incident field" $\tilde{\tilde{\mathbf{E}}}^a_{\mathbf{k}\omega}$ and $\tilde{f}^a_{\mathbf{k}\omega}$, are assumed to be uncorrelated, a restriction that can be readily removed.

The space–time derivatives ∇, $\partial/\partial t$ are assumed in general to operate upon the weak implicit $\mathbf{r}, t$ dependence of the correlation spectrum

$$\mathcal{J}^a_{\mathbf{k}\omega}(\mathbf{v},\mathbf{v}')=\sum_{\mathbf{k}''\omega''}\langle\tilde{f}_{\mathbf{k}\omega}(\mathbf{v})\tilde{f}^{a*}_{\mathbf{k}'\omega'}(\mathbf{v}')\rangle\exp[i(\mathbf{k}''\cdot\mathbf{r}-\omega''t)=\mathcal{J}^a_{\mathbf{k}\omega}(\mathbf{v}'\mathbf{v}). \tag{62}$$

In the steady state where $\mathcal{J}^a_{\mathbf{k}\omega}$ is independent of $\mathbf{r}$ and t, (61b) becomes

$$\mathbf{A}(\mathbf{v})=\frac{e^2}{m}\sum_{\mathbf{k},\omega}\mathcal{R}(\mathbf{k},\omega)\cdot\int\mathbf{v}'\mathcal{J}^a_{\mathbf{k}\omega}(\mathbf{v},\mathbf{v}')d\mathbf{v}' \tag{61c}$$

where $\mathcal{R}(\mathbf{k}, \omega)$ is the plasma resistance dyadic defined in connection with Eq. (54).

With determination in (60), (61) of $\mathcal{D}$ and $\mathbf{A}$ the particle kinetic equation (47b) assumes the form

$$\left(\frac{\partial}{\partial t}+\mathbf{v}\cdot\nabla-\frac{e}{m}\mathbf{E}\cdot\nabla_v-\nabla_v\cdot\mathcal{D}\cdot\nabla_v\right)f(\mathbf{v},\mathbf{r},t)=\nabla_v\cdot\mathbf{A}. \tag{63}$$

Because of the source term on the right, (63) differs from the conventional Fokker–Planck equation frequently employed to define the background particle distribution; only in the near equilibrium case wherein $\mathbf{A}=\mathcal{A}.f$ can (63) be identified with the Fokker–Planck equation.

The kinetic equations (59) and (63) provide the desired, albeit implicit, description of the evolution of the background and fluctuation properties of a simple turbulent plasma. Their use is predicated upon a knowledge of $\bar{\omega}_\beta=\omega_\beta-i\gamma_\beta$, $\mathcal{J}_{\mathbf{k}\omega_\beta}$, $\partial B(\mathbf{k},\omega_\beta)/\partial\omega$, and $\mathcal{D}$, $\mathbf{A}$, which in turn demand a knowledge of $\mathcal{J}_{\mathbf{k}\omega}$, $\mathcal{Z}(\mathbf{k},\omega)$, $\mathcal{J}^a_{\mathbf{k}\omega}$, and these in turn depend on the desired kinetic fields f and $\mathcal{E}_\beta$ in (59) and (63). This tailchasing difficulty reflects the nonlinear nature of plasma turbulence and appears to be resolvable via numerical methods. Computer runs are providing interesting insights into the solution of these kinetic equations both in the transient and steady state.

VI. CONCLUSION

The perturbation formalism discussed above and employed to derive the kinetic equation for particles (63) and for waves (59) has the virtue of explicitly admitting higher order nonlinear corrections to the transport coefficients $\mathcal{D}$ and $\mathbf{A}$ as well as to the collective wave frequency ω_β and growth (or damping) rate γ_β. Although particle trapping is partially taken into account through the nonlinear dependence of the plasma admittance $\mathcal{Y}$ on the field correlation $\mathcal{E}$, the rapidity of convergence of the representation of this dependence has to be explored. The basic result derived herein is contained in Eq. (37). Its use in deriving kinetic equations is illustrated in Sec. V, wherein for an electron plasma some nonlinear wave–wave (parametric) and wave–particle (damping) effects have been omitted for simplicity.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Contract N00014-67-A-0467-0018.

