JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Relationships among Generalized Phase-Space Distributions*

G. C. SUMMERFIELD† AND P. F. ZWEIFEL†
Department of Nuclear Engineering, The University of Michigan, Ann Arbor

(Received 22 May 1968)

The generalized phase-space distributions, including the Wigner distribution, are presented in terms of expected values of generating operators. A generalization of the Weyl correspondence is obtained to provide expressions for generalized Wigner equivalents. Finally, rather simple relationships are obtained connecting the generalized phase-space distributions to the Wigner distribution; and similar relationships are obtained for the generalized Wigner equivalents. In particular, it appears that, among the class considered, there is no reason to use any distribution other than the Wigner for performing any calculations.

I. INTRODUCTION

In 1932, Wigner[1] introduced a method of performing quantum-mechanical ensemble averages in terms of phase-space integrations over c-number variables. Since that time, a number of extensions, modifications, discussions, derivations, applications, etc., have appeared in the literature. We refer the reader to a review[2] in which further references can be found.

Actually, there exist an infinite number of quasi-distribution functions which can be used for the same purpose as the Wigner distribution function. In a recent paper,[3] Cohen described one method for generating such distributions, and showed how the Wigner function, the so-called "symmetric" function, and the Born–Jordan function could be generated. He also obtained equations of motion (quantum Liouville equations) for these distribution functions.

In the present paper we present a particularly simple and elegant manner for generating an infinite class of distribution functions which include, as special cases, the Wigner, symmetric, and Born–Jordan functions. We also show that all of these various distributions can be obtained from the Wigner distribution by a rather trivial transformation.

For the purposes of our later discussion, it is convenient for us to point out several general properties that all of these distributions have in common.

We represent the $6N$-dimensional phase space by the $3N$-dimensional momentum and position vectors r and p. A generalized phase-space distribution is a function of the variables r and p and time, $f(r, p, t)$. These functions satisfy the following conditions:

(A) *Classical Limit:* The function

$$f_c(r, p, t) = \lim_{\hbar \to 0} f(r, p, t) \tag{1}$$

must be the "correct" classical phase-space distribution. That is, $f_c(r, p, t)$ must satisfy the Liouville equation.

(B) *Marginal Distributions:* The integral of f over one of the variables r or p must give the correct distribution in the other variable:

$$\int dr\, f(r, p, t) = \langle \delta(P - p) \rangle, \tag{2}$$

$$\int dp\, f(r, p, t) = \langle \delta(R - r) \rangle, \tag{3}$$

where R and P are the position and momentum operators.

(C) *Generalized Wigner Equivalents:* For any given function $A(R, P)$ of the position and momentum operators, we must be able to determine a generalized Wigner equivalent $a(r, p)$ such that

$$\langle A(R, P) \rangle = \int dr\, dp\, f(r, p, t) a(r, p). \tag{4}$$

We might point out here that the distributions introduced by Cohen[3] do not, in general, provide for a generalized Wigner equivalent. In particular for Cohen's distribution [Eq. (6.2) of Ref. 3], an operator of the form $A(\theta \cdot R + \tau \cdot P)$ does not have a generalized Wigner equivalent.

The most convenient way of finding generalized Wigner equivalents is by first finding the generalized Weyl correspondence. That is, we find the operator $A_g(\theta, \tau, R, P)$ for which the generalized Wigner equivalent is

$$a_g(\theta, \tau, r, p) = e^{i(\theta \cdot r + \tau \cdot p)}. \tag{5}$$

Then if the operators A_g are complete, we can expand any operator as

$$A(R, P) = \int d\theta\, d\tau\, \alpha(\theta, \tau) A_g(\theta, \tau, R, P). \tag{6}$$

(We consider the completeness of the A_g's when we specify the details of the distribution.) Clearly we can

* Work supported by the National Science Foundation.

† Present address: Dept. of Physics, Virginia Polytechnic Institute, Blacksburg, Va.

[1] E. Wigner, Phys. Rev. **40**, 749 (1932).

[2] K. İmre, E. Özizmir, M. Rosenbaum, and P. F. Zweifel, J. Math. Phys. **8**, 1097 (1967).

[3] L. Cohen, J. Math. Phys. **7**, 781 (1966).

determine the Wigner equivalent of $A(R, P)$ by knowing the generalized Wigner equivalent of the right-hand side of (6), that is, using Eq. (6):

$$\langle A(R, P)\rangle = \int d\theta\, d\tau\, \alpha(\theta, \tau)\langle A_g(\theta, \tau, R, P)\rangle,$$

or using (5) and (4):

$$a(r, p) = \int d\theta\, d\tau\, \alpha(\theta, \tau)e^{i(\theta\cdot r+\tau\cdot p)}. \tag{7}$$

It is easily shown that the expected values of the following generating operator:

$$D(R, P, r, p) = \frac{1}{(2\pi)^{6N}}\int d\tau'\, d\theta'\, e^{-i(\theta'\cdot r+\tau'\cdot p)}A_g(\theta', \tau', R, P) \tag{8}$$

gives a distribution for which (4) and (5) hold:

$$f_g(r, p, t) = \langle D(R, P, r, p)\rangle. \tag{9}$$

We show that this distribution also satisfies the other conditions that we listed earlier. Our approach here is related to that followed by Cohen.[3]

II. THE DISTRIBUTIONS

We can specify a distribution by writing the operators $A_g(\theta, \tau, R, P)$. We take, generally,

$$A_g(\theta, \tau, R, P) = g(\hbar\theta\cdot\tau)e^{i(\theta\cdot R+\tau\cdot P)}, \tag{10}$$

where $g(x)$ has a series expansion about zero of the form

$$g(x) = 1 + \sum_{n=1}^{\infty}\frac{x^{2n}}{(2n)!}g^{(2n)}(0). \tag{11}$$

Clearly we must take g to be an even function of $\hbar\tau\cdot\theta$ to insure that D is Hermitian.

The completeness of the operators $e^{i(\theta\cdot R+\tau\cdot P)}$ is shown in Ref. 2.

The Wigner distribution is obtained by taking

$$g(x) = 1.$$

Then,

$$f_w(r, p, t) = \frac{1}{(2\pi)^{6N}}\int d\tau'\, d\theta'\, e^{-i(\theta'\cdot r+\tau'\cdot p)}\langle e^{i(\theta'\cdot R+\tau'\cdot P)}\rangle. \tag{12}$$

This form was obtained by Moyal.[4] If we recall that

$$e^Ae^B = e^{A+B}e^{\frac{1}{2}[B,A]}, \tag{13}$$

for

$$[A, [B, A]] = [B, [B, A]] = 0,$$

and

$$[\theta\cdot R, \tau\cdot P] = i\hbar\theta\cdot\tau, \tag{14}$$

we can write (12) as

$$f_w(r, p, t) = \frac{1}{(2\pi)^{6N}}\int d\tau'\, d\theta'\, e^{-i(\theta'\cdot r+\tau'\cdot p)}\langle e^{i\tau'\cdot P/2}e^{i\theta'\cdot R}e^{i\tau'\cdot P/2}\rangle$$
$$= \frac{1}{(2\pi)^{3N}}\int d\tau'\, e^{-\tau'\cdot p}\langle e^{i\tau'\cdot P/2}\delta(R - r)e^{i\tau'\cdot P/2}\rangle \tag{15}$$

or, alternatively, we can write

$$f_w(r, p, t) = \frac{1}{(2\pi)^{6N}}\int d\tau'\, d\theta'\, e^{-i(\theta\cdot r+\tau'\cdot p)}\langle e^{i\theta'\cdot R/2}e^{i\tau'\cdot P}e^{i\theta'\cdot R/2}\rangle$$
$$= \frac{1}{(2\pi)^{3N}}\int d\theta'\, e^{-i\theta'\cdot r}\langle e^{i\theta'\cdot R/2}\delta(P - p)e^{i\theta'\cdot R/2}\rangle. \tag{16}$$

Using (15) and (16), it is a straightforward matter to derive Eqs. (5a) and (5b) of Ref. 2.

It is clear that the generating operator for the generalized distribution is related to the generating operator for the Wigner distribution by commutators of R and P, since

$$g(\hbar\tau\cdot\theta) = g(-i[\theta\cdot R, \tau\cdot P]). \tag{17}$$

As an example, let us consider the symmetric distribution introduced by Margenau and Hill.[5] As discussed by Cohen,[3] the appropriate $g(x)$ for this case is

$$g(x) = \cos(x/2).$$

In this case the distribution is

$$f_s(r, p, t) = \frac{1}{(2\pi)^{6N}}\int d\theta'\, d\tau'\cos(\hbar\tau'\cdot\theta'/2) \times e^{-i(\theta'\cdot r+\tau'\cdot p)}\langle e^{i(\theta'\cdot R+\tau'\cdot P)}\rangle. \tag{18}$$

When we note that

$$\cos(\hbar\tau'\cdot\theta'/2) = \tfrac{1}{2}\{e^{\frac{1}{2}[\theta'\cdot R,\tau'\cdot P]} + e^{-\frac{1}{2}[\theta'\cdot R,\tau'\cdot P]}\} \tag{19}$$

and use (13), we can write (18) as

$$f_s(r, p, t) = \frac{1}{(2\pi)^{6N}}\int d\theta'\, d\tau'\, e^{-i(\theta'\cdot r+\tau'\cdot p)} \times \langle\tfrac{1}{2}\{e^{i\theta'\cdot R}e^{i\tau'\cdot P} + e^{i\tau'\cdot P}e^{i\theta'\cdot R}\}\rangle$$
$$= \tfrac{1}{2}\langle\delta(R - r)\delta(P - p) + \delta(P - p)\delta(R - r)\rangle. \tag{20}$$

The remaining distributions commonly found in the literature can also be generated by an appropriate choice of $g(x)$.

[4] J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).

[5] H. Margenau and R. N. Hill, Progr. Theoret. Phys. (Kyoto) **26**, 722 (1961).

III. CONNECTIONS AMONG THE DISTRIBUTIONS

First let us show that the three properties of generalized phase-space distributions listed in Sec. I hold for the distributions generated by (8), (9), and (10).

Of course, our choice was made to provide a simple means of determining the generalized Wigner equivalents. Therefore we need not discuss this point further.

To find the classical limit we note[2] that $\langle e^{i(\theta\cdot R+\tau\cdot P)}\rangle$ has a series expansion in $\hbar$ and

$$\lim_{\hbar\to 0}\langle e^{i(\theta\cdot R+\tau\cdot P)}\rangle = \int dr'\,dp'\,f_c(r',p',t)e^{i(\theta\cdot r'+\tau\cdot p')}. \quad (21)$$

Also, we note from (11) that

$$\lim_{\hbar\to 0} g(\hbar\theta\cdot\tau) = 1. \quad (22)$$

Then,

$$\lim_{\hbar\to 0} f_g(r,p,t)$$
$$= \frac{1}{(2\pi)^{6N}}\int d\tau'\,d\theta'\,dr'\,dp'\,f_c(r',p',t)e^{i[\theta'\cdot(r'-r)+\tau'\cdot(p'-p)]}$$
$$= f_c(r,p,t). \quad (23)$$

Now let us consider the marginal distributions

$$\int dr\, f_g(r,p,t)$$
$$= \frac{1}{(2\pi)^{6N}}\int d\tau'\,d\theta'\,dr\,e^{-i(\theta'\cdot r+\tau'\cdot p)}g(\hbar\theta'\cdot\tau')\langle e^{i(\theta'\cdot R+\tau'\cdot P)}\rangle$$
$$= \frac{1}{(2\pi)^{3N}}\int d\tau'\,d\theta'\,\delta(\theta')e^{-ir'\cdot p}\langle e^{ir'\cdot P}\rangle,$$

where we have taken $\theta' = 0$ and noted that $g(0) = 1$. The remaining integrations give Eq. (2) for f_g. It is obviously just as easy to show that Eq. (3) holds for f_g.

To establish the equivalence of the various distributions, we explicitly insert (8) and (10) in (9):

$$f_g(r,p,t)$$
$$= \frac{1}{(2\pi)^{6N}}\int d\tau'\,d\theta'\,g(\hbar\theta'\cdot\tau')e^{-i(\theta'\cdot r+\tau'\cdot p)}\langle e^{i(\theta'\cdot R+\tau'\cdot P)}\rangle. \quad (24)$$

Using the property

$$g(x) = g(-x),$$

we note that

$$g(\hbar\theta'\cdot\tau)e^{-i(\theta'\cdot r+\tau'\cdot p)} = g(\hbar\nabla_r\cdot\nabla_p)e^{-i(\theta'\cdot r+\tau'\cdot p)}. \quad (25)$$

Recalling Eq. (12), we see that

$$f_g(r,p,t) = g(\hbar\nabla_r\cdot\nabla_p)f_w(r,p,t). \quad (26)$$

A form somewhat similar to this was used by von Roos[6] to obtain a distribution function for a molecular gas.

Now let us consider the generalized Wigner equivalent

$$a_g(r,p) = \int d\theta\,d\tau\,\alpha_g(\theta,\tau)e^{i(\theta\cdot r+\tau\cdot p)}, \quad (27)$$

where α_g is obtained from

$$A(R,P) = \int d\theta\,d\tau\,\alpha_g(\theta,\tau)g(\hbar\theta\cdot\tau)e^{i(\theta\cdot R+\tau\cdot P)}.$$

Since $g = 1$ for the Wigner distribution, we must have

$$\alpha_w(\theta,\tau) = \alpha_g(\theta,\tau)g(\hbar\theta\cdot\tau). \quad (28)$$

Applying (28) and (25) in (27), we have

$$a_w(r,p) = g(\hbar\nabla_r\cdot\nabla_p)a_g(r,p). \quad (29)$$

IV. DISCUSSION

Clearly, the generalized phase-space distributions and the generalized Wigner equivalents are different for different choices of $g(x)$. However, the important conclusions regarding these distributions must be concerned with their connections with experiments in terms of Eq. (4). Consider, then,

$$\langle F(R,P)\rangle = \int dr\,dp\,f_g(r,p,t)a_g(r,p,t). \quad (30)$$

Using (26) we have

$$\langle F(R,P)\rangle = \int dr\,dp\,a_g(r,p,t)g(\hbar\nabla_r\cdot\nabla_p)f_w(r,p,t).$$

Integrating by parts gives

$$\langle F(R,P)\rangle = \int dr\,dp\,f_w(r,p,t)g(\hbar\nabla_r\cdot\nabla_p)a_g(r,p,t)$$

and, using (29),

$$\langle F(R,P)\rangle = \int dr\,dp\,f_w(r,p,t)a_w(r,p,t). \quad (31)$$

It is not surprising that both (30) and (31) hold, since we constructed the generalized phase-space distributions to satisfy just these equations. However, the rather trivial connections among the various distributions does not seem to have been pointed out in the literature, and leads one to wonder why more than the Wigner distribution need be considered for any calculations.

Using Eqs. (26) and (29), we can immediately relate the results already obtained for the Wigner distribution (as for example in Ref. 2) to the corresponding results for a generalized phase-space distribution.

[6] O. von Roos, J. Chem. Phys. **31**, 1415 (1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Symmetry of the Ground Level of a Hamiltonian

W. H. KLEINER AND T. A. KAPLAN
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts*

(Received 29 June 1967)

A general connection between nodelessness and symmetry of a function is pointed out. It is proved that a real nodeless energy eigenfunction with energy E has a nonzero part which also is an eigenfunction with energy E and which under coordinate transformations has the full symmetry of the Hamiltonian. This result can be applied to many systems of physical interest for which the ground-state energy eigenfunction is known from the nature of the Hamiltonian to be nodeless. Simple counterexamples are given however, to show that not all Hamiltonians have a nodeless ground-state energy eigenfunction.

1. INTRODUCTION

The nodeless character of ground-state wavefunctions for many systems (Theorem 3 below) is well known[1–4] and has been applied since the earliest days of wave mechanics.[4–8] In this paper two theorems are proved concerning the symmetry of a nodeless function. As a result the symmetry of the ground level of a large class of Hamiltonians can be determined.[9]

The essence of Theorems 1 and 2, proved in Sec. 2, is that if an energy eigenfunction with energy E is nodeless, it must have a nonzero part which also is an eigenfunction with energy E and which has the symmetry of the Hamiltonian.

One can find in the literature results which are tantamount to, or depend in an essential way on, special cases of Theorem 2. Some results of this type and other possible applications are discussed in Sec. 3.

For one-dimensional Hamiltonians another approach has been used to determine the symmetry of the ground level.[9a]

2. THEOREMS

Consider a system of N particles. Assume that the Hamiltonian depends only on the position and momentum coordinates $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$ and $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_N)$ of the particles, so that in the coordinate representation the energy-eigenvalue problem has the form

$$\mathcal{H}\psi_{n\alpha}(\mathbf{r}) = E_n\psi_{n\alpha}(\mathbf{r}), \tag{1}$$

where α runs over a set of degenerate eigenfunctions corresponding to E_n. We assume that $\mathcal{H}$ is real (that is, that $\mathcal{H}$ is time-inversion invariant) so that the $\psi_{n\alpha}(\mathbf{r})$ can all be chosen real. A nonrelativistic spin-independent Hamiltonian of the form

$$\mathcal{H} = \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m_i}\right)\nabla_i^2 + V(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) \tag{2}$$

provides a familiar example. Let G denote a group of operators R representing a group $\mathcal{G}$ of coordinate transformations $\mathcal{R}$. In connection with (1) we restrict G to commute with $\mathcal{H}$. We are interested in the Γ_1 part of a function $\psi(\mathbf{r})$, the part invariant under G, as given by the projection formula

$$P_{\Gamma_1}\psi(\mathbf{r}) = \sum_R R\psi(\mathbf{r}) \Big/ \left(\sum_R 1\right), \tag{3}$$

where the sum is over all operators R in G; in the case of continuous groups the sum denotes the Hurwitz or invariant integral ($\int \cdots dR$) in group space.[10,11] We restrict our considerations to functions ψ and groups G for which the invariant projection $P_{\Gamma_1}\psi$ exists. If ψ is bounded, as is ordinarily the case for wavefunctions, $P_{\Gamma_1}\psi$ exists if $\sum_R 1$ is finite. The condition that $\sum_R 1$ be finite is satisfied for any finite group and for many physically interesting continuous groups—the three-dimensional rotation group, for example.

We consider now three theorems concerning nodeless functions. By a nodeless function in $\mathcal{N}$ dimensions we mean a function which vanishes in no $(\mathcal{N}-1)$-dimensional subregion in the interior of its domain of definition. A nodeless function is thus not identically zero.

* Operated with support from the U.S. Air Force.

[1] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience Publishers, Inc., New York, 1953), Vol. I.

[2] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Co., Inc., New York, 1953), pp. 753–757.

[3] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1958), p. 56.

[4] C. Herring, *Magnetism IV*, G. T. Rado and H. Suhl, Eds. (Academic Press Inc., New York, 1966), Chap. VIII.

[5] W. Heitler and F. London, Z. Physik **44**, 455 (1927).

[6] W. Heisenberg, Z. Physik **49**, 619 (1928).

[7] C. Herring, Rev. Mod. Phys. **34**, 631 (1962).

[8] E. Lieb and D. Mattis, Phys. Rev. **125**, 164 (1962).

[9] W. H. Kleiner and T. A. Kaplan, Bull. Am. Phys. Soc. **12**, 381 (1967).

[9a] B. W. Downs, Am. J. Phys. **31**, 277 (1963).

[10] E. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959), p. 95.

[11] M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Addison-Wesley Publishing Co., Inc., Reading, Mass., 1962), p. 313.

Theorem 1: If a real function $\psi(\mathbf{r})$ is nodeless, then $P_{\Gamma_1}\psi$, the part of ψ invariant under G, is also nodeless.

Proof: This can be proved as follows: Without loss of generality we assume $\psi(\mathbf{r}) \geq 0$. Since $\psi(\mathbf{r}) \geq 0$, so is its transform $R\psi(\mathbf{r}) = \psi(\mathcal{R}^{-1}\mathbf{r})$. Thus $\sum_R R\psi(\mathbf{r})$ is a sum of nodeless functions, each ≥ 0, and is therefore itself nodeless. Consequently, if $\sum_R 1$ is finite, $P_{\Gamma_1}\psi$ is also nodeless.

Now suppose ψ is an eigenfunction of $\mathcal{H}$ with eigenvalue E and the operators R in G commute with $\mathcal{H}$. Then we have

$$\mathcal{H}(P_{\Gamma_1}\psi) = E(P_{\Gamma_1}\psi), \tag{4}$$

since $\mathcal{H}P_{\Gamma_1} = P_{\Gamma_1}\mathcal{H}$. We now apply Theorem 1 to insure that $P_{\Gamma_1}\psi$ is not identically zero. Thus, an immediate consequence of Theorem 1 is:

Theorem 2: If $\psi(\mathbf{r})$ is a real nodeless eigenfunction of $\mathcal{H}$ with eigenvalue E, then so is its Γ_1 part $P_{\Gamma_1}\psi$.

Note that if one chooses the $\psi_{n\alpha}$ of (1) to be orthogonal (which involves no physical restriction), then there is at most one real nodeless eigenfunction in the set of $\psi_{n\alpha}$, since two real nodeless functions cannot be orthogonal.

This derivation applies not only to (1), but also to the momentum-space counterpart of (1), provided $\mathcal{H}$ is invariant under space inversion ($\mathbf{r} \to -\mathbf{r}$) so that $\mathcal{H}$ in the momentum representation will be real. Note that the Fourier transform of a real nodeless function need not be nodeless.

Under what conditions does a Hamiltonian[12] have a real nodeless energy eigenfunction, so that Theorem 2 can be applied? This question can be answered in part by the following theorem, which has, in itself, nothing to do with symmetry.