[1]T. Dupree, Phys. Fluids 9, 1773 (1966); 15, 334 (1972); R.H. Kraichnan, J. Math. Phys. 2, 124 (1961); B.B. Kadomtsev, *Plasma Turbulence* (Academic, London, 1965); V.N. Tsytovich, *An Introduction to the Theory of Plasma Turbulence* (Pergamon, New York, 1972); R.C. Davidson, *Methods in Nonlinear Plasma Theory* (Academic, New York, 1972), Part II; J.J. Thomson and G. Benford, J. Math. Phys. 14, 531 (1973); Phys. Fluids 16, 1505 (1973); K. Nishikawa, Prog. Theor. Phys. 35, 193 (1966); Y.H. Ichikawa, M. Watabe, K. Yarmoda, Prog. Theor. Phys. 31, 153 (1964); 34, 383 (1965).

[2]V.I. Tatarskii and M.E. Gertenshtein, Sov. Phys.-JETP 17, 458 (1963); K. Furutsu, J. Optical Soc. Amer. 62, 240 (1972); V. Frisch, "Wave Propagation in Random Media," in *Probabilistic Methods in Applied Mathematics*, edited by A.T. Bharucha-Reid (Academic, New York, 1968), Vol. I: H. Bremmer, Radio Science 8, 511 (1973); G. Papanicolau and J. Keller, SIAM J. Appl. Math. 21, 287 (1971).

[3]I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience, New York, 1961); R. Balescu, Physica 38, 98 (1968); R. Kubo, J. Phys. Soc. Japan 17, 1100 (1962).

[4]Strictly stated (1) requires a Lorentz force with a magentic field term, subtraction of the self force, etc.

[5]$L_0, \tilde{V}$ are derivative (diagonal) operators whereas $\hat{G}$ is an integral operator. Thus, in $\mathbf{v}, \mathbf{r}, t$ space $\hat{G}\hat{f} \rightarrow \int \hat{g}(\mathbf{v}, \mathbf{r}, t; \mathbf{v}', \mathbf{r}', t') \times \hat{f}(\mathbf{v}', \mathbf{r}', t') d\mathbf{v}' d\mathbf{r}' dt'$ while $(L_0 - \tilde{V})\hat{f}$ is represented as in (1).

[6]R. Newton, *Scattering Theory of Waves and Particles*, Chap. 7 (McGraw-Hill, New York, 1966).

[7]Adjoint operators $L_0^\dagger, \tilde{V}^+, \hat{G}^+$ are implied by the relation $(L_0^\dagger - \tilde{V}^+)\hat{G}^+ = \hat{G}(L_0 - \tilde{V})$. The latter form is more convenient in the subsequent discussion. Also note $L_0^\dagger G_0^\dagger = G_0 L_0$.

[8]F.L. Hinton and C. Oberman, Phys. Fluids 11, 1982 (1968); J. Weinstock, Phys. Fluids 11, 1977 (1968).

[9]For the simple case of "quiet" initial excitation wherein $\tilde{f}^a = 0$, one has $f = f^a$, whence $\hat{f} = f + G\tilde{T}_c f$.

[10]For a representation with continuous wavenumber k and frequency ω, one employs the mode summation

$$\sum_{\mathbf{k},\omega} \rightarrow \int\int\int\int \frac{d\mathbf{k}\, d\omega}{(2\pi)^4}$$

with the three-dimensional k integration extended over the entire real k-space and with the ω contour extending from $-\infty$ to $+\infty$ *above* all singularities of the integrand. For a representation in a bounded volume V with discrete wavenumbers $\mathbf{k}_i$

$$\sum_{\mathbf{k}\omega} \rightarrow \frac{1}{V} \sum_{\mathbf{k}_i} \int_{-\infty+ia}^{+\infty+ia} \frac{d\omega}{2\pi}$$

with $a > 0$.

[11]N. Marcuvitz, IEEE Trans. on Electron Devices ED-17, No. 3, 252 (1970); also, for a preliminary account of the material in Secs. I—III, cf. *Electromagnetic Wave Theory* (Science Press, Moscow, 1971), p. 823—International Symposium Proceedings.

Distribution of the lowest eigenvalue for a class of random matrices

Nazakat Ullah

Tata Institute of Fundamental Research, Bombay 5, India
(Received 28 June 1972)

The random matrix theory is applied to the ensemble of partially diagonalized shell-model Hamiltonian matrices. Using the analogy with the distribution of the poles of the unitary collision matrix, an expression for the joint probability density function of the eigenvalues is derived for the present ensemble. An approximation is used to work out a closed form expression for the probability density function of the lowest eigenvalue. Two simple examples are considered to see what this probability density function looks like.

I. INTRODUCTION

In many of the nuclear spectroscopy calculations one uses a shell-model Hamiltonian which operates in a fairly large space, called the spectroscopic space. This is obtained by distributing a given number of nucleons among some set of single-particle states. It is obvious that the dimensions of such a space will increase quite rapidly as the number of nucleons and the number of available single-particle states increase. Since it is difficult to handle such large spectroscopic spaces exactly one very often uses the approximation of truncation of the given space. Even if one has devised some simple procedure for such a truncation, one would still like to see whether there is some way of knowing how large are the effects of truncation.

We know that random matrix theory[1] has been used quite successfully in the past to study the distribution of the widths and the positions of the compound nucleus resonances where it was difficult to know the detailed knowledge of the matrix elements of the compound nucleus Hamiltonian which are very many in number. We would like to see if these ideas could also be used in nuclear spectroscopy calculations. Unlike the compound nucleus our knowledge of the shell model tells us that we can subdivide the spectroscopic space into smaller subspaces. These subspaces can be conveniently defined, e.g., by specifying the number of nucleons in various single-particle states. Since the dimensions of these subspaces are small, the shell-model Hamiltonian can be diagonalized in them. We can then introduce the statistical methods to handle the interactions between various subspaces.

Earlier we had used[2] the statistical ideas to derive an exact expression in terms of the centroids and the partial width[3] for the mixing of two approximate symmetries of the Hamiltonian. The purpose of the present work is to consider a Hamiltonian matrix with two subspaces and to derive the joint distribution of its eigenvalues. This joint distribution together with an approximation will be used to find a closed-form expression for the distribution of its lowest eigenvalue. A plot of this distribution is shown for two simple examples considered in Sec. IV.

II. JOINT DISTRIBUTION OF THE EIGENVALUES

Let us consider an ensemble of Hamiltonian matrices, the matrix elements of which are given by

$$H_{\mu\nu}=\delta_{\mu\nu}\lambda_\mu+(1-\delta_{\mu\nu})(\delta_{\mu 1}H_{1\nu}+\delta_{\nu 1}H_{1\mu}), \tag{1}$$

where $\lambda_\mu(\mu=1,2,\ldots,N)$ are fixed diagonal elements and $H_{1\nu}(\nu\neq 1)$ are off-diagonal elements. We assume that each of these off-diagonal elements has a Gaussian distribution with the same dispersion σ. We would like to derive the joint distribution of the eigenvalues E_μ of this matrix.

To do this we write the following identity in E which connects the eigenvalues E_μ with the elements λ_μ and $H_{1\mu}$.

$$\prod_{\mu=1}^{N}(E-E_\mu)\equiv\prod_{\mu=1}^{N}(E-\lambda_\mu)-\sum_{\nu=2}^{N}H_{1\nu}^2\prod_{\mu\neq 1,\nu}(E-\lambda_\mu). \tag{2}$$

By equating various powers of E in the above identity, we obtain the following relations:

$$\sum_{\mu=1}^{N}E_\mu=\sum_{\mu=1}^{N}\lambda_\mu, \tag{3a}$$

$$\sum_{\mu<\nu}^{N}E_\mu E_\nu=\sum_{\mu<\nu}\lambda_\mu\lambda_\nu-\sum_{\mu=2}^{N}H_{1\mu}^2, \tag{3b}$$

$$\vdots$$

$$\prod_{\mu=1}^{N}E_\mu=\prod_{\mu=1}^{N}\lambda_\mu-\sum_{\nu=2}^{N}H_{1\nu}^2\Big(\prod_{\mu\neq 1,\nu}\lambda_\mu\Big). \tag{3c}$$

Relations (3) tell us that the problem of finding the joint distribution of E_μ is similar to that of the distribution of the poles of the unitary collision matrix.[4] As in Ref. 4 we first introduce a new set of variables defined by

$$v_1=\sum_{\mu<\nu}(\lambda_\mu\lambda_\nu-E_\mu E_\nu), \tag{4a}$$

$$v_2=\sum(\lambda_\mu\lambda_\nu\lambda_\alpha-E_\mu E_\nu E_\alpha), \tag{4b}$$

$$\vdots$$

$$v_{N-1}=\prod_{\mu=1}^{N}\lambda_\mu-\prod_{\mu=1}^{N}E_\mu, \tag{4c}$$

and derive an expression for their joint distribution. The indices μ, ν, α in relation (4b) are all different.