Theorem 3: Let ρ be a positive function and $L[u] + \lambda\rho u = 0$ be a self-adjoint second-order differential equation for a domain D with arbitrary homogeneous boundary conditions; then

(A) the lowest eigenvalue has a real nodeless eigenfunction[13,14];

(B) the lowest eigenvalue is nondegenerate.[15]

There is a large class of physically interesting Hamiltonians [including (2)] with Schrödinger equations which satisfy the conditions of Theorem 3. For these Hamiltonians we can conclude, using Theorem 2, that *the eigenfunction having the eigenvalue of lowest energy is invariant under G.*[16] Since the lowest eigenvalue is nondegenerate for these Hamiltonians, Theorem 2 serves to select Γ_1 as the symmetry of the lowest eigenvalue from among only the one-dimensional irreducible representations of G.

Not all Hamiltonians have the above properties A and B. In fact, the presence or absence of properties A and B in the four possible combinations can all be realized by appropriately choosing the potential $V(x)$ and the real parameters a, b, and c in the Hamiltonian

$$\mathcal{H} = ap^2 + bp^4 + cp^6 + V(x), \tag{5}$$

where $p = -i(d/dx)$ and we take $\hbar = 1$. For example,

(i) A and B:

$$V = b = c = 0, \quad a > 0;$$

(ii) neither A nor B:

$$V = c = 0, \quad a < 0, \quad b > 0;$$

(iii) A but not B:

$$V = 0, \quad a = 1, \quad b = -2, \quad c = 1;$$

(iv) B but not A:

$$V = 0 \quad \text{for} \quad |x| < \pi$$

and

$$V = \infty \quad \text{for} \quad |x| \geq \pi,$$
$$a = -2n^2b \quad \text{or} \quad a = -2(n + 1/2)^2b,$$
$$b > 0, \quad c = 0,$$
$$n = 1, 2, \cdots .^{17,17a}$$

[12] In this paper, the energy-eigenvalue spectrum is always assumed to have a lower bound.

[13] Reference 1, p. 451. This theorem is derived in Ref. 1 by methods of the calculus of variations. It is a generalization to N dimensions of a result which can be proved very simply in one dimension using Sturm–Liouville theory; the usual Sturm–Liouville proof applies only in the one-dimensional case, however.

[14] It should be pointed out that property A does not hold unless certain restrictions are satisfied. In particular, in the case of the Schrödinger equation the potential cannot have so strong an infinity that it causes the vanishing of all eigenfunctions on a surface interior to D. A different proof of properties A and B has been given in Ref. 8, p. 166, and has been further discussed in Ref. 4, p. 168.

[15] Reference 1, p. 458. Property B is proved by using property A. The statement and proof given can be easily extended to include the class of differential equations and boundary conditions considered in Theorem 3.

[16] In particular, although accidental degeneracy may be present, it cannot be in the lowest level.

[17] Example (iii), which shows the possible existence of a nodeless ground-energy eigenfunction belonging to a degenerate level, provides a counterexample to the "proof" of the contrary in Ref. 3, p. 56. This example shows that a satisfactory proof of nondegeneracy (as given, for example, in Ref. 1) must depend on the particular nature of the Hamiltonian involved, which the "proof" of Ref. 3 does not.

[17a] The invalid proof given by Landau and Lifshitz (see our Footnote 17) has been presented again recently [R. A. Hegstrom and W. N. Lipscomb, Rev. Mod. Phys. **40**, 354 (1968), Footnote 22].

3. PHYSICAL APPLICATIONS

Theorem 2, in conjunction with Theorem 3, is useful in determining the symmetry of the lowest-energy eigenvalue of a large variety of physical systems. It applies to any boson system. However, it does not apply to physical systems containing more than two electrons (or, more generally, more than $2s + 1$ indistinguishable fermions, where s is the fermion spin quantum number). This is because the physical ground energy is not the lowest eigenvalue of $\mathcal{H}$, but is, in fact, an excited eigenvalue of $\mathcal{H}$. This well-known fact[18] is a consequence of the requirement that an electronic wavefunction be antisymmetric with respect to interchange of the space–spin coordinates of any two electrons. A corollary is that for a boson system a position-coordinate wavefunction always has an energy no lower than the physical ground energy, while for a fermion system this energy may be lower than the physical ground energy. Hence, in a variational calculation for a many-fermion system, it is necessary that the trial wavefunction have appropriate permutation symmetry, although other symmetries may be violated.

If for a two-electron system we let G be the group of permutations on the two electron-position coordinates, we conclude (from Theorem 2 together with Theorem 3) that the lowest eigenfunction of (1) must be symmetric under G; for the corresponding space–spin eigenfunction to be antisymmetric, the spin function alone must be antisymmetric, so that the physical ground state is a singlet, a result long well known.[6,7,4]

In order further to illustrate the applicability of Theorem 2, some other immediate results of applying it together with Theorem 3 are listed below. Examples of possible physical interest are given for several groups G; the notation $[G, \Gamma_1]$ is used.[19] One-electron systems are listed before the semicolon, two-electron systems after. In the examples, nuclei are assumed fixed in position, and no attempt has been made to exclude examples with unstable nuclear configurations:

$[O(3), D^{(0+)}]$ One-center systems: H, D, He^+, $Li^{2+}, \cdots$; H^-, D^-, He, Li^+, $Be^{2+}, \cdots$.

$[C_{\infty v}, \Sigma^+]$ Two-center systems: HD^+, $HHe^{2+}, \cdots$; HD, $HHe^+, \cdots$.

$[D_{\infty h}, \Sigma_g^+]$ Two-center systems: H_2^+, D_2^+, $He_2^{3+}, \cdots$; H_2, D_2, $He_2^{2+}, \cdots$.

$[T_d, A_1]$ CH_4^{9+}; CH_4^{8+}.

Note that the applicability of Theorem 2 does not depend on the nuclear charge being an integral multiple of the proton charge, nor even on a "center" being represented by a Coulomb potential. This suggests that Theorem 2 (and/or Theorem 3) may be useful for some systems with more than two electrons when the system can be adequately described by a model in which one or two of the electrons move in an effective field which includes the effect of interaction with the remaining electrons. In such an application it would be important to take account somehow of the exclusion principle; otherwise, for example, for a free V^{4+} ion one might conclude from Theorem 2 together with Theorem 3 that the ground level had S_g symmetry rather than the D_g symmetry which it is known[20] experimentally to have and which is expected from its valence-electron configuration $3d$. Similarly, if a V^{4+} ion is in an ordinary crystal field of octahedral symmetry ($G = O_h$), one might conclude that the ground level has A_{1g} symmetry rather than the symmetry F_{2g} ($= T_{2g}$), which it is known to have.

[18] See, for example, Ref. 3, p. 214, or Ref. 4, p. 168.

[19] The notation used for the group and representation symbols G and Γ_1 can be found in Ref. 11. Spectroscopic notation for $D^{(0+)}$ is S_g.

[20] C. E. Moore, Natl. Bur. Std. (U.S.) Circ. No. 467 (1949), Vol. I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Wave Propagation and Other Spectral Problems in Kinetic Theory

LAWRENCE SIROVICH
Division of Applied Mathematics and the Center for Fluid Dynamics, Brown University
AND
JAMES K. THURBER
Department of Applied Mathematics, Brookhaven National Laboratories

(Received 27 March 1968)

A rigorous discussion of the spectrum of the linear Boltzmann equation and kinetic models is presented. Particular attention is given to plane-wave propagation for a general class of kinetic models. These models in general have a velocity-dependent collision frequency $\nu(\xi)$. The main results of this paper concern the relationship between the complex wavenumber and complex frequency for a plane wave. It is shown that the question of analyticity of this relation is reduced to considering ν in the neighborhood of infinity. Specifically, if

$$\lim_{\xi\to\infty} \nu/\xi = 0,$$

the relationship is not analytic. Otherwise, analyticity is obtained. (Although not specifically considered here, analyticity is closely connected to the convergence of the Chapman–Enskog procedure.) In a general discussion it is shown how the question of analyticity is closely connected with (i) the continuous spectrum of the underlying operator, (ii) the behavior of solutions at large distances from boundaries, and (iii) the nature of the cutoff in an intermolecular interaction.

1. INTRODUCTION

Aside from its own intrinsic interest and importance, the study of plane-wave propagation is a powerful tool for revealing the underlying structure of the equations used in kinetic theory. Of special interest is the functional relation between frequency σ and wavenumber s which is implied by a plane wave. For generality, we permit both these quantities to be complex.[1] The function relationship $\sigma(s)$ [or $s = s(\sigma)$] is usually found as a root of the dispersion relation of the underlying equations and, in general, a number of such relations are possible. A special role is played by those roots for which $\sigma(0) = 0$. These roots are referred to as the hydrodynamical roots since they are the kinetic theory form of the roots found from the hydrodynamical equations.[2]

The analytical form of $\sigma(s)$ [or $s(\sigma)$] in the neighborhood of the origin has a vital role in any discussion of the transition of molecular to continuum theory. More specifically, the convergence of the series expansion of $\sigma(s)$ at the origin is closely connected with the convergence of the Chapman–Enskog procedure.[3] Recently, two conflicting results have been reported in regard to this issue. In an earlier paper,[4] we constructively demonstrated (using a special kinetic model) that this series is, in fact, divergent. On the other hand, for the case of rigid sphere molecules, the series has been found to converge; that is, $\sigma(s)$ is analytic at the origin.[5,6] This conflict may be isolated, and in fact resolved, by considering the molecular collision frequency $\nu(\xi)$. [Section 2 contains a discussion of $\nu(\xi)$ as well as a number of results for and forms of the linear Boltzmann equation.]

The form which $\nu(\xi)$ takes depends greatly on the effective range of the intermolecular potential. For most infinite-range potentials the analytical definition of $\nu(\xi)$ leads to divergent integrals. For this and other reasons, some sort of interaction cutoff seems advisable. From the analytical standpoint the most effective part of the collision frequency is its behavior at large speeds, $\xi \gg 1$. For a rigid sphere gas one finds (Sec. 2):

$$\lim_{\xi\to\infty} \frac{\nu}{\xi} = \pi D^2 n,$$

where D is the molecular diameter and n the number density. The same result is obtained if an arbitrary potential is given a radial cutoff D. If the interaction cutoff is allowed to be velocity-dependent, a wide range of behaviors is obtained. We represent the behavior at infinity by

$$\nu = 0(\xi^\alpha).$$

For example, the recently discussed angular cutoff[7]

[1] For a disturbance in an unbounded media, i.e., a free wave, $s = ik$, with k real. On the other hand, for forced (steady-state) oscillations, $\sigma = i\omega$, ω real.

[2] L. Sirovich, Phys. Fluids **6**, 10 (1963).

[3] See for instance, S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge University Press, London, 1952).

[4] L. Sirovich and J. K. Thurber, J. Math. Phys. **8**, 888 (1967). Hereafter this will be referred to as I. The result quoted above was reported earlier in our lectures in *Statistical Mechanics and Spectral Theory*, J. Pincus, Ed. (Brookhaven National Laboratories, Brookhaven, N.Y., 1965).

[5] J. A. McLennen, Phys. Fluids **8**, 1580 (1965).

[6] A. A. Arseniev, Zh. Vitch. Math. Mat. Phys. (Moscow) **5**, 854 (1965).

[7] H. Grad in *Third International Rarefied Gasdynamics Symposium* (Academic Press Inc., New York, 1963), Vol. 1, p. 26.

leads to $0 \leq \alpha < 1$ (negative α can be obtained by considering interactions which are softer than the Maxwell potential). The kinetic models considered in I correspond to an exponent $\alpha = 0$. Returning to the question of the analyticity of $\sigma(s)$, it was conjectured in I that analyticity is obtained if $\alpha \geq 1$ and that nonanalyticity occurs otherwise. This is proven in Secs. 5 and 6 for a very general class of kinetic models. Section 5 contains the proof for models of the type considered in I, $\alpha = 0$. Section 6 contains the proof for velocity-dependent collision-frequency models.

The velocity-dependent collision-frequency models of Sec. 6 are developed from the Boltzmann equation in Sec. 3. As a starting point we use the model recently introduced by Cercignani.[8] The development then follows the methods[9,10] used in extending the Krook model.[11] Other velocity-dependent collision-frequency models have also been introduced in neutron-diffusion theory.[12] These do not satisfy all the conservation laws and we do not specifically consider them (most of our results still apply, however).

In the past, great use has been made of the BGK model[11] and its extensions.[9,10] In recent years, however, it has become increasingly clear that for many purposes these are too crude an approximation to the Boltzmann equation. In order to produce a more faithful model, it should more accurately mimic the spectrum of the Boltzmann equation. For this reason a general discussion of the spectrum for initial and boundary problems is given in Sec. 4. In the course of preparing this section, it was found that a mathematically rigorous treatment could be given even for the Boltzmann equation. (This too is included in Sec. 4). Although all previous discussions of the spectrum of the Boltzmann equation[13–15] are based only on plausibility arguments, no significantly different results are found here.

Although our main results (Secs. 5 and 6) apply to plane waves, and hence to the discrete modes or spectra, there are a number of interesting and informative connections with other diverse problems and effects. These we discuss now in the introduction. To begin, we consider the continuous spectrum.

To discuss this we write the linear Boltzmann equation and the model equations in the form

$$\left(\frac{\partial}{\partial t} + \boldsymbol{\xi} \cdot \frac{\partial}{\partial \mathbf{x}} + \nu(\xi)\right) g = Kg,$$

where K represents an integral operator. Provided the collision frequency $\nu(\xi)$ exists, this can always be done. Two canonical situations are considered: the initial-value problem in an unbounded domain with sinusoidal initial data of wavenumber k. This leads to a discussion of the spectrum of the operator

$$L_i = K - \nu(\xi) - ik\xi_1 = K - T_i.$$

Second, the case of steady-state oscillations of frequency ω, in a half-space. This leads to a discussion of the spectrum of the operator

$$L_b = \frac{1}{\xi_1} K - \frac{\nu(\xi) + i\omega}{\xi_1} = \frac{1}{\xi_1} K - T_b.$$

Ford[14] has already considered the spectrum of L_i for a rigid sphere gas and for a certain kinetic model (see Sec. 3). Also, Grad[15] considers the spectrum of both L_i and L_b in a number of situations. Both of these studies, however, are based only on plausibility arguments, and in Sec. 4 we give a mathematically rigorous discussion of the spectrum of L_i and L_b for both kinetic models and the Boltzmann equation.

In Sec. 4 we demonstrate that if K is compact, then the continuous spectra of L_i and $-T_i$ are the same. Using the additional information that in all cases K is isotropic, it also follows, then, that L_b and $-T_b$ have the same continuous spectra. Since both $-T_i$ and $-T_b$ are multiplicative operators, their spectra are given by their respective ranges. That is, for the continuous spectra of $-T_i$ (and hence L_i), we hold k (real) fixed and vary $\boldsymbol{\xi}$, and for the continuous spectra of $-T_b$ (and hence L_b), we hold ω (real) fixed and vary $\boldsymbol{\xi}$. Sketches of the continuous spectrum have been given by Ford[14] and Grad.[15] For completeness we repeat some of these.

Figure 1 shows a sketch of the initial-value problem continuous spectrum in the rigid sphere (and radial cutoff) case. As is easily seen, the portion closest to the imaginary axis is due to the slow-moving molecules. The comparable boundary-value case is sketched in Fig. 2. We note that in this case two branches are obtained and that the portion closest to the imaginary axis is due to the fast-moving molecules, $|\xi_1| \to \infty$. As is shown in Figs. 3 and 4, a constant collision frequency leads to a one-dimensional continuous spectrum. In Fig. 4 the contribution in the neighborhood of origin comes from $|\xi_1| \to \infty$. The sketches in Figs. 5 and 6 are comparable to those discussed by Grad.[15] As in Fig. 4 the neighborhood of the origin in Fig. 6 is due to $|\xi_1| \to \infty$.

---

[8] C. Cercignani, Ann. Phys. (N.Y.) **40**, 469 (1966).
[9] E. P. Gross and E. A. Jackson, Phys. Fluids **2**, 432 (1959)
[10] L. Sirovich, Phys. Fluids **5**, 908 (1962).
[11] P. R. Bhatnager, E. P. Gross, and M. Krook, Phys. Rev. **94**, 511 (1954).
[12] N. Corngold, P. Michael, and W. Wollman, Nucl. Sci. Eng. **15**, 13 (1963).
[13] N. Corngold, Nucl. Sci. Eng. **19**, 80 (1964).
[14] G. W. Ford, "Dispersion of Sound in Monatomic Gases," in Proceedings of Midwest Conference on Theoretical Physics, 1963 (unpublished).
[15] H. Grad, SIAM J. Appl. Math. **14**, 932 (1966).

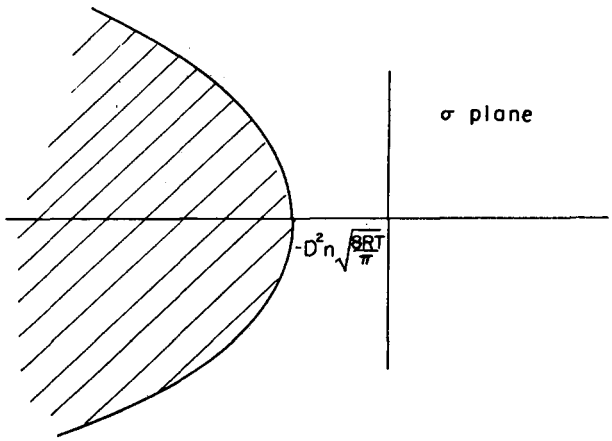


FIG. 1. Continuous spectra for initial-value problem; $\nu \sim |\xi|$.

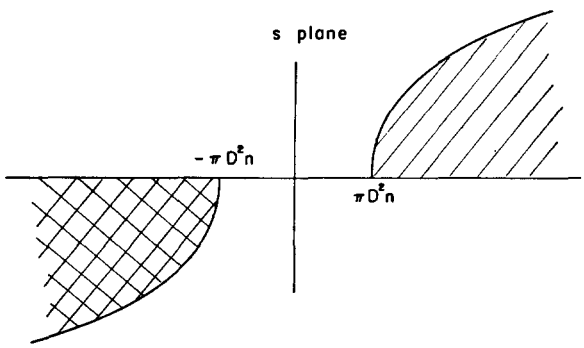


FIG. 2. Continuous spectra for steady-state oscillations; $\nu \sim |\xi|, \omega > 0$.

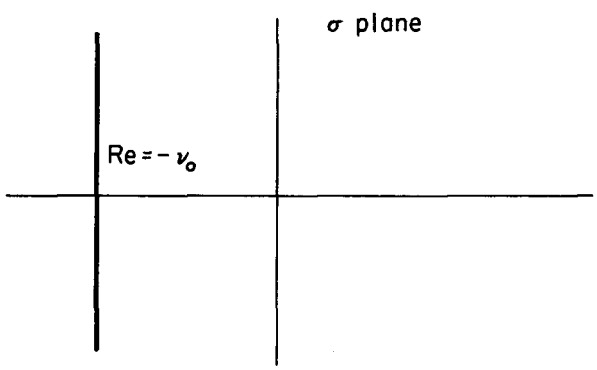


FIG. 3. Continuous spectra for initial-value problem; ν a constant.

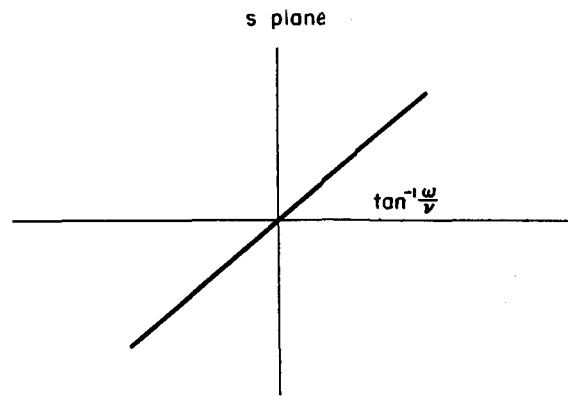


FIG. 4. Continuous spectra for steady-state oscillations; ν a constant, $\omega > 0$.

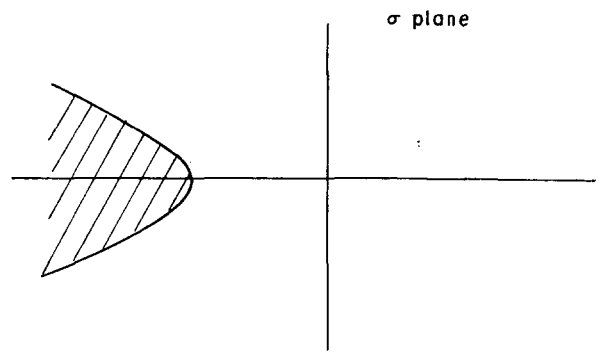


FIG. 5. Continuous spectra for initial-value problem; $\nu \sim |\xi|^\alpha, 0 < \alpha < 1$.

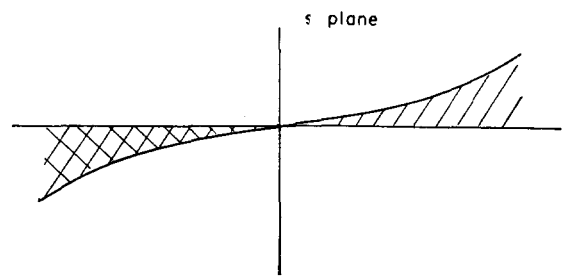


FIG. 6. Continuous pectra for steady-state oscillations; $0 < \alpha < 1, \omega > 0$.

Based on the above discussion, the question of the analyticity of $\sigma(s)$ at $s = 0$ [or $s(\sigma)$ at $\sigma = 0$] can be connected to the continuous spectrum for the steady-state oscillations problem. If the region of the continuous spectrum does not reach the origin, $\sigma(s)$ is analytic; otherwise it is not analytic. (Actually this effect plays a role in the proofs given in Secs. 5 and 6.)

Next we discuss the canonical sound propagation problem, i.e., the problem of an infinite plane oscillating normal to itself in anotherwise unbounded space. A solution is then soughtin one of the half-spaces, say $x > 0$. In addition to th continuous spectra, there is the point spectra whichin Figs. 2, 4, 6 lies in either the first or third quadrant. (As was shown in I at sufficiently high frequncies, $\omega \gg 1$, the discrete modes disappear.) Fc $x > 0$, only the discrete points in the third quadrant enter (which for $\omega > 0$ can be shown to lie of the imaginary axis), and also the continuous spectrain that quadrant. Therefore if we consider steady stte oscillations in a rigid sphere gas, Fig. 2 shows tht the solution falls off at least

exponentially for $x \gg 1$. The same statement cannot be made for the constant collision-frequency case shown in Fig. 4. In fact, from the work on this problem using the Krook model,[16–18] or from the related Rayleigh problem considered by Cercignani,[19] we can expect that, for $x \gg 1$, the solution falls off as $\exp(-kx^{\frac{3}{2}})$ (k a constant). Although no calculations have been performed for the cases depicted in Fig. 6, we can anticipate that the far-field solutions will fall off as $\exp(-kx^{\beta})$, $\frac{1}{2} \leq \beta < 1$ (k a constant). Hydrodynamics leads only to discrete modes, and hence to exponential decay. Therefore, in those cases for which

$$\lim_{\xi \to \infty} \left(\frac{\nu}{\xi}\right) = 0,$$

the far field is not hydrodynamical, even when $0 < \omega \ll 1$. Only for the rigid sphere gas or, more generally, for the finite-interaction case is the far field hydrodynamical in nature for small ω. Although there is no reason to regard hydrodynamics as sacrosanct, there does not seem to be any experimental evidence that it does not give rise to the dominant effect for smooth phenomena. Certainly, examining the far field for $0 < \omega \ll 1$ falls into this category. A careful experiment under these conditions would go far in resolving the question of interaction cutoff for the Boltzmann equation.

The effects just described, and their sensitivity to the collision frequency, may be viewed in another way which to some extent explains, physically, why they occur. Nonhydrodynamical behavior occurs in those cases for which the continuous spectra reaches the origin (Figs. 4 and 6). The continuous spectra in the neighborhood of the origin is, in turn, due to the molecules for which $|\xi_1| \gg 1$, and hence $\xi \gg 1$. Now the free path of a molecule moving with a speed ξ is

$$l(\xi) = \xi/\nu.$$

Therefore, if $\alpha < 1$, then (ξ) becomes unbounded for ξ large. Physically this staes that there are molecules which can travel any distance, no matter how large, and on the average not encounter another molecule. It is clear that such molecules carry signals which are not hydrodynamical in naure. On the other hand, for a finite cutoff D, say, we have

$$l_0 = l(\infty) = 1/\pi D^2 n.$$

In this case, no molecule, no matter how large its speed, can travel on the average a distance larger than l_0 without colliding with another molecule. Therefore, the signals carried by the molecules in this case are collision dominated and, at least for $0 < \omega \ll 1$, are hydrodynamical.

In this introduction we have discussed the question of analyticity of $\sigma(s)$ [or $s(\sigma)$], the topology of the continuous spectra, the range of intermolecular forces, the far-field solution and the free path of fast molecules. It is a remarkable and interesting fact that all of these seemingly diverse effects are so intimately interwoven. So much so that an accurate knowledge of any one aspect resolves for us the nature of the other effects.

2. THE LINEAR BOLTZMANN EQUATION AND THE COLLISION FREQUENCY

Using the notation of I, the linear Boltzmann equation has the form

$$Dg = \left(\frac{\partial}{\partial t} + \boldsymbol{\xi} \cdot \nabla\right) g = \int \Omega_*(g'_* + g' - g_* - g) \times B(\theta, |\boldsymbol{\xi}_* - \boldsymbol{\xi}|)\, d\theta\, d\epsilon\, d\boldsymbol{\xi}_* = Lg. \quad (2.1)$$

g is the dimensionless perturbed distribution function and Ω is the normalized Gaussian:

$$\Omega = [\exp(-\xi^2/2)](2\pi)^{-\frac{3}{2}}. \quad (2.2)$$

Using two common descriptions, the collision frequency is defined by

$$\nu(\xi) = \int \Omega_* B(\theta, |\boldsymbol{\xi}_* - \boldsymbol{\xi}|)\, d\epsilon\, d\theta\, d\boldsymbol{\xi}_* = \int \Omega_* |\boldsymbol{\xi} - \boldsymbol{\xi}_*|\, b\, db\, d\epsilon\, d\boldsymbol{\xi}_*. \quad (2.3)$$

Using operator notation,

$$Lg = kg - \nu g,$$

where

$$kg = \int \Omega_* k(\boldsymbol{\xi}, \boldsymbol{\xi}_*) g_*\, d\boldsymbol{\xi}_*. \quad (2.4)$$

The symmetry of $k(\boldsymbol{\xi}, \boldsymbol{\xi}_*)$ is easily demonstrated and explicit forms for it have been known for some time.[3] In connection with (2.4), it is natural to consider the inner product

$$(f, g) = \int \Omega \bar{f} g\, d\boldsymbol{\xi} \quad (2.5)$$

($\bar{f}$ denotes the complex conjugate of f) and the resulting Hilbert space which we denote by h. For some purposes it is more convenient to remove the weight

[16] H. Weitzner in *Rarefied Gasdynamics*, J. H. de Leeuw, Ed. (Academic Press Inc., New York, 1965), Vol. 1, p. 1.
[17] R. Mason in *Rarefied Gasdyamics*, J. H. de Leeuw, Ed. (Academic Press Inc., New York, 1965), Vol. 1, p. 44.
[18] H. S. Ostrowski and D. J. Kleman, Nuovo Cimento **64b**, 49 (1966).
[19] C. Cercignani, "Elementary Solutions of Linearized Kinetic Models and Boundary Value Problems in the Kinetic Theory of Gases," Brown University Report, 1965 (unpublished).

function Ω. Setting

$$g = \Omega^{-\frac{1}{2}}G$$

in (2.1), we obtain

$$DG = (K - \nu)G,$$

where (2.6)

$$KG = \int \Omega^{\frac{1}{2}}(\xi)k(\xi, \xi_*)\Omega^{\frac{1}{2}}(\xi_*)G(\xi_*)\, d\xi_*.$$

In this case one considers the Hilbert space H defined by the inner product

$$[F, G] = \int FG\, d\xi. \tag{2.7}$$

In certain contexts a further transformation is useful. Set

$$G = \hat{G}\nu^{-\frac{1}{2}}$$

in (2.6) to obtain

$$D\hat{G} = (\nu\hat{K} - \nu)\hat{G}, \tag{2.8}$$

with

$$\hat{K}G = \int \frac{K(\xi, \xi_*)}{[\nu(\xi)\nu(\xi_*)]^{\frac{1}{2}}} G(\xi_*)\, d\xi_*$$
$$= \int \left[\frac{\Omega(\xi)}{\nu(\xi)}\right]^{\frac{1}{2}} k(\xi, \xi_*)\left[\frac{\Omega(\xi_*)}{\nu(\xi_*)}\right]^{\frac{1}{2}} G(\xi_*)\, d\xi_*. \tag{2.9}$$

Similarly we set

$$g = \hat{g}\nu^{-\frac{1}{2}}$$

in (2.4); we obtain

$$D\hat{g} = (\nu\hat{k} - \nu)\hat{g},$$

with (2.10)

$$\hat{k}\hat{g} = \int \frac{\Omega_* k(\xi, \xi_*)}{[\nu(\xi)\nu(\xi_*)]^{\frac{1}{2}}} \hat{g}(\xi_*)\, d\xi_*.$$

For the last form the inner product

$$\langle f, g\rangle = \int \nu\Omega f g\, d\xi \tag{2.11}$$

and the resulting Hilbert Space $\hat{h}$ prove useful.

Hecke[20] first demonstrated that for rigid sphere molecules $\hat{K}$ is compact on H. Subsequent proofs of this have been given by Carleman[21] and Finkelstein.[22] A consequence of this is that

$$\nu^{-1}k$$

is compact with respect to $\hat{h}$. More recently, Dorfman[23] and Grad[7] have demonstrated a stronger result; namely that for rigid sphere molecules, K itself is compact on H. (And hence k is compact on h.) In the latter reference the compactness proof is extended to a somewhat more general class of molecules.

For rigid spheres the impact parameter has for a limit of integration the molecular diameter D. Hence from (2.2)

$$\nu = \int \Omega_* |\xi - \xi_*|\, d(b^2/2)\, d\epsilon\, d\xi_*$$
$$= D^2\sqrt{2\pi}\left[e^{-\xi^2/2} + \left(\xi + \frac{1}{\xi}\right)\int_0^{\xi} e^{-w^2/2} dx\right]. \tag{2.12}$$

In general, most infinite-range potentials do not lead to convergent forms for ν. To obtain convergence, we can, for example, introduce a radial cutoff in the impact parameter.[24] If we denote the radial cutoff by D, it is clear that (2.12) is again obtained for the collision frequency. It remains an open question, however, as to whether K (or $\hat{K}$) under this condition is a compact operator.

We can generalize the above by introducing a velocity dependent cutoff $R(|\xi - \xi_*|)$, which leads to

$$\nu(\xi) = \pi \int \Omega_* |\xi - \xi_*| R^2(|\xi_* - \xi|)\, d\xi_*. \tag{2.13}$$

As is clear, $\nu(\xi)$ is convergent under a wide set of conditions. Angular cutoff[7] is included in this category. This last assumption eliminates collisions which produce only grazing collisions. For example, for repulsive power law potentials

$$V = \kappa/r^p, \quad \kappa > 0,$$

we obtain from the first form of (2.3),

$$\nu(\xi) \equiv 2\pi A(\epsilon) \int \Omega_* |\xi_* - \xi|^{(p-4)/p}\, d\xi_*. \tag{2.14}$$

The constant $A(\epsilon)$ is given by

$$A(\epsilon) = \left[\frac{4K}{m}\right]^{2/p} \int_0^{(\pi/2)-\epsilon} \left|\beta \frac{d\beta}{d\theta}\right| d\theta,$$

and $\beta(\theta)$ implicitly by

$$\theta = \int_0^{\mu_0} \frac{d\mu}{1 - \mu^2 - (\mu/\beta)^p}$$

[μ_0 is such that $1 - \mu_0^2 - (\mu_0/\beta)^p = 0$]. It is easily shown that $A(\epsilon)$ diverges as $\epsilon \to 0$. Comparing (2.14) with (2.13) we obtain

$$R(|\xi - \xi_*|) = [A(\epsilon)]^{\frac{1}{2}}/|\xi_* - \xi|^{2/p}$$

for angular cutoff.

[20] E. Hecke, Math. Z. **12**, 272 (1922).

[21] T. Carleman, *Problèmes mathématiques dans la theorie cinétique des gas* (Almquist & Wiksells, Uppsala, 1957).

[22] L. Finkelstein, thesis, Hebrew University, Jerusalem, 1962.

[23] P. Dorfman, Proc. Natl. Acad. Sci. (U.S.) **50**, 805 (1963).

[24] All accepted demonstrations of the Boltzmann equation, including the modern hierarchy derivations assume, at least implicitly, a finite-range cutoff in the interaction potential. Whether the assumption of angular cutoff (Ref. 7) can be compatibly included in these derivations is not clear. It should be noted that the still open question of correctly terminating intermolecular effects is not a classical problem.

The principal reason for introducing the angular cutoff lies in the fact that $K(\xi, \xi_*)$ may then be shown to be completely continuous.[7] From the mathematical viewpoint, the efficacy of this assumption is therefore clear; however, its physical implications cast some doubt on it. For from (2.14), we see that the fast-moving molecules have almost no encounters. Another way of seeing this is by considering the free-path

$$l(\xi) = \xi/\nu(\xi).$$

For rigid spheres

$$\lim_{\xi\to\infty} \frac{\xi}{\nu} = \frac{1}{\pi D^2 n},$$

(since $\nu \sim n\pi D^2\xi$), whereas for angular cutoff molecules

$$\lim_{\xi\to\infty} \frac{\xi}{\nu} = \infty$$

(since in this case $\nu \sim \xi^{(p-4)/p}$).

3. KINETIC MODELS

In order to learn about the effect of velocity-dependent collision frequencies, it is natural to consider kinetic models exhibiting this effect. Ford[14] in considering the initial value problem introduces

$$Dg = Fg - \nu(\xi)g, \tag{3.1}$$

with

$$Fg = \nu(\xi)\int \Omega_*[1 + \xi\cdot\xi_* + \tfrac{1}{8}(\xi^2 - 3)(\xi_*^2 - 3)]g_*\, d\xi_*.$$

One easily sees that

$$0 = (F - \nu)[1, \xi, \xi^2] = (k - \nu)[1, \xi, \xi^2].$$

Although F shares this property with the exact operator k, (3.1) does not preserve the conservation laws. This may be traced to the fact that F, unlike k, is not symmetric under the inner product (2.5).

In an independent paper, Cercignani[8] avoided this difficulty in a striking manner. (Other models have been introduced in neutron diffusion; these, however, do not satisfy all the conservation laws.) We now give a general derivation of kinetic models.

Since $\hat{K}$ has been shown to be compact in certain cases, it is natural to make use of standard methods for the approximation of $\hat{K}$. From the formal viewpoint, this is most elegantly done in terms of a finite dyadic expansion in the eigenfunctions of $\hat{K}$. However, except for $(1, \xi, \xi^2)$ (which are eigenfunctions), nothing else is known of the eigenfunctions of $\hat{K}$. Therefore for practical reasons, it is more advantageous to expand in terms of known functions. We will also consider $\nu^{-1}k$, (which is compact with respect to $\hat{h}$) instead of $\hat{K}$.

Denote by $\chi_n(\xi)$ the orthonormal set of polynomials (say generated by the Gram–Schmidt process) which form a basis in $\hat{h}$, i.e.,

$$\langle \chi_n, \chi_m \rangle = \delta_{nm}.$$

For convenience we take χ_i, $i = 1, 5$, to be constructed from $(1, \xi, \xi^2)$. We formally write

$$g = \sum_{n=1}^{\infty} a_n\chi_n, \tag{3.2}$$

with

$$a_n = \langle \chi_n, g \rangle. \tag{3.3}$$

Next we define the projection operator

$$P_N = \sum_{n=1}^{N} \chi_n\langle \chi_n, \rangle. \tag{3.4}$$

Then considering (2.1) and (2.3), we define the kinetic models

$$\begin{aligned} Dg + \nu(\xi)g &= \nu P_N\nu^{-1}kP_Ng \\ &= k_Ng = \nu \sum_{m,n\le N} a_n\kappa_{nm}\chi_m, \end{aligned} \tag{3.5}$$

with

$$\kappa_{mn} = (\chi_m, k\chi_n).$$

(Note since k_N is finite-dimensional, it is compact with respect to $\hat{h}$ and h.)

It should be noted that κ_{mn} is defined in terms of the inner product in h, and not $\hat{h}$. This is a great convenience since the κ_{mn} are then closely related to the so-called bracket integrals for which there is a large literature.[3,25]

The symmetry of k_N under the inner product (2.5) now follows from the symmetry of k. For

$$\begin{aligned} (f, k_Ng) &= \langle f, P_N\nu^{-1}kP_Ng \rangle = \langle P_Nf, \nu^{-1}kP_Ng \rangle \\ &= (P_Nf, kP_Ng) = (kP_Nf, P_Ng) = \langle \nu^{-1}kP_Nf, P_Ng \rangle \\ &= (k_Nf, g). \end{aligned}$$

From these, one easily obtains that $\nu - k_N$ is non-negative, and the remaining Boltzmann-like property

$$(\nu - k_N)(1, \xi, \xi^2) = 0$$

has already been built into the models for $N \ge 5$.

For $N = 5$,

$$Dg = Cg = \nu(\xi)\sum_{i=1}^{5} \langle \chi_i, g \rangle\chi_i, \tag{3.6}$$

which is the model introduced and studied by Cercignani.[8,26]

[25] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954).

[26] C. Cercignani, Ann. Phys. (N.Y.) **40**, 469 (1966).

For later purposes we exhibit the simplest non-trivial model; i.e., if $N = 1$, we have

$$Dg + \nu(\xi)g = \frac{\nu(\xi)}{\nu_0}\int \nu\Omega g \, d\boldsymbol{\xi},$$

$$\nu_0 = \int \nu\Omega \, d\xi. \tag{3.7}$$

This equation, which preserves only the continuity equation, is a generalization of the isosteric model discussed in I.

4. SPECTRUM OF THE LINEAR BOLTZMANN EQUATION AND MODELS

The purpose of this section is two-fold. On the one hand we wish to examine the continuous spectrum of the models of the last sections. Secondly we wish to compare this with the spectrum of the exact linear Boltzmann. The agreement of these should lend further support to the use of the models in the next two sections.

The spectral problem for the Boltzmann equation has already been considered by Ford[14] and Grad,[15] and for neutron diffusion by Corngold.[12] These previous discussions are, however, strictly formal. Below, we also give a mathematically rigorous discussion for the Boltzmann equation, as well as for the model equations. Although the rigorous analysis is more complete than the previous studies, it does not reveal any significant differences from these formal discussions.[12,14,15] Other discussions of the spectrum for certain kinetic models have appeared in the course of problem solving.[16–19]

To initiate the discussion we consider

$$\left(\frac{\partial}{\partial t} + \boldsymbol{\xi}\cdot\frac{\partial}{\partial \mathbf{x}} + \nu\right)\tilde{G} = K\tilde{G}, \tag{4.1}$$

where we leave open the nature of ν and of the linear operator K. We formally write

$$\tilde{G} = e^{\sigma t + s x}G(\xi)$$

(simple arguments show that there is no loss of generality in assuming 1-dimensionality for either of the two problems discussed below),so that

$$(\sigma + s\xi_1 + \nu)G = KG.$$

The two main problems which naturally arise are: (1) the initial-value problem for which $s\,(= i\kappa)$ is pure imaginary, and (2) the boundary-value problem (actually steady-state oscillations) for which $\sigma\,(= i\omega)$ is pure imaginary.

We therefore write

$$\sigma G = (K - i\kappa\xi_1 - \nu)G = L_i G, \tag{4.2}$$

and

$$sG = \left(\frac{1}{\xi_1}K - \frac{i\omega + \nu}{\xi_1}\right)G = L_b G, \tag{4.3}$$

for the initial- and boundary-value problems, respectively. Our aim is then to discuss the spectrum of the operators L_i and L_b. Aspects of the discrete spectra are discussed in Secs. 5 and 6. The present discussion will be devoted to the continuous spectrum.

We will make repeated use of the following generalization of a theorem by Weyl, given by Kato in his recent book.[27]

Theorem: If T is a closed operator from a Banach space to itself, and A is compact, relative to T (i.e., A is T-compact), then T and $T + A$ have the same essential spectrum.

Actually, we are not justified in identifying the continuous spectrum with the essential spectrum. For example, a dense set of discrete eigenvalues and a continuous spectrum show up in the same way in the essential spectrum. Under the circumstances this seems unlikely and we will refer to the essential spectrum as the continuous spectrum. In all cases we assume that $\nu(\xi)$ is real, positive, and continuous for $\boldsymbol{\xi}\epsilon R^3$, and also that K is compact on H. (We could consider h equally well.)

We start our discussion by considering

$$1/[\nu(\xi) + i\kappa\xi_1].$$

This by the assumptions on ν, is bounded and hence defined on all H. Therefore it is closed, and hence its inverse

$$T_i = \nu(\xi) + i\kappa\xi_1 \tag{4.4}$$

is also closed. By the Weyl–Kato theorem, T_i and L_i have the same continuous spectrum.

Next consider

$$P = \xi_1/[\nu(\xi) - i\omega].$$

Since P is defined on all functions of H which vanish outside a set of finite measure, its domain $D(P)$ is dense in H. Hence its adjoint

$$P^* = \xi_1/[\nu(\xi) + i\omega]$$

exists and is closed. From this it follows that

$$(P^*)^{-1} = T_b = [i\omega + \nu(\xi)]/\xi_1 \tag{4.5}$$

is closed.

[27] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, New York, 1966). For the result quoted see Theorem 5.35. For definitions and other results used in this section, see Chap. IV of this book.

We next wish to show that $(1/\xi_1)K$ is T_b compact. To this end it is convenient to introduce the Banach space B, defined by the graph norm

$$\|u\|_T = \{\|u\|^2 + \|T_b u\|^2\}^{\frac{1}{2}},$$

where $\|\ \|$ refers to the norm in H. The compactness of K in H is not sufficient to induce the compactness of $(1/\xi_1)K$ on bounded sets of B. [For example, if $\phi \notin D(\xi_1^{-1})$, then $\phi\mathfrak{L}$, with $\mathfrak{L}$ a linear functional, is compact and $\xi_1^{-1}\phi\mathfrak{L}$ does not even exist.] The additional property of K which we use is that for $u = \xi_1 w$ with $w \in H$,

$$K\xi_1 w = \xi_1 y, \tag{4.6}$$

with $\|y\| \in H$. Since K is compact, this is only a condition at the origin. This property of K is an immediate consequence of the isotropy of the Boltzmann equation (Sec. 3) and its models (Sec. 4).

Let M denote a bounded set of functions of B, i.e., $\phi \in M$ implies

$$\|\phi\|_T < C,$$

when C is a constant. It is immediate that M is bounded in H. Hence we may choose a sequence $\{\phi_i\} \in M$ such that K maps it into a convergent sequence. Since $\|\phi_i\|_{T_b} < C$, it follows that we can write

$$\phi_i = \xi_1 \psi_i,$$

with ψ_i such that $\psi_i \in H$ (since $\|T_b\phi\|$ exists). Now consider

$$\left\| \frac{1}{\xi_1} K(\phi_i - \phi_j) \right\|^2 = \left[\frac{1}{\xi_1} K(\phi_i - \phi_j), \frac{1}{\xi_1} K(\phi_i - \phi_j) \right].$$

But by the isotropy condition (4.6), we have

$$K\phi_i = K\xi_i\psi_i = \xi_1 y_i,$$

with $y_i \in H$ is $\psi_i \in H$. Hence

$$\left\| \frac{1}{\xi_1} K(\phi_i - \phi_j) \right\|^2 = \|y_i - y_j\|^2.$$

The vanishing of the right-hand side as $i, j \to \infty$ is a consequence of

$$\lim_{i,j\to\infty} \|K(\phi_i - \phi_j)\| = 0,$$

and the boundedness of $\|y_i\|$. Hence K is T_b-compact and by the Weyl–Kato theorem, T_b and L_b have the same continuous spectrum.