From relations (3) and (4) we can write the following matrix relation

$$\begin{pmatrix} 1 & 1 & 1 \\ \sum_{\alpha\neq 1,2}\lambda\alpha & \sum_{\alpha\neq 1,3}\lambda\alpha & \sum_{\alpha\neq 1,N}\lambda\alpha \\ \vdots & \vdots & \vdots \\ \prod_{\alpha\neq 1,2}\lambda\alpha & \prod_{\alpha\neq 1,3}\lambda\alpha & \prod_{\alpha\neq 1,N}\lambda\alpha \end{pmatrix}\begin{pmatrix} H_{12}^2 \\ H_{13}^2 \\ \vdots \\ H_{1N}^2 \end{pmatrix}=\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{N-1} \end{pmatrix}, \tag{5}$$

between the variables v_μ and the squares of the off-diagonal elements $H_{1\mu}^2$. To find the distribution of the variables v_μ, we assume the probability density function of the variables $H_{1\mu}$ to be

$$P(H_{12}, H_{13}, \ldots, H_{1N}) \prod_{\mu=2}^{N} dH_{1\mu} = K\left[\exp\left(-\frac{1}{2\sigma^2}\sum_{\mu=2}^{N} H_{1\mu}^2\right)\right]\prod_{\mu=2}^{N} dH_{1\mu}, \tag{6}$$

where K is the normalization constant. We now use the well-known method[5] of finding the distribution of one set of variables which are related to another set by a transformation given by expressions (5). This gives us the joint probability density function of the variables v_μ as

$$P(v_1, v_2, \ldots, v_{N-1}) \prod_{\mu=1}^{n-1} dv\mu = K\left(\left|\prod_{\nu=2}^{N}\sum_{\alpha=1}^{N} M_{\nu\alpha}^{-1} v_\alpha\right|\right)^{-1/2} \times\left[\exp\left(-\frac{v_1}{2\sigma^2}\right)\right]\prod_{\mu=1}^{N-1} dv_\mu, \tag{7}$$

where K is the new normalization constant and M^{-1} is the inverse of the matrix in expression (5). It is given by

$$M^{-1} = \begin{pmatrix} \frac{\lambda_2^{N-2}}{\prod_{\mu\neq1,2}(\lambda_2-\lambda_\mu)} & -\frac{\lambda_2^{N-3}}{\prod_{\mu\neq1,2}(\lambda_2-\lambda_\mu)} & \cdots & \frac{(-1)^N}{\prod_{\mu\neq1,2}(\lambda_2-\lambda_\mu)} \\ \vdots & \vdots & & \vdots \\ \frac{\lambda_N^{N-2}}{\prod_{\mu\neq1,N}(\lambda_N-\lambda_\mu)} & -\frac{\lambda_N^{N-3}}{\prod_{\mu\neq1,N}(\lambda_N-\lambda_\mu)} & \cdots & \frac{(-1)^N}{\prod_{\mu\neq1,N}(\lambda_N-\lambda_\mu)} \end{pmatrix}. \tag{8}$$

In writing the inverse given by relation (8), we have assumed that $\lambda_\mu(\mu=2,3,\ldots,N)$ are all different. The case where some of them are equal will be given later.

The next step is to transform the probability density function from the set of the variables v_μ to the desired set E_μ. We use relations (3a), (4), (7), and (8) to arrive at the following joint probability density function for the eigenvalues E_μ:

$$P(E_1, E_2, \ldots, E_N) \prod_{\mu=1}^{N} dE_\mu = K\delta\left[\sum_{\mu=1}^{N}(E_\mu-\lambda_\mu)\right] \times\left[\exp-\left(\frac{1}{4\sigma^2}\sum_{\mu=1}^{N} E_\mu^2\right)\right]\left(\prod_{\mu<\nu}|E_\mu-E_\nu|\right) \times\left(\left|\prod_{\nu=2}^{N}\prod_{\mu=1}^{N}(\lambda_\nu-E_\mu)\right|\right)^{-1/2}\prod_{\mu=1}^{N} dE_\mu. \tag{9}$$

Before we conclude this section, we also consider the case when some of $\lambda_\mu(\mu=2,\ldots,N)$ are equal. We see from the structure of the matrix elements given by expression (1) that for such a case some of the eigenvalues E_μ are independent of $H_{1\mu}$ and are equal to those λ_μ which are equal. All one has to do then is to exclude these values of μ in the sums and products in expression (9) and add additional δ functions to take this fact into account.