It only remains for us to identify the above v and K with the various possibilities open to us. There are three cases to which the above analysis immediately applies. (1) Rigid-sphere molecules, since K is compact.[20] (2) Intermolecular potentials with angle cutoff, again since K is compact.[7] (3) The models of Sec. 3, since K is finite-dimensional and hence compact. It is highly plausible that the above discussion also applies to intermolecular potentials with radial cutoff. But thus far no compactness proof for K has been given.[28]

Since T_i and T_b are multiplicative operators their continuous spectra are given by their ranges when these are regarded as functions. Various situations are sketched in Figs. 1–6. The discussion of these has been given in the introduction and we do not repeat it.

5. LOW-FREQUENCY EXPANSIONS (CONSTANT COLLISION-FREQUENCY MODELS)

Using the notation of I, we first consider models of the form

$$\left(\frac{\partial}{\partial t} + \xi_1 \frac{\partial}{\partial x} + 1\right) g = kg = \sum_{m,n\le N} a_n \beta_{mn} \psi_m, \tag{5.1}$$

where

$$a_m = (\psi_m, g), \tag{5.2}$$

and then β_{mn} are constants. A plane wave is of the form

$$g = e^{\sigma t + s x} G(\xi), \tag{5.3}$$

where σ and s are in general complex.[1] One then easily shows that a necessary and sufficient condition for a plane wave is that

$$G = \frac{\sum_{m,n\le N} \alpha_n \beta_{nm} \psi_m}{\sigma + s\xi_1 + 1},$$

with

$$\alpha_r = (\psi_r, G),$$

such that

$$(\mathbf{1}s - \mathbf{C})\boldsymbol{\alpha} = 0, \tag{5.4}$$

where

$$C_{mn} = s \sum_{r\le N} \beta_{nr} \left(\psi_m, \frac{\psi_r}{1 + \sigma + s\xi_1} \right). \tag{5.5}$$

(The purpose of the factor of s will be seen shortly.) Equation (5.4) therefore implies that σ and s must satisfy

$$D = \det(\mathbf{1}s - \mathbf{C}) = 0. \tag{5.6}$$

This in turn implies a functional dependence of σ on s (or vice versa), and we shall focus attention on $\sigma = \sigma(s)$. In I it is shown, by direct construction, that $\sigma(s)$ is not analytic for $N = 1$ in (5.5). We now consider the general case.

The entries of $\mathbf{C}$, (5.5), are composed of terms of the type

$$C_{ij} = s \int \frac{\bar{\psi}_i \psi_j \Omega}{1 + \sigma + s\xi_1} \, d\boldsymbol{\xi}.$$

[28] For an interesting paper which casts doubt on this conjecture, see C. Cercignani, Phys. Fluids **10**, 2097 (1967).

Since the ψ_i are polynomials, each such term may be reduced to the form[29]

$$C_{ij}(z) = p_{ij}(z) + g_{ij}(z)F(z),$$

where p_{ij} and g_{ij} are polynomials and

$$F(z) = \int_{-\infty}^{\infty} \frac{e^{-(x^2/2)}}{(2\pi)^{\frac{1}{2}}} \frac{dx}{x - z},$$
$$z = -(1 + \sigma)/s. \tag{5.7}$$

We can therefore expand (5.6) in the form

$$D = \sum_{n=0}^{N} A_n(z, s)[F(z)]^n = 0, \tag{5.8}$$

where the A_n are polynomials in z and s.

Our primary interest is in the roots of D for which[30]

$$\sigma = s = 0, \tag{5.9}$$

which correspond to the hydrodynamical modes of propagation. (In Sec. 7 other branches are briefly discussed.) In this case $z \to \infty$, and we easily have that in this limit

$$F(z) \sim -\frac{1}{z}\left(1 + \frac{1}{z^2} + \frac{1 \cdot 3}{z^4} + \frac{1 \cdot 3 \cdot 5}{z^6} + \cdots\right). \tag{5.10}$$

Actually (5.7) defines two different analytic functions $F^+(z)$, for $\operatorname{Im} z > 0$, and $F^-(z)$, for $\operatorname{Im} z < 0$. Then a careful argument[29] shows (5.10) to be valid in a region larger than the half-plane given by $-\pi/4 < \arg z < 5\pi/4$ for F^+, and for $-5\pi/4 < \arg z < \pi/4$ for F^-.

If (5.10) is substituted into (5.8), the asymptotic expansion for $\sigma(s)$,

$$\sigma = b_1 s + b_2 s^2 + \cdots,$$

can be obtained. The explicit form for the b_r's has been discussed previously,[2,29,31] and in I we have shown that for $N = 1$ in (5.1), this expansion is divergent.

We now demonstrate that $\sigma(s)$ is not analytic for arbitrary N. To prove this we assume the contrary, i.e., that $\sigma = \sigma(s)$ is analytic in some small neighborhood of the origin of the s plane. Then, in the deleted neighborhood of the origin,

$$z(s) = -[1 + \sigma(s)]/s$$

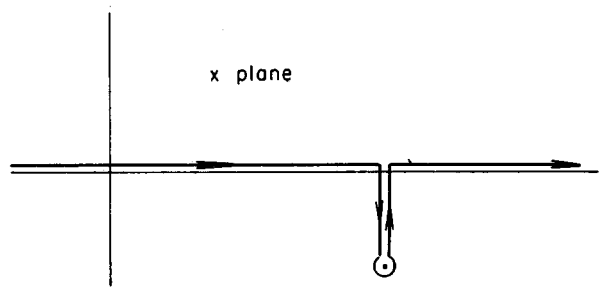


FIG. 7. Path of integration for $F^+(z)$, when $\operatorname{Im} z < 0$.

is analytic, and since $F^+(z)$ [and $F^-(z)$] is entire, $F^+[z(s)]$ is also analytic in the deleted neighborhood of the origin in the s plane. On first restricting attention to s such that $\operatorname{Im} z > 0$, we have by hypothesis that

$$D(F^+(z(s)); z(s), s) = 0. \tag{5.11}$$

But since this holds for a continuous point set, it is an identity in s and by analytic continuity holds in the entire neighborhood of the origin. For s such that $\operatorname{Im} z < 0$, the analytically continued form of $F^+(z)$ must be used. Clearly, this is obtained by choosing the path of integration for (5.7) as shown in Fig. 7. From this we immediately have

$$F^+(z) = F^-(z) + i(2\pi)^{\frac{1}{2}} e^{-z^2/2}.$$

Substituting (5.12) into (5.11) [and hence into (5.8)], and taking $s = -i\epsilon$ with $\epsilon \to 0$, we have a contradiction. For by (5.10), $F^-(z)$ is bounded as $z \to \infty$, and dividing (5.8) by $(F^+)^N$, the contradiction is shown.

Therefore, $\sigma = \sigma(s)$ is not analytic in the neighborhood of $s = 0$. [Hence $s = s(\sigma)$ is not analytic for $\sigma \sim 0$.]

6. LOW-FREQUENCY EXPANSIONS (VELOCITY-DEPENDENT COLLISION FREQUENCY MODELS)

A similar discussion now follows for velocity-dependent collision frequencies. For simplicity we start our discussion with the simple model, (3.7), already introduced in Sec. 3,

$$\left[\frac{\partial}{\partial t} + \xi_1 \frac{\partial}{\partial x} + \nu(\xi)\right] g = \frac{\nu}{\nu_0} \int \Omega \nu(\xi) g \, d\boldsymbol{\xi} = \frac{\nu}{\nu_0} \hat{\rho}. \tag{6.1}$$

We make the following assumptions on $\nu(\xi)$, all of which are in keeping with its interpretation of being a collision frequency:

$$\nu(\xi) \geq \nu(\xi = 0) = 1 > 0$$

and for ξ large

$$\nu(\xi) \sim \xi^{\alpha}, \quad 0 \leq \alpha \leq 1 \tag{6.2}$$

[29] The closed form for arbitrary $C_{ij}(z)$ is given explicitly in L. Sirovich and J. K. Thurber in *Rarefied Gasdynamics*, J. J. de Leeuw, Ed. (Academic Press Inc., New York, 1965), Vol. 1, p. 21.

[30] That (5.9) satisfies (5.8) follows from the properties of the β_{nm}. Actually, this follows from the collisional invariants of the collision integral.

[31] L. Sirovich and J. K. Thurber in *Rarefied Gasdynamics*, J. A. Laurmann, Ed. (Academic Press Inc., New York, 1963), Vol. 1, p. 159.

(the first condition is merely a normalization). We shall also assume that $\nu(\xi)$ may be piecewise analytically continued in the complex plane.

Repeating the discussion which led to (5.6), we now arrive at the dispersion relation

$$\pi(\sigma, s) = 1 - \frac{1}{\nu_0}\int_{-\infty}^{\infty} \frac{\Omega\nu^2(\xi)\, d\xi}{\sigma + s\xi_1 + \nu(\xi)} = 0, \quad (6.3)$$

and we seek $\sigma = \sigma(s)$ such that $\sigma(0) = 0$. [It is clear that since (6.1) preserves continuity, $\sigma = s = 0$ is a root of (6.3).] The formal expansion

$$\sigma = b_1 s + b_2 s^2 + \cdots$$

may be obtained directly from (6.3). For we may write

$$\begin{aligned}\hat{\mathcal{F}} &= \int_{-\infty}^{\infty} \frac{\nu}{\nu_0} \frac{\Omega\nu}{\sigma + s\xi_1 + \nu}\, d\xi \\ &= \int_{-\infty}^{\infty} \frac{\nu}{\nu_0} \frac{\Omega}{1 + (\sigma + \xi_1 s)/\nu}\, d\xi \\ &= \int_{-\infty}^{\infty} \frac{\nu}{\nu_0}\Big[1 - \frac{\sigma + s}{\nu} \pm \cdots + \Big(-\frac{\sigma + s\xi_1}{\nu}\Big)^n \\ &\quad + \frac{[-(\sigma + s\xi_1)/\nu]^{n+1}}{1 + (s\xi_1 + \sigma)/\nu}\Big]\Omega\, d\xi \\ &= 1 - \frac{\sigma}{\nu_0}\int_{-\infty}^{\infty}\Omega\, d\xi - \frac{s}{\nu_0}\int_{-\infty}^{\infty}\xi_1\Omega\, d\xi + \cdots \\ &\quad + \int\Big(-\frac{\sigma + s\xi_1}{\nu}\Big)^n \frac{\nu}{\nu_0}\,\Omega\, d\xi + O[(|\sigma| + |s|)^{n+1}].\end{aligned} \quad (6.4)$$

From this we immediately obtain

$$\sigma = \frac{s^2}{\nu_0}\int_{-\infty}^{\infty} \frac{\Omega\xi_1^2}{\nu}\, d\xi + O(s^4). \quad (6.5)$$

For the further investigation of $\sigma(s)$, we first reduce (6.4). Introducing spherical coordinates, we have, after some straightforward manipulations,

$$\begin{aligned}\hat{\mathcal{F}} &= \frac{1}{s}\mathcal{F} \\ &= \frac{1}{\nu_0 s(2\pi)^{\frac{1}{2}}}\int_0^{\infty} xe^{-x^2/2}\nu^2(x) \ln\Big[\frac{\sigma + sx + \nu(x)}{\sigma - sx + \nu(x)}\Big]\, dx,\end{aligned} \quad (6.6)$$

and the dispersion relation (6.3) is

$$s - \mathcal{F} = 0. \quad (6.7)$$

As was the case with (5.7), more than one analytic function is defined by (6.6). Now, however, the different functions are determined by the location of the branch points of the logarithm in (6.6). For σ and s small, $\sigma + sx + \nu(x)$ has a branch point x_+ such that

$$x_+ \sim (-s)^{-1/(1-\alpha)}, \quad (6.8)$$

and $\sigma - sx + \nu(x)$ a branch point x_-, such that

$$x_- \sim s^{-1/(1-\alpha)}. \quad (6.9)$$

In addition to (6.8) and (6.9) there are branch points which are a distance, at most, $O(1)$ from the origin.

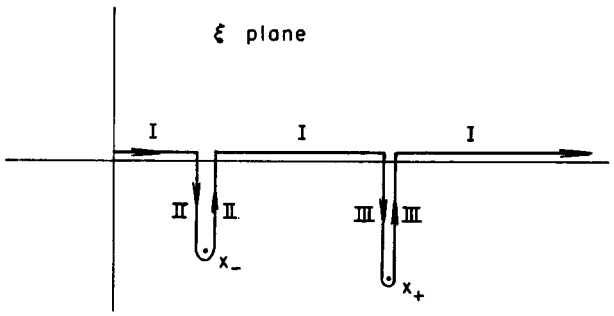


FIG. 8. Path of integration for $\mathcal{F}$.

Case 1: $\alpha < 1$

In this case the additional branch points are immaterial as we shall see.

Assume that $\sigma = \sigma(s)$ is analytic in a neighborhood of the origin. Inserting this into the dispersion relation (6.7), we have

$$\begin{aligned}0 = s &- \frac{1}{\nu_0(2\pi)^{\frac{1}{2}}} \\ &\times \int_0^{\infty} xe^{-x^2/2}\nu^2(x) \ln\Big[\frac{\sigma(s) + sx + \nu(x)}{\sigma(s) - sx + \nu(x)}\Big]\, dx.\end{aligned} \quad (6.10)$$

Let us suppose that s in (6.10) is such that both x_+ and x_- are in the upper half-plane.[32] We will denote the integral in (6.10) by $\mathcal{F}^+$. Then since $\sigma(s)$ is analytic by assumption, and $\mathcal{F}^+(\sigma, s)$ is analytic, (6.10) is an identity in s, and by continuation it holds in a neighborhood of the origin, $s = 0$. If the branch points pass through the positive-real axis, the path of integration shown in Fig. 8 must be used for the integral in (6.10). [That both x_+ and x_- can be made to pass through $\xi > 0$ is clear from (6.8) and (6.9).] In this case, i.e., when x_- and x_+ are in the lower half-plane, as shown in the figure, we immediately have that

$$\begin{aligned}\mathcal{F}^+(\sigma(s), s) = \mathcal{F}^-(\sigma(s), s) &+ i\frac{(2\pi)^{\frac{1}{2}}}{\nu_0}\int_{\mathrm{Re}\, x_+}^{x_+} xv^2(x)e^{-x^2/2}\, dx \\ &+ i\frac{(2\pi)^{\frac{1}{2}}}{\nu_0}\int_{\mathrm{Re}\, x_-}^{x_-} xv^2(x)e^{-x^2/2}\, dx,\end{aligned} \quad (6.11)$$

[32] The configuration of x_+ and x_- depends on α, e.g., if $\alpha = 0$ then $x_+ = -x_-$ and the branch points are in opposite quadrants. Our subsequent discussion is uneffected by the actual configuration of x_+ and x_- and a choice is solely made for the purpose of illustration.

where $\mathcal{F}^-$ denotes the integral in (6.10) when x_+ and x_- are in the lower half-plane. (As pointed out in Footnote 32, the actual choice of a configuration of x_+ and x_- is immaterial.) But $\mathcal{F}^-$ is bounded as is shown by (6.4), whereas simple estimates show that the last two integrals in (6.11) can be made unbounded by allowing $s \to 0$. (Since $x_+ \neq x_-$, these two integrals do not identically cancel.) Hence we are lead to a contradiction and $\sigma(s)$ is not analytic [similarly, $s(\sigma)$ is not analytic].

Case 2: $\alpha = 1$

In this case the previous argument fails since x_+ and x_- are no longer in the neighborhood of infinity. In this case a direct estimate of the terms in (4.4) shows that the infinite expansion converges for $(|\sigma| + |s|) < 1$ and hence $\mathcal{F}(\sigma, s)$ is analytic in the neighborhood of $\sigma = 0$, $s = 0$. Furthermore $\partial\mathcal{F}(0, 0)/\partial\sigma \neq 0$, and hence by the implicit function theorem, $\sigma(s)$ is analytic for $s \sim 0$. [Since $\partial\mathcal{F}(0, 0)/\partial s = 0$ and $\partial\mathcal{F}(0, 0)/\partial s^2 \neq 0$, s is an analytic function of $\sigma^{\frac{1}{2}}$.] We mention in passing that if $\alpha > 1$, analyticity is again obtained; however, such values of α seem to be unphysical.

Therefore, we have proven that for plane wave propagation as described by (6.1), $\sigma(s)$ is not analytic for $0 \leq \alpha < 1$, and is analytic for $\alpha = 1$.

Finally we point out that exactly the same formalism as was used in Sec. 5 applies to the general velocity-dependent collision frequency models discussed in Sec. 3. Further, it is clear that the analysis of this section applies directly to these general models. Therefore, for the models discussed in Sec. 3, $\sigma(s)$ is not analytic for collision frequencies such that

$$\nu(\xi) = 0(\xi^\alpha) \quad \text{as} \quad \xi \to \infty \quad \text{for} \quad 0 \leq \alpha < 1,$$

and analytic when $\alpha = 1$.

7. DISCUSSION

It has been pointed out that modes of propagation other than the hydrodynamical ones exist.[2,33,34] For example, if we set $s = 0$ in (5.6), we obtain N real nonpositive values of σ. Each of these yield a new branch $\sigma(s)$. Then provided $\sigma(0) \neq -1$, the same arguments used in Secs. 5 and 6 again apply, and $\sigma(s)$ is not analytic for $\alpha < 1$, and analytic for $\alpha = 1$. [And similarly for $s = s(\sigma)$.]

In I we showed that for the $N = 1$ model, (5.6), that $\sigma(s)$ is not analytic by the actual construction of the series for $\sigma = \sigma(s)$. In Secs. 5 and 6 we show that $\sigma = \sigma(s)$ is not analytic for $\alpha < 1$ by nonconstructive means. This leaves a gap in our discussion, for the asymptotic expansion of $\sigma(s)$ may still be convergent. [For example, if $D(\sigma, s) = \sigma - s - e^{-1/|s|}$, then $D(0, 0) = 0$, and $D(\sigma, s) = 0$ has the asymptotic root $\sigma \sim s$, whereas the actual root is certainly not analytic.] In view of the construction given in I, however, it is unlikely that this is actually the case.

Throughout our discussion we have only considered collision frequencies which are bounded away from zero. This assumption is of course based on the discussion in Sec. 2 of the actual forms which the collision frequency can take. For completeness, and since they have appeared in neutron diffusion, we briefly comment on velocity-dependent collision frequencies which can vanish. In this case, the continuous spectrum of the initial-value problem touches the origin. This immediately signals the non-analyticity of $\sigma(s)$ [or $s(\sigma)$]. It is interesting to note that this can now arise because of the slow-moving molecules [e.g., if $\nu(\xi) = 0(\xi)$ for ξ small] instead of the fast molecules as was the case in Secs. 5 and 6.

An interesting connection with the Chapman–Enskog procedure may be pointed out. In the present context the Chapman–Enskog procedure may be understood to be a series expansion for the distribution function in ik or in $i\omega$. Under mild conditions on the data of a problem, it is seen that this expansion is convergent if the continuous spectrum does not pass through the origin (for either the boundary- or initial-value problems). Otherwise it is at most an asymptotic expansion. It may even fail to be asymptotic if the continuous spectrum dominates the discrete modes. This, for example, is the case for forced oscillations when $\nu = 0(\xi^\alpha)$, $\alpha < 1$.

ACKNOWLEDGMENT

The results presented in this report were obtained in the course of research sponsored by the office of Naval Research under Contract Nonr 562(39) with Brown University, Providence, R.I., and by Brookhaven National Laboratories.

[33] L. Sirovich, Phys. Fluids **6**, 218 (1963).
[34] L. Sirovich, Phys. Fluids **6**, 1428 (1963).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Closed-Form Solution of the Differential Equation $\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0$ Subject to the Initial Condition $P(x, y, t = 0) = \Phi(x, y)$

J. L. NEURINGER
AVCO, Missile Systems Division, 201 Lowell Street, Wilmington, Massachusetts

(Received 25 March 1968)

A closed-form solution of the differential equation

$$\left(\frac{\partial^2}{\partial x \partial y} + ax\frac{\partial}{\partial x} + by\frac{\partial}{\partial y} + cxy + \frac{\partial}{\partial t}\right)P = 0,$$

subject to the initial condition $P(x, y, t = 0) = \Phi(x, y)$, is presented.

1. INTRODUCTION

Lambropoulos[1] presented a solution of the above initial-value problem in the form of a series of powers of two of the variables (x and y) with coefficients depending on the third (t). The solution obtained is very cumbersome and even in those cases where some or all of the constants a, b, and c are zero, the solution involves complicated triple infinite sums. It is the intent here to obtain a formal closed-form solution of the above system by first transforming the dependent and independent variables to a new set of variables, so that the system is separable in the new set, and then using standard Fourier-transform techniques to obtain the solution. Lambropoulos' claim that it does not seem feasible to solve the system by conventional methods is debatable, as the following demonstrates.

2. TRANSFORMATION OF THE SYSTEM

We seek a formal solution of

$$\frac{\partial P}{\partial t} + ax\frac{\partial P}{\partial x} + by\frac{\partial P}{\partial y} + cxyP + \frac{\partial^2 P}{\partial x \partial y} = 0, \quad (1)$$

subject to the initial condition

$$P(x, y, 0) = \Phi(x, y). \quad (2)$$

Introduce a new dependent variable Q defined by

$$P(x, y, t) = e^{\alpha xy}Q(x, y, t), \quad (3)$$

where the unknown parameter α is so chosen as to eliminate the product variable-coefficient term $cxyP$. Substituting Eq. (3) into the system of Eqs. (1) and (2) yields

$$\frac{\partial Q}{\partial t} + (a + \alpha)x\frac{\partial Q}{\partial x} + (b + \alpha)y\frac{\partial Q}{\partial y} + \alpha Q + \frac{\partial^2 Q}{\partial x \partial y} = 0, \quad (4)$$

subject to

$$Q(x, y, 0) = e^{-\alpha xy}\Phi(x, y), \quad (5)$$

where

$$\alpha = \frac{-(a + b) \pm [(a + b)^2 - 4c]^{\frac{1}{2}}}{2} \quad (6)$$

The appropriate sign to be used in Eq. (6) depends on the convergence of the solution which, in turn, depends on the values of the constants (see Sec. 3).

Now, Eq. (4) is a linear second-order partial differential equation with variable coefficients. The variable coefficients correspond to variables associated with a second-order differential coefficient. If we are able to effect a transformation to new independent variables, so that the new differential equation contains at most only variable coefficients associated with the first-order differential coefficient t, then the resulting equation is separable and hence much easier to solve.

Introducing new independent variables, X, Y, and T defined by

$$X = xe^{-(a+\alpha)t}, \quad Y = ye^{-(b+\alpha)t}, \quad T = t, \quad (7)$$

then

$$\frac{\partial Q}{\partial t} = \frac{\partial Q}{\partial T} - (a + \alpha)xe^{-(a+\alpha)t}\frac{\partial Q}{\partial X} - (b + \alpha)ye^{-(b+\alpha)t}\frac{\partial Q}{\partial Y},$$

$$\frac{\partial Q}{\partial x} = e^{-(a+\alpha)t}\frac{\partial Q}{\partial X},$$

$$\frac{\partial Q}{\partial y} = e^{-(b+\alpha)t}\frac{\partial Q}{\partial Y},$$

$$\frac{\partial^2 Q}{\partial x \partial y} = e^{-(a+b+2\alpha)t}\frac{\partial^2 Q}{\partial X \partial Y}.$$

Substituting into the system of Eqs. (4) and (5) yields

$$\frac{\partial Q}{\partial t} + \alpha Q + e^{-(a+b+2\alpha)t}\frac{\partial^2 Q}{\partial X \partial Y} = 0, \quad (8)$$

[1] P. Lambropoulos, J. Math. Phys. 8, 2167 (1967).

subject to

$$Q(X, Y, 0) = e^{-\alpha XY}\Phi(X, Y) = e^{-\alpha xy}\Phi(x, y), \quad (9)$$

since $X = x$; $Y = y$, when $t = 0$.

It is noted that the variable coefficient in Eq. (8) is a function at most of the variable t only as required.

3. SOLUTION

Equation (8) is separable. Seeking a solution of the form

$$Q = \tau(t)F(X, Y), \quad (10)$$

substituting into Eq. (8), and deliberately setting the separation constant equal to $k_x k_y$, yields

$$\frac{d\tau}{dt} + (\alpha - k_x k_y e^{-(a+b+2\alpha)t})\tau = 0, \quad (11)$$

and

$$\frac{\partial^2 F}{\partial X \partial Y} + k_x k_y F = 0. \quad (12)$$

Integrating Eq. (11) yields

$$\tau = \exp\left\{-\alpha t - \frac{k_x k_y}{a + b + 2\alpha} \times \exp\left[-(a + b + 2\alpha)t\right]\right\}, \quad (13)$$

and when $t = 0$,

$$\tau(0) = \exp\left(-\frac{k_x k_y}{a + b + 2\alpha}\right). \quad (14)$$

Integrating Eq. (12) yields

$$F(X, Y) = e^{-ik_x X}e^{-ik_y Y}. \quad (15)$$

Substituting the solutions, Eqs. (13) and (15), back into Eq. (10) yields

$$Q(X, Y, t) = e^{-\alpha t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{k_x k_y}{a + b + 2\alpha}e^{-(a+b+2\alpha)t}\right] \times \Theta(k_x, k_y)e^{-ik_x X}e^{-ik_y Y}\,dk_x\,dk_y, \quad (16)$$

where $\Theta(k_x, k_y)$ is an unknown function of k_x, k_y to be determined from the initial condition. When $t = 0$, we obtain

$$e^{-\alpha XY}\Phi(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\frac{k_x k_y}{a + b + 2\alpha}\right) \times \Theta(k_x, k_y)e^{-ik_x X}e^{-ik_y Y}\,dk_x\,dk_y. \quad (17)$$

From Eq. (17), and under the assumption that the Fourier transform exists, it is recognized that

$$\exp\left(-\frac{k_x k_y}{a + b + 2\alpha}\right)\Theta(k_x, k_y)$$

is the Fourier transform of $e^{-\alpha XY}\Phi(X, Y)$. Call it $\Psi(k_x, k_y)$. Hence,

$$\Theta(k_x, k_y) = \exp\left(\frac{k_x k_y}{a + b + 2\alpha}\right)\Psi(k_x, k_y).$$

Substituting for $\Theta(k_x, k_y)$ into Eq. (16) and returning to the old dependent and independent variables, we obtain

$$P(x, y, t) = e^{\alpha(xy-t)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Psi(k_x, k_y) \times \exp\left\{\frac{k_x k_y}{a + b + 2\alpha}\left[1 - e^{-(a+b+2\alpha)t}\right] - i[k_x x e^{-(a+\alpha)t} + k_y y e^{-(b+\alpha)t}]\right\} dk_x\,dk_y, \quad (18)$$

where

$$\Psi(k_x, k_y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\alpha xy}\Phi(x, y)e^{ik_x x}e^{ik_y y}\,dx\,dy,$$

and α, from Eq. (6), is either α_1 or α_2, where α_1 is associated with the top sign, and α_2 with the bottom sign.

It is easily verified by direct substitution of Eq. (18) into Eq. (1), that it satisfies the differential equation and the initial condition Eq. (2).

It was previously mentioned in passing that the behavior of the solution depends on the particular choice of the parameter α whose value, in turn, depends on the values assigned to the constants a, b, and c. As an example, let us consider the particular case $a > 0$, $b > 0$, and $c < 0$. For this case, α_1 is positive and α_2, negative. Also, the linear expression $(a + b + 2\alpha)$ appearing in the solution yields $(a + b + 2\alpha_1) = [(a + b)^2 - 4c]^{\frac{1}{2}}$ and is positive, and $(a + b + 2\alpha_2) = -[(a + b)^2 - 4c]^{\frac{1}{2}}$ and is negative. Hence, with α_1, used in Eq. (18), the exponential time factors both outside and inside the integral both approach zero when $t \to \infty$, while they both approach infinity when α_2 is used. With $\Psi(k_x, k_y)$ preassigned, whether neither or both of the individual solutions (corresponding to α_1 or α_2) converge for large time depends, of course, on the limiting form that Eq. (18) takes once the integration is performed.

Finally, we list without derivation two particular integrals of Eq. (1), corresponding, respectively, to the two roots α_1 and α_2:

$$P_1(x, y, t) = \frac{e^{-\alpha_1 t}}{[e^{-\beta t} - 1]}\exp\left\{-\left[\frac{\beta e^{-\beta t}}{e^{-\beta t} - 1} - \alpha_1\right]xy\right\}, \quad (19)$$

$$P_2(x, y, t) = \frac{e^{-\alpha_1 t}}{[e^{-\beta t} - 1]} \times \exp\left\{-\left[\frac{-\beta e^{\beta t}}{e^{\beta t} - 1} - \alpha_1 + \beta\right]xy\right\}, \quad (20)$$

where

$$\beta = [(a + b)^2 - 4c]^{\frac{1}{2}}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Inner and Restriction Multiplicity for Classical Groups

R. M. DELANEY AND B. GRUBER
Department of Physics, Saint Louis University, St. Louis, Missouri

(Received 28 December 1967)

For the classical compact Lie groups G, a formula for the multiplicity of weights (called inner multiplicity) is given. This formula relates the inner multiplicity of a group G to the inner multiplicity of a naturally embedded subgroup G'. For the $SU(n)$ groups the formula can be brought into a particularly simple form—namely, a sum over Kronecker symbols—by choosing the group $SU(2)$ for G'. The multiplicity of irreducible representations of a subgroup G' into which an irreducible representation of a group G decomposes if G is restricted to G'—called restriction multiplicity of G with respect to G'—is related to the inner multiplicity of the group G.

1. INTRODUCTION

The multiplicity of weights for the classical compact Lie groups G—called inner multiplicity[1]—can be calculated by various methods. One way is by making use of the elegant formulas given by Kostant[2] and Freudenthal.[3] These formulas are completely general and hold for all classical compact Lie groups. Another approach to the problem has been proposed by Racah.[4] Racah points out that the inner multiplicity of the weights of an irreducible representation can be obtained by solving a set of difference equations. For the group $SU(4)$ Wigner[5] has obtained recurrence relations which allow the calculation of the inner multiplicity of weights of $SU(4)$.

While all these approaches permit in principle the calculation of the inner multiplicity of weights, none of them is really simple enough to allow an easy calculation, except for groups of rank two, and even then not too large representations of them.[6,7] [In fact, it is only Kostant's formula which allows a simple calculation of the inner multiplicity of $SU(3)$ for any representation.[8,9] Of course, $SU(2)$ is trivial.] Kostant's formula rests on the knowledge of the partition function $\bar{M}$ which has been calculated so far for the rank two groups and $SU(4)$ only.[10] The partition function has a complicated structure which has prevented its calculations for other groups. A simplification of this problem has been achieved by relating the partition function for any of the classical compact Lie groups G to the very simple partition function of $SU(3)$.[6] This simplification allows one to calculate the values of the partition function for groups of not too high rank. (Of course, the higher the rank the more complicated the partition function, and one cannot expect simple results.) In spite of the knowledge of the partition function for the groups of low rank or at least the possibility of calculating its values, Kostant's formula is, as was already mentioned, in general not easy to use [except for $SU(3)$ and $SU(2)$]. This is due to the fact that Kostant's formula involves a summation over the elements S of the Weyl group W. This implies that the value of the partition function $\bar{M}$ has to be calculated for each element of this sum. Thus, for instance, in the case of $SU(4)$ one has to calculate the value of the partition function $\bar{M}$ twenty-four times in order to obtain the multiplicity of one single weight. (Some of these terms are always zero. This simplifies the problem ,however, only slightly.[7])

Freudenthal's formula,[3] on the other hand, is a recurrence formula. In order to obtain the multiplicity for a particular weight, in general one has to calculate the multiplicities of other weights. Namely, all the multiplicities of the weights have to be known which can be reached from the weight under consideration by positive multiples of the positive roots. It is clear from this that, even for groups of low rank, this involves considerable work except for the simplest irreducible representations.

Also, the method of calculating the inner multiplicity by means of difference equations, which was developed by Racah, is, unfortunately, not simple. Wigner's results apply to $SU(4)$ and its generalization to higher rank groups does not seem to be easy.

Thus in spite of the presence of the formulas and methods quoted, it seems to be desirable to develop, if possible, a simpler method for the calculation of the

[1] A. J. Macfarlane, L. O'Raifeartaigh, and P. S. Rao, J. Math. Phys. **8**, 536 (1967).
[2] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1965), p. 261.
[3] N. Jacobson, *Lie Algebras* (Interscience Publishers, Inc., New York, 1965), p. 247.
[4] G. Racah, *Group Theoretical Concepts and Methods in Elementary Particle Physics*, F. Gursey, Ed. (Gordon and Breach, Science Publishers, New York, 1962).
[5] E. Wigner, Phys. Rev. **51**, 106 (1937).
[6] B. Gruber and F. Zaccaria, Nuovo Cimento Suppl. **5**, 914 (1967).
[7] B. Gruber, Nuovo Cimento **48**, 23 (1967).
[8] B. Gruber and T. S. Santhanam, Nuovo Cimento **45A**, 1046 (1966).
[9] B. Gruber and H. J. Weber, Proc. Roy. Irish Acad. **66A**, 31 (1968).
[10] J. Tarski, J. Math. Phys. **3**, 569 (1963).

inner multiplicity. This is also desirable since the "outer" multiplicity[1] of the Clebsch–Gordan series is related to the inner multiplicity and thus can be calculated from known inner multiplicities.[1,11–14]

There exists, however, a rather simple method for the calculation of the inner multiplicity of weights of $SU(n)$ by means of Gel'fand patterns.[8] This method for the calculation is implicitly contained in an article by Baird and Biedenharn[15] and consists in a prescription for counting all distinct Gel'fand patterns which belong to the same weight. It is the intention of the first part of this paper to supplement this prescription with a *formula*, from which, in turn, the prescription can be read off. Moreover, the formula obtained holds for all classical compact Lie groups and it becomes restricted to $SU(n)$ only when choosing the particular chain of subgroups, $SU(n) \supset U(1) \otimes SU(n-1)$.[15a] While in Sec. 2 this formula is derived from general properties of the characters and the Weyl group, it is shown in Sec. 3 that for $SU(n)$ this formula can be derived from the properties of the Gel'fand patterns alone. This latter derivation is particularly useful since it avoids the explicit use of properties of weight diagrams and the Weyl group. Moreover, in Sec. 3, there is given a precise prescription for the calculation of the inner multiplicities for the $SU(n)$ groups which is based on the derived formula.

Finally, in Sec. 4, a relation between restriction multiplicity and inner multiplicity is derived for the classical compact Lie groups G. This relation then allows the calculation of the restriction multiplicity from known inner multiplicities. In order to demonstrate the usefulness of this formula, the Weyl branching law for $SU(3)$ is rederived from it. Moreover, an expression for the upper limit for the multiplicity of $SU(2)$ multiplets (irreducible representations) in a given $SU(3)$ (irreducible) representation is obtained.

Before concluding this section it should be remarked that the subgroups G' are subgroups which are naturally embedded into the group G. This means their root diagram is part of the root diagram of G. Thus for instance, for $SU(n)$ the inner multiplicity can be related to the inner multiplicity of the subgroup $G' = SU(n-1)$ by restricting $SU(n)$ to the subgroup $G^{l-1} = SU(n-1) \otimes U(1)$, $n = l + 1$. Thereby $SU(n-1)$ is naturally embedded in $SU(n)$.

2. INNER MULTIPLICITY

In this section a relation is derived between the inner multiplicity of a group G, the inner multiplicity of a subgroup G' of G, and the restriction multiplicity of irreducible representations of G' occurring in an irreducible representation of G if G is restricted to G'. For the $SU(n)$ groups this relation is utilized to reduce the inner multiplicity of $SU(n)$ weights to the inner multiplicity of $SU(2)$ weights, i.e., essentially to a sum over Kronecker symbols [the inner multiplicity in $SU(2)$ is either zero or one]. However, before doing this it is necessary to introduce the notation first.

Let G denote a classical compact Lie group. Its irreducible representation, whose highest weight is $M = (M_1, M_2, \cdots, M_l)$, l being the rank of G, is denoted by $D(M)$. $m = (m_1, m_2, \cdots, m_l)$ denotes an arbitrary weight of G. The (inner) multiplicity of a weight m is denoted by $\gamma(m)$ and, if necessary for reasons of clarity, by $\gamma^M(m)$. Thus $\gamma^M(m)$ is the multiplicity of the weight m of $D(M)$.

G' denotes a subgroup of G which is embedded naturally in G. As was said, this means that the root system of G' forms a subsystem of the roots of G. Thus, the subgroup G' is either a simple compact Lie group or a direct product of simple compact Lie groups. $D(M')$ denotes an irreducible representation of the group G'. If G' is a direct product of simple groups, then M' is simply the direct sum of the highest weights M' of the simple groups. [This means, for instance, that if G' is the direct product of two simple groups $\bar{G}$ and $\bar{\bar{G}}$ with rank r and t, respectively, a weight m' of G' is given as $m' = (\bar{m}_1, \bar{m}_2, \cdots, \bar{m}_r; \bar{\bar{m}}_1, \bar{\bar{m}}_2, \cdots, \bar{\bar{m}}_t)$, where $\bar{m}$ is a weight of $\bar{G}$ and $\bar{\bar{m}}$ a weight of $\bar{\bar{G}}$.] The rank of the group G' is defined to be the sum of the ranks of its simple subgroups. Again, $\gamma(m')$ or $\gamma^{M'}(m')$ denotes the multiplicity of the weight m' contained in $D(M')$.

The restriction multiplicity is denoted by $\tilde{\gamma}(M')$ or $\tilde{\gamma}^M(M')$. Thus $\tilde{\gamma}(M')$ is the number of times the irreducible representation $D(M')$ of the subgroup G' occurs in an irreducible representation $D(M)$, if G is restricted to G'. Under this restriction a weight m of G goes over into a weight $m' = L(m)$ of G', where the L indicates that the weight m has to be restricted in order to become a weight of G'. In fact, L is a mapping of weights m of G onto weights m' of G'.

---

[11] D. Speiser, Ref. 4; J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).

[12] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **5**, 1730 (1964).

[13] B. Gruber, J. Math. Phys. **7**, 1797 (1966).

[14] B. Gruber, Ann. Inst. Henri Poincaré, **8**, 43 (1968).

[15] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963); L. C. Biedenharn (private communication).

[15a] *Note added in proof:* In fact the subgroup to be restricted to is the group $U(n-1)$. The group $SU(n-1) \otimes U(1)$ is locally isomorphic to the group $U(n-1)$ [see for instance, L. C. Biedenharn, *Lectures on Theoretical Physics, Cargèse, Summer School, Corsica, 1965*] and it should be kept in mind that throughout this article we are in fact dealing with the group $SU(n-1) \otimes U(1)/Z_{n-1}$ whenever the group $SU(n-1) \otimes U(1)$ appears in the text. We are indebted to Professor Biedenharn for a discussion on this point.

For the $SU(n)$ groups the $(l = n - 1)$-dimensional weight space is embedded in the familiar fashion into an n-dimensional Euclidean space R^n such that the weights are n-dimensional vectors

$$m = (m_1, \cdots, m_l, m_{l+1}), \tag{1}$$

for which the auxiliary condition holds

$$m_1 + m_2 + \cdots + m_l + m_{l+1} = 0. \tag{2}$$

The reason for doing this is twofold. For the components of the weights of $SU(n)$ we then have

$$m_i = k_i/(l + 1), \quad k_i = \text{integer}. \tag{3}$$

Moreover, the operation of the elements S of the Weyl group W on such a weight takes on a particularly simple form. The elements S of W simply permute the components m_i of the weight m. [See, for instance, Eq. (6) and Sec. 4.]

After these preliminaries it is now possible to go over to the actual derivation of the expression for the inner multiplicity.

The character $\chi(M)$ of an irreducible representation $D(M)$ of the group G is given as

$$\chi(M) = \sum_{m \in D(M)} \gamma(m) e^{i(m,\psi)}, \tag{4}$$

where (m, ψ) is the linear form

$$(m, \psi) \equiv m_1\psi_1 + m_2\psi_2 + \cdots + m_n\psi_n, \tag{5}$$

with $n = l$, except for $SU(l + 1)$, where $n = l + 1$ according to the convention made above. The ψ_i are the group parameters in the toroid T^l (also called the Cartan subgroup of G). The sum goes over all weights m of the irreducible representation $D(M)$, $\gamma(m)$ being their multiplicity. Another expression for the character, known as Weyl's formula, is given as

$$\chi(M) = \frac{\sum_{S \in W} \delta_S e^{i[S(M+R_0),\psi]}}{\sum_{S \in W} \delta_S e^{i(SR_0,\psi)}} \tag{6}$$

The sums in Eq. (6) go over all elements S of the Weyl group W; $\delta_S = -1$ if $S \in W$ is a reflection; $\delta_S = +1$ otherwise. The vector R_0 is one-half of the sum over the positive roots; i.e.,

$$R_0 = \frac{1}{2} \sum_{\alpha_i > 0} \alpha_i, \tag{7}$$

where the α_i denote the roots and $\alpha_i > 0$ indicates that the sum goes over the positive roots only.

Now, let G' be a subgroup of G which is naturally embedded. Then the root system of G' forms a subsystem of the root system of G. The root systems, however, determine, up to equivalence, the space of the toroids T^l and T^s of the groups G and G' respectively.[16,17] Since the root system of G' forms a subset of roots of G, so necessarily the space of T^s forms a subspace of T^l. This in turn implies that the subspace of T^s is given in terms of linear relations among the independent parameters ψ_i, $i = 1, \cdots, l$ of the toroid T^l. Under the restriction of G to G' the toroid T^l goes over into the toroid T^s of G'. Thus under this restriction the parameters $\psi_i, i = 1, \cdots, l$ go over into linear relations among themselves, and thus

$$\psi_i \to \tilde{L}_i(\psi_1, \psi_2, \cdots, \psi_s), \quad i = 1, \cdots, n, \tag{8}$$

where s of the parameters ψ_i have been taken to be the *independent* parameters of the toroid T^s (relabeling them if necessary). Therefore, under the restriction to the subgroup G',

$$(m, \psi) \to (m, \psi)_r = (m, \tilde{L}(\psi)). \tag{9}$$

However, inserting the linear forms $\tilde{L}_k(\psi)$ yields

$$\begin{aligned}(m, \psi)_r = (m', \psi) &\equiv (L(m), \psi) \\ &= m'_1\psi_1 + \cdots + m'_s\psi_s,\end{aligned} \tag{10}$$

where m' is a weight of G' whose components m'_k are given by linear relations of the components m_i of the weight m. Thus, instead of m', one can write $L(m)$, where m is a weight of G such that the linear relations among its components induced by the restriction to G' just form the weight m'. [An example is given in Sec. 4, following Eq. (64).]

Then, if the character $\chi(M)$ is restricted, from Eq. (4) one obtains

$$\chi(M)_r = \sum_{m \in D(M)} \gamma(m) e^{i[L(m),\psi]}. \tag{11}$$

On the other hand,

$$\chi(M)_r = \sum_{M'} \tilde{\gamma}(M')\chi(M'), \tag{12}$$

where the sum goes over all weights M' occurring in $D(M)$ under the restriction of G to G'; i.e., the M' are weights $L(m)$ which are highest weights.