III. LOWEST EIGENVALUE DISTRIBUTION

We would first like to consider the case for $N=2$. It is a straightforward calculation to obtain the probability density function of the lowest eigenvalue E from expression (9) by integrating out E_2 from E_1 to ∞. It is given by

$$P(E)dE = (2\pi\sigma^2)^{-1/2}[(\lambda_1-E)(\lambda_2-E)]^{-1/2}[\lambda_1+\lambda_2-2E] \times\exp\left(-\frac{1}{2\sigma^2}(\lambda_1-E)(\lambda_2-E)\right)dE, \tag{10}$$

where the range of E is determined from the condition that $(\lambda_1-E)(\lambda_2-E)\geq 0$.

Assuming $\lambda_1<\lambda_2$, it follows from expression (10) that the average and the mean-square deviation of the lowest eigenvalue E are given by

$$\langle E\rangle = \tfrac{1}{2}(\lambda_1+\lambda_2) - (8\sqrt{2\pi}\sigma)^{-1}(\lambda_1-\lambda_2)^2[\exp(16\sigma^2)^{-1}(\lambda_1-\lambda_2)^2] \times[K_0((16\sigma^2)^{-1}(\lambda_1-\lambda_2)^2)+K_1((16\sigma^2)^{-1}(\lambda_1-\lambda_2)^2)], \tag{11}$$

$$\langle E^2\rangle - \langle E\rangle^2 = \tfrac{1}{4}(\lambda_1-\lambda_2)^2+\sigma^2-(128\pi\sigma^2)^{-1}(\lambda_1-\lambda_2)^4 \times[\exp(8\sigma^2)^{-1}(\lambda_1-\lambda_2)^2][K_0((16\sigma^2)^{-1}(\lambda_1-\lambda_2)^2) +K_1((16\sigma^2)^{-1}(\lambda_1-\lambda_2)^2)]^2, \tag{12}$$

where K_0, K_1 are the modified Bessel functions.[6]

We next consider the N-dimensional case. We have not been able to work out a closed form expression for the distribution of the lowest eigenvalue in general. To make further progress we introduce the approximation of replacing each $\lambda_\mu(\mu=2,\ldots,N)$ by their centroid.[3] We then find that the probability density function of the lowest eigenvalue E is given by

$$P(E)dE = C_N[\lambda_1+\lambda_2-2E][(\lambda_1-E)(\lambda_2-E)]^{(N-3)/2} \times[\exp-(2\sigma^2)^{-1}(\lambda_1-E)(\lambda_2-E)]dE, \tag{13}$$

where λ_2 is the centroid of $\lambda_\mu(\mu=2,\ldots,N)$ and the range of E is again determined from the condition $(\lambda_1-E)(\lambda_2-E)\geq 0$. The normalization constant C_N is given by

$$C_N = (2\sigma^2)^{-(N-1)/2}[\Gamma(\tfrac{1}{2}(N-1))]^{-1}. \tag{14}$$

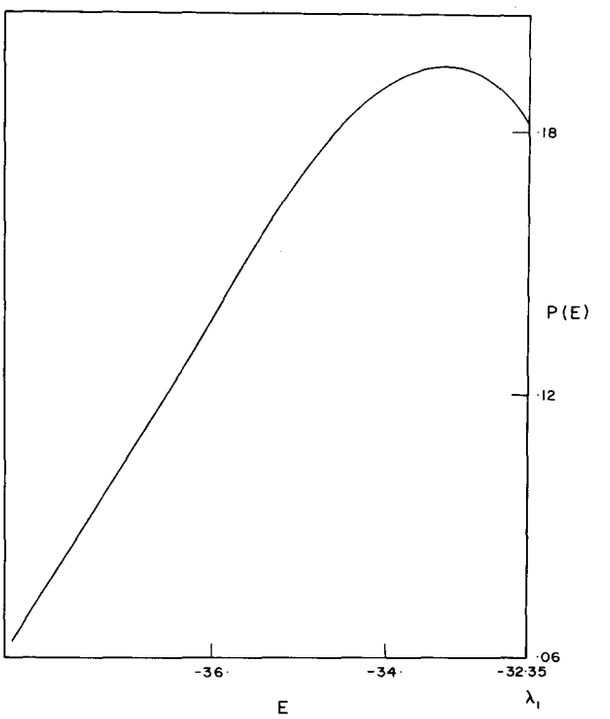


FIG. 1. Plots of probability density function for the lowest eigenvalue given by expression (13) for various values of λ_1, λ_2, σ^2, N. $\lambda_1=-32.35$ MeV, $\lambda_2=-26.17$ MeV, $\sigma^2=16.93$ M (MeV)^2, $N=3$.

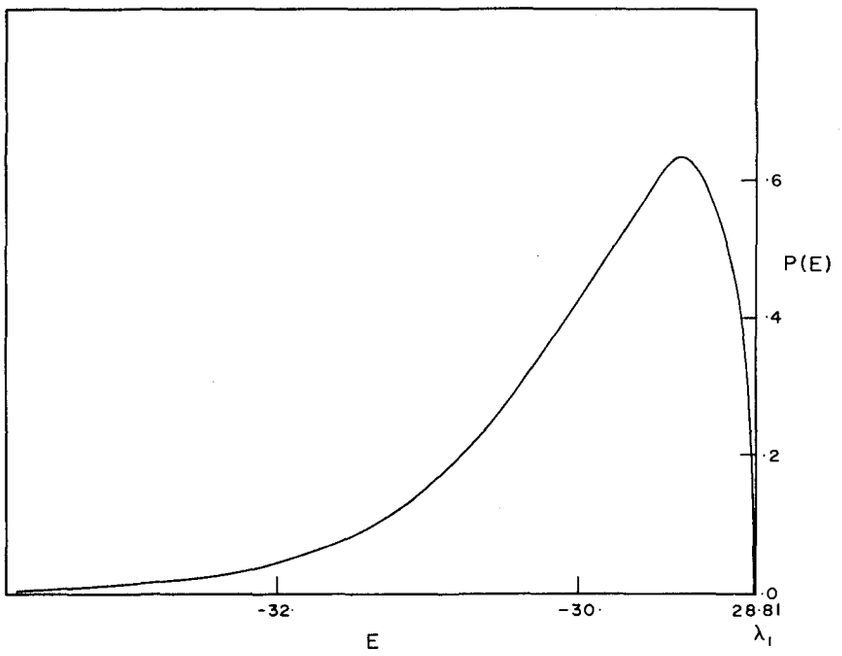


FIG. 2. Plots of probability density function for the lowest eigenvalue given by expression (13) for various values of λ_1, λ_2, σ^2, N. $\lambda_1 = -28.81$ MeV, $\lambda_2 = -17.31$ MeV, $\sigma^2 = 4.88$ (MeV)^2, $N = 4$.

For the distribution (13), the average value and the mean-square deviation can be determined from the following relations:

$$\langle [E - \tfrac{1}{2}(\lambda_1 + \lambda_2)] \rangle = -(8\sqrt{2}\sigma)^{-1}(\lambda_1 - \lambda_2)^2 [\Gamma(\tfrac{1}{2}(2m+1))]^{-1} \times \left[-\left(\frac{\partial}{\partial \alpha}\right)^m (\exp \tfrac{1}{2}\xi\alpha)[K_0(\tfrac{1}{2}\xi\alpha) + K_1(\tfrac{1}{2}\xi\alpha)] \right]_{\alpha=1}, \quad N = 2m+2, \tag{15a}$$

$$\langle [E - \tfrac{1}{2}(\lambda_1 + \lambda_2)] \rangle = -\sqrt{2}\,\sigma(m!)^{-1} \left[\left(-\frac{\partial}{\partial \alpha} \right)^m \alpha^{-1} \times [\sqrt{\xi} + \tfrac{1}{2}\sqrt{\pi}/\alpha\, (\exp \alpha\xi)\operatorname{erfc}(\sqrt{\alpha\xi})] \right]_{\alpha=1}, \quad N = 2m+3, \tag{15b}$$

$$\langle [E - \tfrac{1}{2}(\lambda_1 + \lambda_2)]^2 \rangle = \tfrac{1}{4}(\lambda_1 - \lambda_2)^2 + (n-1)\sigma^2, \tag{15c}$$

where $\xi = (8\sigma^2)^{-1}(\lambda_1 - \lambda_2)^2$,

and erfc is the complementary error function.[6]