Now, from Eq. (12), using Eq. (11) and the expression of $\chi(M')$, it follows that

$$\sum_{m \in D(M)} \gamma(m) e^{i[L(m),\psi]} = \sum_{M'} \tilde{\gamma}(M') \sum_{m' \in D(M')} \gamma(m') e^{i(m',\psi)}. \tag{13}$$

Multiplying Eq. (13) by $e^{i(\bar{m}',\psi)}$, where $\bar{m}'$ is some

[16] E. Stiefel, Commun. Math. Helvet. **14**, 350 (1942).

[17] B. Gruber, *Matscience Symposia*, A. Ramakrishnan, Ed. (Plenum Press, Inc., New York), Vol. 6 (to be published).

weight of G', and integrating over all parameters ψ_i of T^s, it follows that

$$\sum_{m\in D(M)} \gamma(m)\delta_{L(m),\bar{m}'} = \sum_{M'} \sum_{m'\in D(M')} \tilde{\gamma}(M')\gamma(m')\delta_{m',\bar{m}'}, \tag{14}$$

where

$$\delta_{m',\bar{m}'} = 1 \quad \text{if} \quad m' = \bar{m}',$$
$$= 0 \quad \text{if} \quad m' \neq \bar{m}'.$$

In deriving Eq. (14) the orthogonality relations for the $e^{i(m,\psi)}$ have been used together with the fact that the components m_i of the weights m are of the form[18,19]

$$\text{for } SU(l+1): m_i = \frac{k_i}{l+1},$$
$$\text{for } SO(2l+1): \text{(a) } m_i = k_i \text{ or (b) } m_i = k_i/2, \tag{15}$$
$$\text{for } SO(2l): \text{(a) } m_i = k_i \text{ or (b) } m_i = k_i/2,$$
$$\text{for } Sp(2l): m_i = k_i,$$

where k_i is some integer.

Moreover, the fact has been used that, for $SO(2l)$ and $SO(2l+1)$, the weights of an irreducible representation are all of the same form, i.e., either all of the form (a) or all of the form (b).

Equation (14) relates the multiplicity $\gamma^M(m)$ to the multiplicities $\tilde{\gamma}^M(M')$ and $\gamma^{M'}(m')$. While Eq. (14) holds for all compact semisimple Lie groups G, the groups G will henceforth be restricted to the $SU(n)$ groups (except for Sec. 4).

Before simplifying Eq. (14) for the case of the $SU(n)$ groups, a few words have to be said concerning the toroid T^s of the subgroup G'. First of all, if G' is a direct product of simple groups $G' = G_1 \otimes \cdots \otimes G_p$, then the toroid T^s of G' is the direct product $T^r \otimes \cdots \otimes T^t$ of the toroids of the simple groups. Since, however, any k-dimensional toroid T^k is the direct product of k one-dimensional toroids T^1, it is clear that one can speak of *the* (s-dimensional) toroid T^s of G', where $s = r + \cdots + t$ is the sum of the dimensions of the toroids of the simple groups $G_1, \cdots, G_p$.

The simplification one wants to achieve in Eq. (16) is to make the sum on the left-hand side of the equation trivial. This sum is nontrivial, since the rank s of any naturally embedded subgroup G' is smaller than the rank l of G [$= SU(n)$]. Therefore, in general, several weights m are mapped onto a single weight m' by $L(m)$. Thus, what one wants is a subgroup G^s of rank l which, in turn, contains G' as a subgroup in such a manner that the inner multiplicity structure of G^s is the same as for G'. For such a subgroup G^s the sum on the left-hand side of Eq. (14) reduces to a single term. The mapping $\mathbf{L}^s(m)$ of weights m of G onto weights m^s of the subgroup G^s induced by the restriction of G to G^s is one-to-one, and therefore, to a weight m, there corresponds just one weight, $\mathbf{L}^s(m)$. On the other hand, this single term will be expressed in terms of the inner multiplicity structure of the naturally embedded subgroup G'. Both goals can be achieved by choosing as subgroup G^s the group[15a]

$$G^s = G' \otimes \{U(1) \otimes \cdots \otimes U(1)\}_{l-s}, \tag{16}$$

where $\{U(1) \otimes \cdots \otimes U(1)\}_{l-s}$ indicates the direct product taken $l - s$ times. The toroid T of this group G^s is l-dimensional; there have been added just as many toroids $T^1 \sim U(1)$ as to make T an l-dimensional toroid T^l. Since, however, G^s is a subgroup of G, the toroid of G^s is contained in the toroid of G.

This, however, implies that the mapping

$$m \to \mathbf{L}^s(m)$$

of the weights of G onto weights $\mathbf{L}^s(m)$ of the subgroup G^s is nonsingular. On the other hand, the group G^s of Eq. (16) has the same inner multiplicity structure as the group G'. The inner multiplicity of the $U(1)$ groups is zero or one; therefore they do not influence the multiplicity of the weights m' of G'. In fact, the weights of the $U(1)$ groups can be looked upon as parameters which resolve the singularity in the mapping $\mathbf{m} \to L(m)$. [The weights λ differ only by a scale factor from the projections of a weight m of G onto the one-dimensional weight spaces of the $U(1)$ subgroups.] Thus the subgroup $G^s = G' \otimes U(1) \otimes \cdots \otimes U(1)$ satisfies our requirements.

The weights of the subgroup G^s are denoted by

$$\mathbf{m}^s = (m_1^s, \cdots, m_s^s; \lambda_1, \cdots, \lambda_{l-s}) \equiv (m^s, \lambda), \tag{17}$$

where $m^s = (m_1^s, \cdots, m_s^s)$ is a weight of the subgroup G' while the λ_i, $i = 1, \cdots, l - s$, are the weights of the $U(1)$ subgroups of G^s. Then the mapping $\mathbf{L}^s(m)$ of weights of G onto weights of the subgroup G^s is given as

$$\mathbf{L}^s(m) = (L^s(m), \lambda), \tag{18}$$

where $L^s(m)$ is the mapping of the weight m onto some weight m^s of the subgroup G' of G^s [previously denoted by $L(m)$].

Using this notation, Eq. (14), for the subgroup G^s, goes over into

$$\gamma^M(m) = \sum_{\mathbf{M}^s} \tilde{\gamma}(\mathbf{M}^s)\gamma^{\mathbf{M}^s}(\mathbf{L}^s(m))\delta_{\mathbf{L}^s(m),D(\mathbf{M}^s)}, \tag{19}$$

$$\delta_{\mathbf{L}^s(m),D(\mathbf{M}^s)} = 1 \quad \text{if} \quad \mathbf{L}^s(m) \in D(\mathbf{M}^s)$$
$$= 0 \quad \text{if} \quad \mathbf{L}^s(m) \notin D(\mathbf{M}^s),$$

[18] G. Racah, *Ergebnisse der exakten Naturwissenschaften* (Springer Verlag, Berlin, 1965), Vol. 37.
[19] F. Zaccaria, J. Math. Phys. 7, 1548 (1966).

where

$$\gamma^{\mathbf{M}^s}(\mathbf{L}^s(m)) = \gamma^{M^s}(L^s(m)), \tag{20a}$$

and where the Kronecker symbol δ has been added to express merely the fact that

$$\gamma^{\mathbf{M}^s}(\mathbf{L}^s(m)) = 0 \quad \text{if} \quad \mathbf{L}^s(m) \notin D(\mathbf{M}^s). \tag{20b}$$

Equation (19) can be simplified further if the group $SU(n-1)$ is chosen for G'. Then the subgroup G^s of the group $SU(n)$ is given as

$$G^{l-1} = SU(l) \otimes U(1), \quad n = l+1. \tag{21}$$

This subgroup however is familiar from Weyl's branching law[20,21] and it is exactly a property of the Weyl branching law which is to be used. Namely, that under the restriction of $SU(n)$ to $SU(n-1) \otimes U(1)$, there holds

$$\tilde{\gamma}(\mathbf{M}^{l-1}) = 1 \tag{22}$$

for all $\mathbf{M}^{l-1}$ occurring in the sum of Eq. (19). Thereby the weight $\mathbf{M}^{l-1}$, according to Eq. (17), is of the form

$$\mathbf{M}^{l-1} = (M^{l-1}; \lambda_1),$$

where M^{l-1} is a highest weight of $SU(l)$,while λ_1 is a weight of $U(1)$. Thus, with the help of Eq. (22), one obtains for Eq. (19)

$$\gamma^M(m) = \sum_{\mathbf{M}^{l-1}} \gamma^{\mathbf{M}^{l-1}}(\mathbf{L}^{l-1}(m)) \delta_{\mathbf{L}^{l-1}(m), D(\mathbf{M}^{l-1})}, \tag{23}$$

with

$$\gamma^{\mathbf{M}^{l-1}}(\mathbf{L}^{l-1}(m)) = \gamma^{M^{l-1}}(L(m)), \tag{24}$$

where the subgroup to be restricted to is $SU(l) \otimes U(1)$.

The last step is now obvious. What was done for $\gamma(m)$ is now done for $\gamma(\mathbf{m}^{l-1}) = \gamma(m^{l-1})$, and so on. If $\mathbf{M}^{l-1-i}$, $\mathbf{L}^{l-1-i}$ is written for the restriction of the group $SU(l+1-i) \otimes \{U(1) \otimes \cdots \otimes U(1)\}_i$ to the subgroup $SU(l-i) \otimes \{U(1) \otimes \cdots \otimes U(1)\}_{i+1}$, $i = 0, 1, \cdots, l-2$, ($l-1-i$ then denotes the rank of G') the final result is obtained as

$$\gamma^M(m) = \sum_{\mathbf{M}^{l-1}(M)} \sum_{\mathbf{M}^{l-2}(\mathbf{M}^{l-1})} \cdots \sum_{\mathbf{M}^1(\mathbf{M}^2)} \delta_{\mathbf{L}^{l-1}(m), D(\mathbf{M}^{l-1})} \times \delta_{\mathbf{L}^{l-2}\mathbf{L}^{l-1}(m), D(\mathbf{M}^{l-2})} \cdots \delta_{\mathbf{L}^1\mathbf{L}^2 \cdots \mathbf{L}^{l-1}(m), D(\mathbf{M}^1)}, \tag{25a}$$

or as

$$\gamma^M(m) = \sum_{\mathbf{M}^{l-1}(M)} \sum_{\mathbf{M}^{l-2}(\mathbf{M}^{l-1})} \cdots \times \sum_{\mathbf{M}^1(\mathbf{M}^2)} \delta_{\mathbf{L}^1\mathbf{L}^2 \cdots \mathbf{L}^{l-1}(m), D(\mathbf{M}^1)}, \tag{25b}$$

where the Kronecker symbols are defined as in Eq. (19). The summation "indices" $\mathbf{M}^{l-i}(\mathbf{M}^{l-i-1})$ denote that the sum is to be taken over all irreducible representations $D(\mathbf{M}^{l-i})$ occurring in the restriction of $D(\mathbf{M}^{l-i+1})$ to the subgroup $SU(l-i+1) \otimes \{U(1) \otimes \cdots \otimes U(1)\}_i$. Thereby, $\mathbf{M}^l = (M^l; 0) \equiv M$. Equation (25b) follows from Eq. (25a), since

$$\sum_{\mathbf{M}^{l-i}(\mathbf{M}^{l-i+1})} \delta_{\mathbf{L}^{l-i} \cdots \mathbf{L}^{l-1}(m), D(\mathbf{M}^{l-i})} = 0,$$

if

$$\delta_{\mathbf{L}^{l-i+1} \cdots \mathbf{L}^{l-1}(m), D(\mathbf{M}^{l-i+1})} = 0, \quad i = 2, 3, \cdots, l-1.$$

Equations (25a) and (25b) can be rewritten more explicitly if the definitions given in Eqs. (17) and (18) are used. Then

$$\begin{aligned} &\delta_{\mathbf{L}^{l-k}\mathbf{L}^{l-k+1} \cdots \mathbf{L}^{l-1}(m), D(\mathbf{M}^{l-k})} \\ &= \delta_{(L^{l-k}L^{l-k+1} \cdots L^{l-1}(m); \lambda_1, \lambda_2, \cdots, \lambda_k), D(M^{l-k}; \lambda_1', \lambda_2', \cdots, \lambda_k')} \\ &= \delta_{\lambda_1, \lambda_1'} \delta_{\lambda_2, \lambda_2'} \cdots \delta_{\lambda_k, \lambda_k'} \delta_{L^{l-k}L^{l-k+1} \cdots L^{l-1}(m), D(M^{l-k})}. \end{aligned} \tag{26}$$

Using Eq. (26), the expressions for the inner multiplicity $\gamma^M(m')$, Eq. (25), can be rewritten as

$$\gamma^M(m) = \sum_{\mathbf{M}^{l-1}(M)} \cdots \sum_{\mathbf{M}^1(\mathbf{M}^2)} \delta_{\lambda_1, \lambda_1'} \delta_{\lambda_2, \lambda_2'} \cdots \delta_{\lambda_{l-1}, \lambda_{l-1}'} \times \delta_{L^{l-1}(m), D(M^{l-1})} \cdots \delta_{L^1 \cdots L^{l-1}(m), D(M^1)} \tag{27a}$$

and

$$\gamma^M(m) = \sum_{\mathbf{M}^{l-1}(M)} \cdots \sum_{\mathbf{M}^1(\mathbf{M}^2)} \delta_{\lambda_1, \lambda_1'} \cdots \delta_{\lambda_{l-1}, \lambda_{l-1}'} \times \delta_{L^1 \cdots L^{l-1}(m), D(M^1)}, \tag{27b}$$

where M^1 is a $SU(2)$ highest weight, $\mathbf{M}^1 = (M^1; \lambda_1', \cdots, \lambda_{l-1}')$.

The meaning of Eq. (25b) and Eq. (27b) can be understood easily. If the group $SU(n)$ is restricted to the subgroup $G^1 = SU(2) \otimes \{U(1) \otimes \cdots \otimes U(1)\}_{l-1}$, the representation $D(M)$ decomposes into a number of irreducible representations $D(\mathbf{M}^1)$ of G^1. The multiplicity $\gamma^M(m)$ of a weight m of $D(M)$ then is equal to the number of irreducible representations $D(\mathbf{M}^1)$ of G^1 which contain the weight

$$\mathbf{L}^1\mathbf{L}^2 \cdots \mathbf{L}^{l-1}(m) = \mathbf{m}^1 = (m^1; \lambda_1, \cdots, \lambda_{l-1}),$$

where m^1 is an $SU(2)$ weight.

On the other hand, from Eq. (25a) and Eq. (27a), the relationship to the Gel'fand patterns can be read off. In order to do so, it is first necessary to relate the notation used for the Gel'fand patterns to the weight notation for the $SU(n)$ groups [described in Eqs. (1), (2), (3)].

[20] H. Weyl, *Theory of Groups and Quantum Mechanics* (Dover Publications, Inc., New York, 1931).

[21] H. Boerner, *Representations of Groups* (North-Holland Publ. Company, Amsterdam, 1963).

A Gel'fand pattern is given as[15]

$$\begin{bmatrix} m_{1n} & m_{2n} & m_{3n} & \cdots & m_{nn} \\ & m_{1\,n-1} & m_{2\,n-1} & \cdots & \\ & & m_{1\,n-2} & \cdots & \\ & & \cdots & & \\ & & & m_{11} & \end{bmatrix}, \tag{28}$$

where the m_{in} characterize an irreducible representation of $U(n)$. The irreducible representations of $SU(n)$ are also characterized by a set m_{in}; however, $m_{nn} = 0$ holds for them. (The m_{in} are the partition numbers $\{\lambda_i\}$ of the corresponding Young tableaux.) Since the m_{in} with $m_{nn} = 0$ characterize an irreducible representation $D(M^{n-1}) \equiv D(M_1, \cdots, M_n)$ of $SU(n)$ as well as the highest weights do, there must be a unique relation among the two labelings. The relation is

$$\begin{aligned} m_{1n} &= M_1 - M_n, \\ M_{2n} &= M_2 - M_n, \\ &\cdots \\ m_{n-1\,n} &= M_{n-1} - M_n, \\ m_{nn} &= 0. \end{aligned} \tag{29}$$

All the m_{ik} of Eq. (28) are nonnegative integers for which the usual betweenness relations hold:

$$m_{i\,k+1} \geq m_{ik} \geq m_{i+1\,k+1}. \tag{30}$$

Now, consider an irreducible representation $D(M^{n-1}) = D(M_1, \cdots, M_n) = D(m_{in})$ of $SU(n)$, whose partition numbers are m_{in}, $i = 1, \cdots, (n-1)$, $m_{nn} = 0$. Then each admissible set $m_{i\,n-1}$, $i = 1, 2, \cdots, (n-1)$, i.e., each set of integers $m_{i\,n-1}$ satisfying the relations of Eq. (30), characterizes an irreducible representation $D(\mathbf{M}^{l-1})$ of the subgroup $SU(n-1) \otimes U(1)$ of $SU(n)$. All the sets $\{m_{i\,n-1}\}$ satisfying the relations of Eq. (30) then correspond to the subgroup content of the irreducible representation $D(m_{in})$ if $SU(n)$ is restricted to $SU(n-1) \otimes U(1)$.

Now, in turn, the subgroup content of the representations $D(m_{i\,n-1}) \equiv D(\mathbf{M}^{l-1})$, $i = 1, \cdots, (n-1)$ of $SU(n-1) \otimes U(1)$ with respect to $SU(n-2) \otimes U(1) \otimes U(1)$ is given by all sets $m_{i\,n-2}$, $i = 1, 2, \cdots, n-2$, satisfying the relations of Eq. (30). Proceeding in this fashion one finally arrives at the subgroup content of $D(M)$ with respect to the subgroup $SU(2) \otimes U(1) \otimes \cdots \otimes U(1)$.

From the above, the relationship of Eqs. (25a) and (27a) to the Gel'fand patterns is easy to see; moreover, it can be seen that the formulas of Eq. (25a) and Eq. (27a) contain the prescription of how to obtain the multiplicity $\gamma^M(m)$ by means of Gel'fand patterns. Namely, for a given weight m and a given representation $D(M)$ of $SU(n)$ containing this weight, the subgroup content of $D(M)$ with respect to $SU(n-1) \otimes U(1)$ is considered. The weight M determines the first line of a Gel'fand pattern, and the subgroup content $D(\mathbf{M}^{l-1})$ of $D(M)$ is given by all possible second lines of this Gel'fand pattern. All the representations $D(\mathbf{M}^{l-1})$ of $SU(n-1) \otimes U(1)$ containing the projected weight $\mathbf{L}^{l-1}(m)$ contribute to the multiplicity $\gamma^M(m)$. For all these representations $D(\mathbf{M}^{l-1})$ containing $\mathbf{L}^{l-1}(m)$, the same procedure is applied as for $D(M)$; i.e., for each such $D(\mathbf{M}^{l-1})$ the subgroup content $D(\mathbf{M}^{l-2})$ is determined. This content is given by all possible third lines for a second line which corresponds to such a $D(\mathbf{M}^{l-1})$. Again, if the projected weight $\mathbf{L}^{l-2}\mathbf{L}^{l-1}(m)$ is contained in a representation $D(\mathbf{M}^{l-2})$, this representation contributes to the multiplicity $\gamma^M(m)$. Continuing in this fashion until the subgroup $SU(2) \otimes U(1) \otimes \cdots \otimes U(1)$ is reached, the multiplicity $\gamma^M(m)$ is equal to the number of all representations $D(\mathbf{M}^1)$ reached in this manner which contain the "projected" weight $\mathbf{L}^1 \cdots \mathbf{L}^{l-1}(m)$. A prescription to be used for the actual calculation of $\gamma^M(m)$ will be given at the end of Sec. 3.

3. INNER MULTIPLICITY AND GEL'FAND PATTERNS

In the preceding section the recursion formula (23) for the multiplicity of a weight of $SU(n)$ was derived using the properties of the characters of $SU(n)$ and some of its subgroups. In this section essentially the same recursion relation is derived by a more direct, but less general, method using Gel'fand patterns. As a side result of this derivation, a simple scheme is obtained by which the eigenvalues of the diagonal generators of $SU(n)$ can be computed using Young's diagrams.

With $U(\alpha)$ denoting the group of unitary, unimodular, n-dimensional matrices, and labeling the group parameters by α_a ($a = 1, \cdots, n^2 - 1$), one has

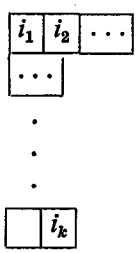
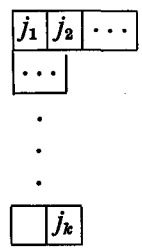
$$U(\alpha) = e^{i\alpha_a f_a}, \tag{31}$$

where f_a denotes the generators of the self-representation of $SU(n)$.

A contravariant tensor of rank k transforms as[21]

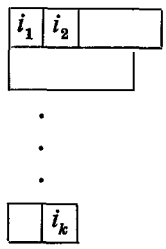
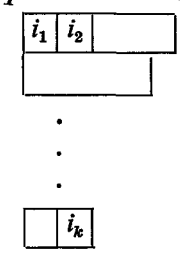
$$T'_{i_1 i_2 \cdots i_k} = U_{i_1 j_1} U_{i_2 j_2} \cdots U_{i_k j_k} T_{j_1 j_2 \cdots j_k}, \; i, j = 1, \cdots, n. \tag{32}$$

As is known, the irreducible parts of this tensor can be characterized by Young's tableaux,[21] and because of the bisymmetric property of the transformation, Young-symmetrized tensors transform in the same way, namely,

T'  $= U_{i_1 j_1} \cdots U_{i_k j_k} T$  , (33)

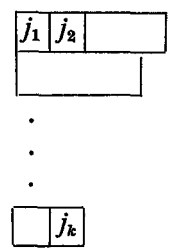
the sum on the right in Eq. (33) being carried out over all values of $(j_1, j_2, \cdots, j_k)$.