For large values of N, it is useful to derive the following expression for the average value of E:

$$\langle [E - \tfrac{1}{2}(\lambda_1 + \lambda_2)] \rangle = -\tfrac{1}{2}|\lambda_1 - \lambda_2| \{ \tfrac{1}{12}(N-1)(\xi[\xi + \tfrac{1}{2}(N+1)])^{-1/2} + \xi^{1/2}[\xi + \tfrac{1}{2}(N-1)]^{-1/2} \}. \tag{15d}$$

For given values of λ_1, λ_2, σ^2, and N, expressions (15) can be used to calculate the quantity $\langle E \rangle \pm (\langle E^2 \rangle - \langle E \rangle^2)^{1/2}$, which gives us the most probable range of the lowest eigenvalue. It is this quantity which should give us some idea about the possible correction[7] which has to be applied to the lowest eigenvalue when we enlarge our basis set of states.

IV. ILLUSTRATIVE EXAMPLES

To see how the probability density function of the lowest eigenvalue E given be expression (13) looks like, we choose the nucleus ^{16}O with ^{12}C as a core.

As a first example, we consider its ground state which has the angular momentum and parity $J^\pi = 0^+$. Since ^{12}C is assumed to be a core, we are left with four active nucleons which we assume to move in the single-particle orbits $|p_{1/2}$ and $|d_{5/2}$. As usual, we take the single-particle energies to be $\epsilon_{p_{1/2}} = -4.95$ MeV, $\epsilon_{d_{5/2}} = -3.3$ MeV and the interaction between the nucleons to be a 40-meV Rosenfeld interaction. Had there been no configuration mixing, all the four nucleons have occupied the $p_{1/2}$ level. Because of the two-body interaction the $(p_{1/2})^4$ configuration will mix with the various particle–hole configurations $(p_{1/2})^{4-2n}(d_{5/2})^{2n}$. On calculating the values of λ_1, λ_2, σ^2, and N for the present example, we find that they are given by:

$\lambda_1 = -32.35$ MeV, $\lambda_2 = -26.17$ MeV, $\sigma^2 = 16.93$ (MeV)^2, $N = 3$.

We note that λ_2 is the centroid of the $J^\pi = 0^+$ states belonging to two-particle–two-hole configuration. The plot of $P(E)$ versus E is shown in Fig. 1. Expressions (15) give us the most probable range of the lowest eigenvalue to be from -33.17 to -37.72 MeV.

In the second example we consider the excited state $J^\pi = 3^-$ in ^{16}O. λ_1 is now the diagonal matrix element of the Hamiltonian with respect to the single particle–hole configuration $(p_{1/2})^{-1}(d_{5/2})$, which interacts with the three-particle–three-hole configuration $(p_{1/2})^{-3}(d_{5/2})^3$. The values of λ_1, λ_2, σ^2, N are now given by

$\lambda_1 = -28.81$ MeV, $\lambda_2 = -17.31$ MeV, $\sigma^2 = 4.88$ (MeV)^2, $N = 4$.

The probability density function for the lowest eigenvalue is shown in Fig. 2. The most probable range in this case is from -29.08 to -30.76 MeV.

Had there been no configuration mixing, then, it is obvious that in both the examples the mean-square deviation $\langle E^2 \rangle - \langle E \rangle^2$ would have been zero. Therefore, as we had remarked earlier, the most probable range gives us some idea as to how large are the effects of configuration mixing on the lowest eigenvalue.

ACKNOWLEDGMENT

The author would like to thank Professor Virendra Singh for interesting discussions and helpful comments on the present work.

[1]C.E. Porter, *Statistical Theories of Spectra: Fluctuations* (Academic, New York, 1965).
[2]N. Ullah, Nucl. Phys. A **164**, 658 (1971).
[3]J.B. French and K.F. Ratcliff, Phys. Rev. C **13**, 94 (1971).
[4]N. Ullah, J. Math. Phys. **10**, 2099 (1969).
[5]T.W. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1958), p. 10.
[6]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I.A. Stegun (Dover, New York, 1965), Chaps. 7, 9.
[7]F.S. Chang, Ph.D. thesis (University of Rochester, 1970); F.S. Chang *et al.*, Ann. Phys. (N.Y.) **66**, 137 (1971).