If a transformation infinitesimally close to the identity ($\alpha_a = \epsilon_a$) is performed, one has, to first order in ϵ_a,

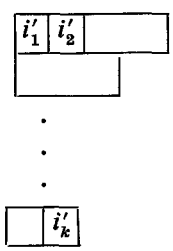
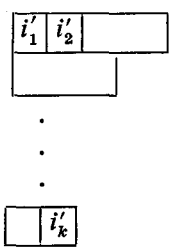
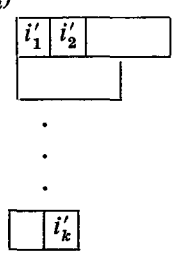
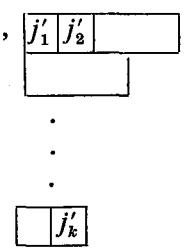
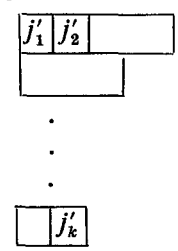
T'  $= T$  $+ i\epsilon_a [(f_a)_{i_1 j_1} \delta_{i_2 j_2} \cdots \delta_{i_k j_k}$

$+ \delta_{i_1 j_1} (f_a)_{i_2 j_2} \cdots \delta_{i_k j_k}$

$\cdots$ (34)

$+ \delta_{i_1 j_1} \delta_{i_2 j_2} \cdots (f_a)_{i_k j_k}] T$  .

Now with F_a labeling the generators of the irreducible representation corresponding to the above Young tableaux, one also has to first order in ϵ_a

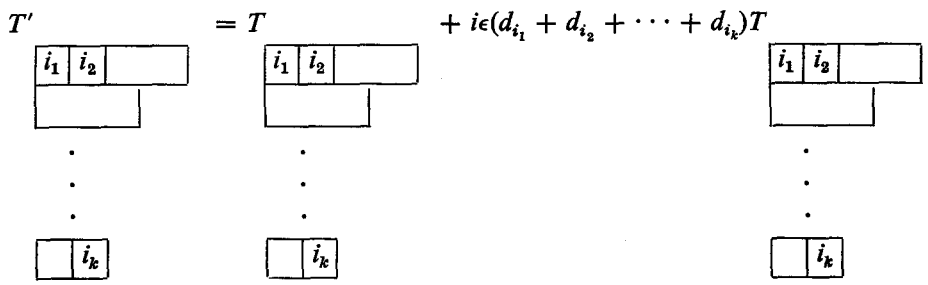
T'  $= T$  $+ i\epsilon_a (F_a)$  ,  T  , (35)

where the primed indices range over only those values consistent with standard Young diagrams[21] (i.e., so that only linearly independent tensor components occur).

Suppose one chooses a transformation $e^{i\alpha_a f_a}$ belonging to the toroid of $SU(n)$ such that $\alpha_a f_a = \epsilon_a f_a = \epsilon d$, where d is chosen as diagonal in the self-representation; i.e.,

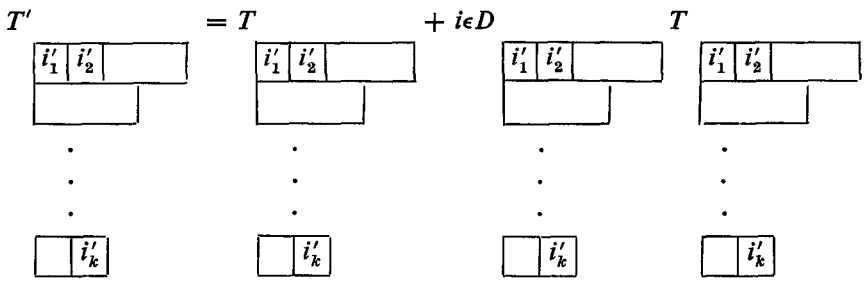
$$d_{ij} = d_i \delta_{ij}. \tag{36}$$

Then, by Eq. (34),



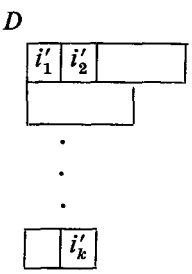
(no sum over $i_1, i_2, \cdots, i_k$). (37)

Now if $\epsilon_a F_a = \epsilon D$, where D is chosen as diagonal in the irreducible representation, according to Eq. (35) one has



(no sum over $i'_1, i'_2, \cdots, i'_k$). (38)

Then choosing $(i_1, i_2, \cdots, i_k) = (i'_1, i'_2, \cdots, i'_k)$ in Eq. (37) and comparing with Eq. (38), one sees that



(39)

Thus the diagonal elements of the commuting generators in any irreducible representation of $SU(n)$ are given by simple sums of the diagonal elements of the corresponding generators in the self-representation, the elements to be added being determined by the standard Young diagrams of the irreducible representation.

The prescription, Eq. (39), will now be used to construct the weight vectors of $SU(n)$. Using the Okubo notation for the generators of $SU(n)$, one has in the self-representation

$$(A^i_j)^k_l = \delta^i_l \delta^k_j - n^{-1} \delta^i_j \delta^k_l. \tag{40}$$

Thus the diagonal elements of the commuting generators in the self-representation are given by

$$\begin{aligned} A^1_1 &= n^{-1}(n-1, -1, -1, \cdots, -1), \\ A^2_2 &= n^{-1}(-1, n-1, -1, \cdots, -1), \\ &\cdots \\ A^n_n &= n^{-1}(-1, -1, -1, \cdots, n-1). \end{aligned} \tag{41}$$

In any irreducible representation the diagonal elements of the $A^{(i')}_{(i')}$, $i = 1, \cdots, n$, corresponding to a simultaneous eigenvector form the components of a vector. The vectors thus obtained are called weight vectors. In the self-representation, these weight vectors are given by the columns of Eq. (41).

Let us consider a particular Young tableaux corresponding to an irreducible representation characterized by the partition numbers $\lambda_1, \lambda_2, \cdots, \lambda_n = 0$.[22] Further, let us consider a particular standard Young diagram of that tableaux. In this standard diagram let us denote by

$$\begin{aligned} &N_1 \quad \text{the number of 1's,} \\ &N_2 \quad \text{the number of 2's,} \\ &\cdot \\ &\cdot \\ &\cdot \\ &N_n \quad \text{the number of } n\text{'s.} \end{aligned} \tag{42}$$

Then, from Eq. (39), when inserting the d_i values of Eq. (41), it follows that the components of the general weight of $SU(n)$ are given by

$$\begin{aligned} m_1 &= \frac{1}{n}[(n-1)N_1 - N_2 - \cdots - N_n], \\ m_2 &= \frac{1}{n}[-N_1 + (n-1)N_2 - \cdots - N_n], \\ &\cdot \qquad\qquad \cdot \\ &\cdot \qquad\qquad \cdot \\ &\cdot \qquad\qquad \cdot \\ m_n &= \frac{1}{n}[-N_1 - N_2 - \cdots + (n-1)N_n], \end{aligned} \tag{43}$$

$$\sum_{i=0}^{n} m_i = 0.$$

Moreover, since

$$\sum_{i=1}^{n} N_i = \sum_{i=1}^{n-1} \lambda_i \tag{44}$$

holds, the components m_i of the weights of Eq. (43) can be rewritten as

$$m_j = N_j - \frac{1}{n}\sum_{k=1}^{n-1} \lambda_k. \tag{45}$$

Equation (45) is the starting point for our considerations concerning the multiplicity of a weight $m = (m_1, m_2, \cdots, m_n)$. This equation expresses the fact that the multiplicity $\gamma^{\{\lambda_i\}}(m)$ of the weight m of the irreducible representation $\{\lambda_1, \lambda_2, \cdots, \lambda_n = 0\}$ is equal to the number of standard Young diagrams that can be constructed from the same set of occupation numbers $N_1, N_2, \cdots, N_n$. Thus, if the standard Young diagrams of an irreducible representation were given, this set of diagrams could be separated into subsets, each subset being characterized by the same set of numbers $N_1, N_2, \cdots, N_n$. Then, each such subset corresponds to a weight, the number of elements of the subset being its multiplicity. From this it is clear that the multiplicity of a weight m can be obtained by a counting process.

However, instead of carrying through the counting process in terms of standard Young diagrams, it is of advantage to use instead the Gel'fand patterns described in Eqs. (28) and (30). This is possible due to the one-to-one correspondence between Gel'fand patterns and standard Young diagrams.[15] This one-to-one correspondence is given by the fact that in a Gel'fand pattern:

(A) the m_{ii} are equal to the number of i's in the ith row of the Young diagram; and

(B) the $m_{ij} - m_{ij-1}$ are equal to the number of j's in the ith row of the Young diagram, $i < j, i = 1, 2, \cdots, n$; or conversely, the occupation numbers N_j are given in terms of the m_{ik} as

$$N_j = m_{jj} + \sum_{i=1}^{j-1}(m_{ij} - m_{i,j-1}), \quad j = 1, 2, 3, \cdots, n. \tag{46}$$

From Eq. (46) it follows that

$$\sum_{i=1}^{j} m_{ij} = \sum_{i=1}^{j} N_i \equiv S_j, \quad j = 1, 2, \cdots, n. \tag{47}$$

Thus, for a particular weight vector $m = (m_1, m_2, \cdots, m_n)$ of an irreducible representation $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ given by Eq. (45), there exists a unique set of sums along the rows of the Gel'fand patterns $\{S_j\}$.

The multiplicity $\gamma^{\{\lambda_i\}}(m) = \gamma^{\{\lambda_i\}}(N)$ of this vector m is then given by the number of Gel'fand patterns which possess the same set of sums:

$$S_j = \sum_{i=1}^{j} m_i + \frac{j}{n}\sum_{k=1}^{n-1} \lambda_k, \; m_{in} = \lambda_i, \quad j, i = 1, 2, \cdots, n. \tag{48}$$

From what has been said it is now possible to derive a recurrence relation for the multiplicity $\gamma^{\{\lambda\}}(N_j)$. This relation is obtained as follows: For a given irreducible representation $\{\lambda_1, \lambda_2, \cdots, \lambda_n = 0\}$, the sets $\{\lambda'_1, \lambda'_2, \cdots, \lambda'_{n-1}\}$ are determined. These sets satisfy

$$\lambda_1 \geq \lambda'_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda'_{n-1} \geq \lambda_n = 0, \tag{49}$$

as well as the supplementary condition

$$\lambda'_1 + \lambda'_2 + \cdots + \lambda'_{n-1} = \sum_{i=1}^{n-1} N_i \equiv S_{n-1}. \tag{50}$$

For each of the sets $\{\lambda'_1, \lambda'_2, \cdots, \lambda'_{n-1}\}$ obtained in this fashion, let

$$\gamma(\lambda'_1 - \lambda'_{n-1}, \lambda'_2 - \lambda'_{n-1}, \cdots, 0; \\ N_1 - \lambda'_{n-1}, \cdots, N_{n-1} - \lambda'_{n-1})$$

[22] These λ's should not be confused with the $U(1)$ weights of Sec. 2.

denote the number of Gel'fand patterns

$$\begin{bmatrix} \lambda_1' & \lambda_2' & \cdots & \lambda_{n-1}' \\ m_{1\,n-2} & m_{2\,n-2} & \cdots & m_{n-2\,n-2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ & m_{11} & & \end{bmatrix} \tag{51}$$

satisfying

$$\sum_{i=1}^{j} m_{ij} = S_j, \quad j = 1, 2, \cdots, n-2. \tag{52}$$

Then the recurrence relation is obtained as [for reasons of convenience, $\gamma(\lambda, N)$ is now written instead of $\gamma^{\lambda}(N)$, as previously]

$$\gamma(\lambda_1, \lambda_2, \cdots, \lambda_{n-1}; N_1, \cdots, N_n) = \sum_{\lambda_1', \cdots, \lambda'_{n-1}} \gamma(\lambda_1' - \lambda_{n-1}', \lambda_2' - \lambda_{n-1}', \cdots, 0; N_1 - \lambda_{n-1}', \cdots, N_{n-1} - \lambda_{n-1}'), \tag{53}$$

where the sum runs over all sets $\{\lambda_i'\}$ satisfying

$$\lambda_1 \geq \lambda_1' \geq \lambda_2 \geq \cdots \geq \lambda_{n-1}' \geq 0,$$
$$\lambda_1' + \lambda_2' + \cdots + \lambda_{n-1}' = \sum_{i=0}^{n-1} N_i. \tag{54}$$

The reason for writing

$$\gamma(\lambda_1 - \lambda_{n-1}', \cdots, 0; N_1 - \lambda_{n-1}', \cdots, N_{n-1} - \lambda_{n-1}')$$

for the number of Gel'fand patterns given by Eq. (51) under the restrictions imposed by Eq. (52) is that this number is equal to the multiplicity of the weight $m = (m_1, m_2, \cdots, m_{n-1})$ of the irreducible representation $\{\lambda_1' - \lambda_{n-1}', \lambda_2' - \lambda_{n-1}', \cdots, 0\}$ of $SU(n-1)$ whose occupation numbers are

$$(N_1 - \lambda_{n-1}', N_2 - \lambda_{n-1}', \cdots, N_{n-1} - \lambda_{n-1}').$$

From the aforesaid it is clear that

$$\gamma(\lambda_1' - \lambda_{n-1}', \lambda_2' - \lambda_{n-1}', \cdots, 0; N_1 - \lambda_{n-1}', \cdots, N_{n-1} - \lambda_{n-1}') = 0, \tag{55}$$

if any

$$N_i - \lambda_{n-1}' < 0, \quad i = 1, \cdots, n-1, \tag{56}$$

since the occupation numbers cannot be negative.

The condition in Eq. (56) can be sharpened. Using Eq. (46) and the properties of the Gel'fand pattern

$$N_i \geq m_{ii} \geq m_{jj}, \quad i = 1, \cdots, j,$$

we have

$$N_i \geq m_{jj}, \qquad i = 1, \cdots, j. \tag{57}$$

Setting $j = n-1$ yields condition Eq. (56).

Equations (53), (54), (55), and (56) correspond to Eqs. (23) and (25) of the previous section.

The multiplicity $\gamma^M(m) = \gamma(\lambda, N)$ of a weight m, with occupation numbers $(N_1, N_2, \cdots, N_n)$, which belongs to the irreducible representation $D(M) \equiv D(\lambda)$, is therefore obtained as follows.

Method A

1. Find all sets $\{\lambda_i'\}$, $i = 1, \cdots, n-1$ satisfying

$$\lambda_1 \geq \lambda_1' \geq \lambda_2 \geq \lambda_2' \geq \cdots \geq \lambda_{n-1}' \geq \lambda_n = 0,$$
$$\lambda_1' + \lambda_2' + \cdots + \lambda_{n-1}' = \sum_{i=1}^{n-1} N_i,$$

and

$$N_j - \lambda_{n-1}' \geq 0 \quad j = 1, \cdots, n-1.$$

2. For each of the sets of 1, perform 1 again, where everywhere $n \to n-1$ (now $\lambda_{n-1} \neq 0$ in general).

3. Continue in this fashion until $i = 1, 2$. The number of all sets obtained is the multiplicity $\gamma(\lambda_1, \cdots, \lambda_n = 0, N_1, \cdots, N_n) = \gamma^M(m)$ of the weight m of $D(M)$. The relations between $\{\lambda_i\}$, M, and $\{N_i\}$, m are given by Eqs. (29) and (45), respectively. An example is given in Appendix A.

Since, in general, the weights m are not given in terms of the occupation numbers (N_i), it may be advantageous to give the prescription of the calculation of $\gamma^M(m)$ in terms of the weight notation and the partition numbers alone.

However, the number of sets to be considered now becomes larger, since sets $\{\lambda'\}$ which are ruled out by the last condition of 1 are carried along. These sets will be ruled out only by the last step. The multiplicity $\gamma^M(m)$ is then obtained as follows.

Method B

1. Form the set $\{\lambda_i\} = M_i - M_n$, $i = 1, \cdots, n$, and

$$\Lambda = \sum_{k=1}^{n-1} \lambda_k.$$

2. Determine all sets $\{\lambda_i'\}$, $i = 1, \cdots, n-1$, satisfying

$$\lambda_1 \geq \lambda_1' \geq \lambda_2 \geq \lambda_2' \geq \cdots \geq \lambda_{n-1}' \geq 0$$

and

$$\lambda_1' + \lambda_2' + \cdots + \lambda_{n-1}' = S_{n-1} = \sum_{i=1}^{n-1} m_i + \frac{n-1}{n} \Lambda.$$

3. With all sets obtained in this manner, perform 2. Now $(n-1) \to (n-2)$.

4. Continue in this fashion until $i = 1$. The number of sets obtained by the last step is the multiplicity $\gamma^M(m)$.

An example is given in Appendix B.

4. REDUCTION MULTIPLICITY

In this section a formula for the reduction multiplicity is derived. Given an irreducible representation $D(M)$ of a group G, the reduction multiplicity $\tilde{\gamma}(M')$ of irreducible representations $D(M')$ into which $D(M)$ decomposes, if the group G is restricted to a naturally embedded subgroup G', is related to the inner multiplicity $\gamma(m)$ of the weights m of $D(M)$. Then from this formula the reduction multiplicity can be calculated for $SU(n)$ by means of the prescriptions given for the calculation of the inner multiplicities at the end of the previous section.

The relation between reduction and inner multiplicity is obtained in the same fashion as Eq. (14). From Eq. (11), (12), and (6) it follows that

$$\sum_{m\in D(M)} \gamma(m) e^{i[L(m),\psi]} \sum_{S'\in W'} \delta_{S'} e^{i(S'R_0',\psi)} = \sum_{M'} \tilde{\gamma}(M') \sum_{S'\in W'} \delta_{S'} e^{i[S'(M'+R_0'),\psi]}. \quad (58)$$

The primed quantities belong to the subgroup G'; thus S' is an element of the Weyl group W' of G', etc. Then, multiplying by some $e^{i(M',\psi)}$ and using the orthogonality relations holding for the trigonometric functions, one obtains, for the reduction multiplicity with respect to the subgroup G', the relation

$$\tilde{\gamma}(M') = \sum_{S'\in W'} \sum_{m\in D(M)} \gamma(m)\delta_{S'}\delta_{L(m)+S'R_0',M'+R_0'}, \quad (59)$$

with $\delta_{S'} = \mp 1$, depending on whether S' is a reflection or not and with

$$\delta_{L(m)+S'R_0',M'+R_0'} = 1 \quad \text{if} \quad L(m) + S'R_0' = M' + R_0', \qquad = 0 \quad \text{otherwise.} \quad (60)$$

By using

$$\gamma(S'm) = \gamma(m), \quad (61)$$

an alternate form is obtained for Eq. (59), namely,

$$\tilde{\gamma}(M') = \sum_{S'\in W'} \sum_{m\in D(M)} \gamma(m)\delta_{S'}\delta_{S'(L(m)+R_0'),M'+R_0'}, \quad (62)$$

where, again for the two δ's, what was said following Eq. (59) holds. The relation given in Eq. (61) holds since W' is a subgroup of the Weyl group W of G. An example is treated in Appendix D.

In order to demonstrate the usefulness of Eq. (59) or Eq. (62), the known result of the Weyl branching law is rederived for $SU(3)$. Namely, that if $SU(3)$ is restricted to the subgroup $G^1 = SU(2) \otimes U(1)$, the multiplicity $\tilde{\gamma}(M^1, \lambda)$ is 0 or 1.

In the Okubo notation introduced in Eq. (40), the toroid of $SU(3)$ is given as

$$\exp i(\psi_1 A_1^1 + \psi_2 A_2^2 + \psi_3 A_3^3). \quad (63)$$

Equation (63) can be rewritten as

$$\exp i(\psi_1 \bar{A}_1^1 + \psi_2 \bar{A}_2^2) \cdot \exp [i\tfrac{3}{2} \cdot \psi_3 A_3^3] \quad (64)$$

where

$$\bar{A}_1^1 = A_1^1 + \tfrac{1}{2}A_3^3,$$
$$\bar{A}_2^2 = A_2^2 + \tfrac{1}{2}A_3^3,$$

so that $\bar{A}_1^1 + \bar{A}_2^2 = 0$ and the parameters satisfy $\psi_1 + \psi_2 + \psi_3 = 0$. Then the generators $\bar{A}_1^1$, $\bar{A}_2^2$, A_1^2, A_2^1, and A_3^3 form a subgroup $SU(2) \otimes U(1)$ of $SU(3)$ and it follows that

$$m \to \mathbf{L}^1(m) = (m_1 + \tfrac{1}{2}m_3, m_2 + \tfrac{1}{2}m_3; \tfrac{3}{2}m_3). \quad (65)$$

From the above it is now easy to understand the meaning of the mappings $\tilde{L}$, L, and $\mathbf{L}$ of Eqs. (8), (10), and (18). The root system of $SU(3)$ is given as

$$\pm(e_i - e_j),\ i < j,\ i,j = 1, 2, 3.$$

Then $SU(2)$ can be embedded into $SU(3)$ by choosing $\pm(e_1 - e_2)$ as the roots of $SU(2)$. With this choice, the parameters ψ_1, ψ_2, ψ_3 of the toroid T^2 of $SU(3)$ with

$$\psi_1 + \psi_2 + \psi_3 = 0, \quad (66)$$

go over into the parameters ψ_1, ψ_2, ψ_3 of the Toroid T^1 of the subgroup $SU(2)$ with

$$\psi_1 + \psi_2 + \psi_3 = 0, \quad (67a)$$

$$(\psi_1, \psi_2, \psi_3)(1, 1, -2) = \psi_1 + \psi_2 - 2\psi_3 = 0. \quad (67b)$$

Equation (67b) is the additional linear relation among the parameters ψ_i, $i = 1, 2, 3$; the two planes of Eqs. (67a) and (67b) just describe the one-dimensional subspace of T^1. Then

$$\psi_i \to \tilde{L}_i(\psi_1, \psi_2),\ i = 1, 2, 3,$$

where

$$\tilde{L}_1(\psi_1, \psi_2) = \psi_1, \quad \tilde{L}_2(\psi_1, \psi_2) = \psi_2, \quad \tilde{L}_3(\psi_1, \psi_2) = \tfrac{1}{2}(\psi_1 + \psi_2), \quad (68a)$$

with

$$3(\psi_1 + \psi_2)/2 = 0 \pmod{2\pi}. \quad (68b)$$

In Eq. (68a) the linear forms $\tilde{L}_i$ have been expressed in terms of the first two parameters ψ_1, ψ_2. Equation (68b) expresses the fact that ψ_1 and ψ_2 are not independent. [As was said in Sec. 2, it is preferrable that the $SU(n)$ groups have embedded their $(n-1)$-dimensional weight space in an n-dimensional Euclidean space, having then a supplementary condition for the parameters ψ_i, $i = 1, \cdots, n$.] Then, from Eq. (68a), for the mapping $L = L^1$ it follows that

$$m = (m_1, m_2, m_3) \to L^1(m),$$
$$L^1(m) = (m_1 + \tfrac{1}{2}m_3, m_2 + \tfrac{1}{2}m_3, 0), \quad (69)$$

where now $L(m)$ is a $SU(2)$ weight. What is wanted,

however, are the weights of G^2. Thus the weights for $U(1)$ have yet to be found. The space of $U(1)$, perpendicular to the toroid of $SU(2)$, is given as the intersection of the two planes

$$\psi_1 - \psi_2 = 0, \quad \psi_1 + \psi_2 + \psi_3 = 0. \tag{70}$$

Thus, when expressing the ψ_i in terms of ψ_3, one obtains

$$\psi_1 \to -\tfrac{1}{2}\psi_3,$$
$$\psi_2 \to -\tfrac{1}{2}\psi_3,$$
$$\psi_3 \to \psi_3.$$

Therefore

$$(m, \psi) \to (\tfrac{3}{2}m_3)\psi_3.$$

Thus the mapping $\mathbf{L}^1$ is given as

$$m \to \mathbf{L}^1(m) = \left(m_1 + \frac{m_3}{2}, m_2 + \frac{m_3}{2}; \frac{3m_3}{2}\right). \tag{71}$$

Equation (59) for the reduction multiplicity then reads

$$\begin{aligned}
&\tilde{\gamma}(M_1^1, M_2^1; \lambda) \\
&\quad = \sum_{S' \in W'} \sum_{m \in D(M)} \gamma(m_1, m_2, m_3)\delta_{S'} \\
&\qquad \times \delta_{[m_1+(m_3/2), m_2+(m_3/2); 3m_3/2]+S'(\frac{1}{2},-\frac{1}{2}),(M_1^1+\frac{1}{2}, M_2^1-\frac{1}{2};\lambda)} \\
&\quad = \sum_{m \in D(M)} \delta_{3m_3/2,\lambda}\{\gamma(m_1, m_2, m_3) \\
&\qquad \times \delta_{[m_1+(m_3/2), m_2+(m_3/2)],(M_1^1, M_2^1)} - \gamma(m_1, m_2, m_3) \\
&\qquad \times \delta_{[m_1+(m_3/2)-\frac{1}{2}, m_2+(m_3/2)+\frac{1}{2}],(M_1^1+\frac{1}{2}, M_2^1-\frac{1}{2})}\}.
\end{aligned} \tag{72}$$

Now, the mapping $\mathbf{L}^1$ of Eq. (71) is one-to-one. The inverse mapping $(\mathbf{L}^1)^{-1}$ is given by

$$(\mathbf{L}^1)^{-1}(M_1^1, M_2^1; \lambda) = \left(M_1^1 - \frac{\lambda}{3}, M_2^1 - \frac{\lambda}{3}, \frac{2\lambda}{3}\right), \tag{73}$$

where the weight on the right-hand side of Eq. (73) is the $SU(3)$ weight which is mapped by $\mathbf{L}^1$ onto the weight $(M_1^1, M_2^1; \lambda)$ of $SU(2) \otimes U(1)$. Using Eq. (73) in Eq. (72), the following expression is obtained:

$$\begin{aligned}
&\tilde{\gamma}(M_1^1, M_2^1; \lambda) \\
&\quad = \gamma^M\left(M_1^1 - \frac{\lambda}{3}, M_2^1 - \frac{\lambda}{3}, \frac{2\lambda}{3}\right) \\
&\qquad - \gamma^M\left(M_1^1 - \frac{\lambda}{3} + 1, M_2^1 - \frac{\lambda}{3} - 1, \frac{2\lambda}{3}\right).
\end{aligned} \tag{74}$$

However, in a $SU(3)$ weight diagram, when going along the vector $R_0 = (1, 0, -1)$, the multiplicity either changes in each step by ± 1 or remains constant.[22] Thus, the values of Eq. (74) can be ± 1 or 0. [What was said about R_0 is also true for the vector $(1, -1, 0)$, which is obtained from R_0 by a reflection S of the Weyl group W.]

The value -1, however, can be ruled out since the weight $(M_1^1 - \lambda/3, M_2^1 - \lambda/3, 2\lambda/3)$, when restricted to $SU(2)$, is a highest weight for $SU(2)$. Thus

$$\tilde{\gamma}^M(M_1^1, M_2^1; \lambda) = 0, 1. \tag{75}$$

Equation (75) states the well-known result of the Weyl branching law for $SU(3)$. Of course, Eq. (75) can also be obtained from Eq. (74) by using Kostant's formula.[7]

Knowing the result of Eq. (75), another useful formula can be derived, namely, a formula for the upper limit of the *maximal* multiplicity $\tilde{\gamma}(M^1)$ of $SU(2)$ representations $D(M^1)$ which are contained in a $SU(3)$ representation $D(M^3)$, if $SU(3)$ is restricted to $SU(2)$.

Any two weights of an $SU(3)$ representation $D(M^2)$ related by the vector $(1, 1, -2)$ go over into the same $SU(2)$ weight when $SU(3)$ is restricted to $SU(2)$. Namely, the vector $(1, 1, -2)$ is parallel to the weight-space of the $U(1)$ subgroup and is such that it connects weights inside the diagram. In particular, $(1, 1, -2)$ connects $\tilde{\gamma}(M^1, \lambda)$ with different $\lambda_1\lambda_1' \cdots$. Thus an upper limit $\tilde{\gamma}_{Ul}$ for the maximal multiplicity $\tilde{\gamma}^M(M^1)$ is obtained by counting the maximal number of weights of $D(M)$ which are related by $(1, 1, -2)$. This number is obtained as follows: If $\lambda_{\max}$ is the maximal $U(1)$ weight of $D(M)$ and $\lambda_{\min}$ is the minimal $U(1)$ weight of $D(M)$, then

$$\tfrac{1}{2}(\lambda_{\max} - \lambda_{\min}) = \frac{M_1}{2} - \frac{M_3}{2},$$

and the upper limit $\tilde{\gamma}_{Ul}$ for the maximal multiplicity $\tilde{\gamma}^M(M^1)$ is given by

$$\tilde{\gamma}_{Ul} = n + 1,$$

with

$$n \leq \frac{M_1}{2} - \frac{M_3}{2}, \quad n \text{ closest integer.}$$

Thus

$$\tilde{\gamma}^M(M^1) \leq n + 1.$$

APPENDIX A

The following is an example of an application of the Method A given at the end of Sec. 3 for the determination of the inner multiplicity of a weight belonging to a particular irreducible representation of $SU(n)$.

We take $n = 6$ and choose the weight $(N_1, \cdots, N_6) = (4, 5, 4, 4, 0, 4)$ of the irreducible representation $(\lambda_1, \cdots, \lambda_6) = (7, 5, 4, 3, 2, 0)$, which is of dimensionality 190512.

Imposing the conditions

$$\sum_{i=1}^{5} \lambda_i' = \sum_{i=1}^{5} N_i = 17,$$

and $(N_1, \cdots, N_5)_{\min} - \lambda'_5 \geq 0$ (which yields $\lambda'_5 = 0$), one finds the partitions

λ'_1	λ'_2	λ'_3	λ'_4	λ'_5
7	5	3	2	0
7	4	4	2	0
6	5	4	2	0
7	4	3	3	0
6	5	3	3	0
6	4	4	3	0
5	5	4	3	0

Applying step 1 again with $n \rightarrow n-1$ gives the partitions

7	5	3	2
7	4	4	2
6	5	4	2
7	4	3	3
6	5	3	3
6	4	4	3
5	5	4	3

and a third application gives the partitions

7	4	2	(2)
6	5	2	(2)
7	3	3	(2)
6	4	3	(6)
5	5	3	(4)
5	4	4	(2)

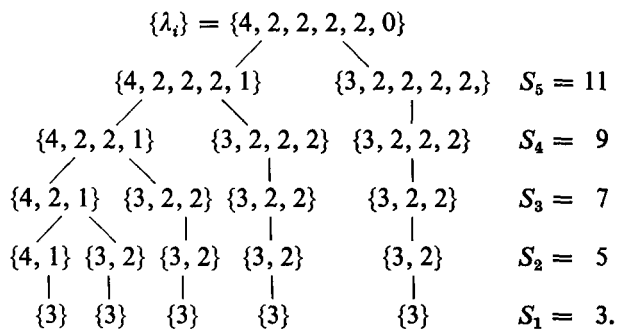
with the numbers in parentheses being the multiplicity of the particular partition.

Step 1 applied to (7, 4, 2) yields 3 partitions, each of which is consistent with $m_{ii} = N_1 = 4$, giving a contribution of $3 \times 2 = 6$ to the multiplicity. Continuing this process for all the partitions, one finds

$$\gamma(N) = 3 \times 2 + 2 \times 2 + 1 \times 2 + 2 \times 6 + 1 \times 4 + 1 \times 2 = 30.$$

APPENDIX B

The multiplicity $\gamma(1, 0, 0, 0, 0, -1)$ of the weight $m = (1, 0, 0, 0, 0, -1)$ of the $SU(6)$ representation $D(2, 0, 0, 0, 0, -2)$ is calculated using Method B. Then,



Therefore,

$$\gamma(1, 0, 0, 0, 0, -1) = 5.$$

APPENDIX C

In the following the diagonal generators $(I_3 Y)$ of $SU(3)$ are constructed for the irreducible representations $\bar{3}$ and 8. Using for the self-representation

$$I_3 = \begin{pmatrix} \frac{1}{2} & & \\ & -\frac{1}{2} & \\ & & 0 \end{pmatrix} \quad Y = \begin{pmatrix} \frac{1}{3} & & \\ & \frac{1}{3} & \\ & & -\frac{2}{3} \end{pmatrix},$$

by Eq. (39) it follows that the diagonal elements of I_3 and Y in any irreducible representation are given by

$$I_3 = \tfrac{1}{2}(N_1 - N_2),$$

$$Y = \tfrac{1}{3}(N_1 + N_2 - 2N_3),$$

where the occupation numbers (N_1, N_2, N_3) are defined by Eq. (42).

Thus for $\bar{3}$, one has:

Young diagram	N_1	N_2	N_3	I_3	Y
$\begin{array}{|c|}\hline 2 \\ \hline 3 \\ \hline \end{array}$	0	1	1	$-\frac{1}{2}$	$-\frac{1}{3}$
$\begin{array}{|c|}\hline 1 \\ \hline 3 \\ \hline \end{array}$	1	0	1	$\frac{1}{2}$	$-\frac{1}{3}$
$\begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline \end{array}$	1	1	0	0	$\frac{2}{3}$

and for 8:

Young diagram	N_1	N_2	N_3	I_3	Y
$\begin{array}{\|c\|c\|}\hline 1 & 1 \\ \hline 2 \\ \cline{1-1}\end{array}$	2	1	0	$\frac{1}{2}$	1
$\begin{array}{\|c\|c\|}\hline 1 & 2 \\ \hline 2 \\ \cline{1-1}\end{array}$	1	2	0	$-\frac{1}{2}$	1
$\begin{array}{\|c\|c\|}\hline 1 & 1 \\ \hline 3 \\ \cline{1-1}\end{array}$	2	0	1	1	0
$\begin{array}{\|c\|c\|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}$	1	1	1	0	0
$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}$	1	1	1	0	0
$\begin{array}{\|c\|c\|}\hline 2 & 2 \\ \hline 3 \\ \cline{1-1}\end{array}$	0	2	1	−1	0
$\begin{array}{\|c\|c\|}\hline 1 & 3 \\ \hline 3 \\ \cline{1-1}\end{array}$	1	0	2	$\frac{1}{2}$	−1
$\begin{array}{\|c\|c\|}\hline 2 & 3 \\ \hline 3 \\ \cline{1-1}\end{array}$	0	1	2	$-\frac{1}{2}$	−1.

APPENDIX D

The multiplicity $\tilde{\gamma}^{(1,0,-1)}(\frac{1}{2}, -\frac{1}{2})$ of the $SU(2)$ representation $D(\frac{1}{2}, -\frac{1}{2})[\equiv D(\frac{1}{2})]$, which is contained in the restriction of the $SU(3)$ representation $D(1, 0, -1)$ (octet) to a $SU(2)$ subgroup, is calculated. Then, from Eq. (62), it follows that

$$\begin{aligned}\tilde{\gamma}(\tfrac{1}{2}, -\tfrac{1}{2}) &= \sum_{S'\in W'} \sum_{m\in D(1,0,-1)} \gamma(m) \\ &\quad \times \delta_{S'}\delta_{S'[L^1(m)+(\frac{1}{2},-\frac{1}{2})],(1,-1)} \\ &= \sum_{S'\in W'} \sum_{m\in D(1,0,-1)} \gamma(m) \\ &\quad \times \delta_{S'}\delta_{S'[m_1+(m_3/2)+\frac{1}{2},-m_1-(m_3/2)-\frac{1}{2}],(1,-1)} \\ &= \sum_{m\in D(1,0,-1)} \{\gamma(m)\delta_{m_1+(m_3/2)+\frac{1}{2},1} \\ &\quad - \gamma(m)\delta_{m_1+(m_3/2)+\frac{1}{2},-1}\} \\ &= \gamma(1, 0, -1) + \gamma(0, -1, 1) = 1 + 1 = 2.\end{aligned}$$

For $\tilde{\gamma}^{(1,0,-1)}(0, 0)$, one obtains:

$$\begin{aligned}\tilde{\gamma}(0, 0) &= \sum_{m\in D(1,0,-1)} \{\gamma(m)\delta_{m_1+(m_3/2)+\frac{1}{2},\frac{1}{2}} \\ &\quad - \gamma(m)\delta_{m_1+(m_3/2)+\frac{1}{2},-\frac{1}{2}}\} \\ &= \gamma(0, 0, 0) - \gamma(-1, 1, 0) = 2 - 1 = 1.\end{aligned}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Theory of the Propagation of the Plane-Wave Disturbances in a Distribution of Thermal Neutrons

JAMES J. DUDERSTADT
Division of Engineering and Applied Science, California Institute of Technology
Pasadena, California

(Received 8 January 1968)

The propagation of plane-wave disturbances in a neutron gas in thermal equilibrium with a moderator is studied using the linearized Boltzmann equation. The eigenvalue spectrum of the appropriate Boltzmann transport operator is investigated for both polycrystalline and noncrystalline media, and several necessary conditions for the existence of a point spectrum (and hence for the existence of plane-wave propagation) are discussed. Techniques are presented which allow the solution of various boundary-value problems using this spectral representation.

I. INTRODUCTION

Recent experiments[1] with time-dependent neutron sources have stimulated detailed investigations of the linearized Boltzmann equation describing neutron transport. Most attention has been directed towards the initial-value problem[2] in which one studies the time relaxation of a neutron pulse injected into a sample of the material of interest. However, there has also been considerable experimental effort[3] directed towards the use of modulated sources to excite wavelike disturbances in the neutron distribution within a material. Since the propagation of these disturbances depends upon material composition and geometry, measurements of the relative attenuation and phase lag of the disturbance at various locations within the medium can be used to obtain information about the neutron-transport properties of the medium. Such "neutron-wave" experiments have failed to receive detailed theoretical study to date.

We propose to use the linearized Boltzmann equation to study the propagation of plane-wave disturbances in a neutron gas in thermal equilibrium with various moderating materials (gaseous, liquid, and solid). Such a study takes on an added significance when it is realized that forced sound-wave propagation in rarefied monatomic gases (with truncated intermolecular potentials) can be described by equations with a similar mathematical structure.[4] Much of the theory seems similarly applicable to the description of longitudinal ion and electron waves in low-density plasmas.[5]

Let us begin by considering the steady-state response of a neutron distribution to an oscillating neutron source of fixed frequency ω. The linearized Boltzmann equation under the assumption of plane symmetry (which will be relaxed in Sec. II) becomes

$$\frac{\partial f}{\partial t} + \mu v \frac{\partial f}{\partial x} + v\Sigma_t(v) f(x, \mu, v, t) = \int_{-1}^{+1} d\mu' \int_0^{\infty} dv' v' \Sigma_s(v' \to v, \mu' \to \mu) f(x, \mu', v', t) + S(x, \mu, v) e^{i\omega t}, \quad (1)$$

where $f(x, \mu, v, t)$ is the neutron-distribution function of space x, direction cosine μ from the x axis, speed v, and time t; $S(x, \mu, v)e^{i\omega t}$ is the source distribution; $\Sigma_t(v)$ is the macroscopic total cross section; and $\Sigma_s(v' \to v, \mu' \to \mu)$ is the differential-scattering cross section ("scattering kernel").

II. SPECTRAL ANALYSIS OF THE BOLTZMANN WAVE OPERATOR

A very powerful technique for analyzing such boundary-value problems involves the use of separation of variables to arrive at an eigenvalue problem. To be more precise, we seek elementary solutions to (1) in the form of plane waves

$$f(x, \mu, v, t) = F(\kappa; \mu, v) \exp(-\kappa x + i\omega t), \quad (2)$$

where κ is an arbitrary complex constant, $F(\kappa; \mu, v)$ is an as yet undetermined function of μ and v, and ω is real and *fixed* at the source frequency. Substituting this ansatz into the homogeneous version of (1) yields a restriction on κ and $F(\kappa; \mu, v)$:

$$[i\omega + v\Sigma_t(v) - \kappa\mu v] F(\kappa; \mu, v) = \int_{-1}^{+1} d\mu' \int_0^{\infty} dv' v' \Sigma_s(v' \to v, \mu' \to \mu) F(\kappa; \mu', v'). \quad (3)$$

But recognize that this two-dimensional integral equation is just an eigenvalue problem for the eigenvalues κ and the corresponding eigenfunctions

[1] R. E. Uhrig, IAEA Symp. Pulsed Neutron Res. **2**, 659 (1965).
[2] K. Beckurts, IAEA Symp. Pulsed Neutron Res. **1**, 1 (1965).
[3] R. B. Perez and R. S. Booth, IAEA Symp. Pulsed Neutron Res. **2**, 701 (1965).
[4] H. Grad, J. SIAM Appl. Math. **14**, 935 (1966).
[5] N. G. van Kampen, Physica **21**, 949 (1955).

$F(\kappa; \mu, v)$. It is convenient to define

$$\psi_\kappa(\mu, v) \equiv (\mu v/M(v))^{\frac{1}{2}} F(\kappa; \mu, v),$$

$$\tilde{\Sigma}_s(v' \to v, \mu' \to \mu) \equiv (v' M(v')/v M(v))^{\frac{1}{2}} \Sigma_s(v' \to v, \mu' \to \mu), \quad (4)$$

where $M(v)$ is the Maxwellian speed distribution. Then (3) can be rewritten in the more standard form

$$A\psi_\kappa = \kappa \psi_\kappa, \quad (5)$$

where the *Boltzmann wave operator* A is defined as:

$$A = A_1 + A_2,$$

$$A_1 \equiv \left[\frac{i\omega}{\mu v} + \frac{\Sigma_t(v)}{\mu}\right], \quad (6)$$

$$A_2 \equiv \left[-\int_{-1}^{+1} d\mu' \int_0^\infty dv' \frac{\tilde{\Sigma}_s(v' \to v, \mu' \to \mu)}{(\mu'\mu)^{\frac{1}{2}}} \cdot\right].$$

Notice that the principle of detailed balance[6]

$$v' M(v') \Sigma_s(v' \to v, \mu' \to \mu) = v M(v) \Sigma_s(v \to v', -\mu \to -\mu')$$

implies that $\tilde{\Sigma}_s(v' \to v, \mu' \to \mu)$ is a symmetric kernel.

In principle, then, if (5) could be solved for the eigenvalues and eigenfunctions of A, and it could then be demonstrated that these eigenfunctions formed a complete biorthogonal set, one would have available a powerful tool for treating problems involving not only infinite sourceless media, but also media of finite dimension and containing distributed sources. To treat these latter problems one would merely expand their general solution in the eigenfunctions of A, and then use boundary conditions to evaluate the expansion coefficients. Thus in some sense the problem of neutron-wave propagation has been reduced to a study of the eigenvalue spectrum of the Boltzmann wave operator A.

Unfortunately, (5) is more complicated than the standard eigenvalue problems of mathematical physics. The operator A is (1) not completely continuous, (2) unbounded, and (3) non-self-adjoint. However, some information can be obtained about its spectrum $\sigma(A)$. First define the sets

$$I_\mu \equiv \{\mu : \mu \in [-1, +1]\}, \quad I_v \equiv \{v : v \in [0, \infty)\},$$

$$G \equiv I_\mu \otimes I_v.$$

Now consider A to operate on elements of the Hilbert space $\mathfrak{L}_2(G)$ of complex-valued square-integrable functions defined over G with inner product

$$(f, g) \equiv \int_{-1}^{+1} d\mu \int_0^\infty dv\, \overline{f(\mu, v)}\, g(\mu, v).$$

We now discuss $\sigma(A)$ for several specific cases:

A. Noncrystalline Media

We characterize noncrystalline media by

(i) monotonically decreasing $\Sigma_t(v)$ such that

$$\lim_{v \to 0} \Sigma_t(v) \to \Sigma_t^0/v \quad \text{and} \quad \lim_{v \to \infty} \Sigma_t(v) = \Sigma_M > 0;$$

(ii) a scattering operator

$$\tilde{S} \equiv \int_{-1}^{+1} d\mu' \int_0^\infty dv' \tilde{\Sigma}_s(v' \to v, \mu' \to \mu),$$

which is completely continuous.[7]

Under these restrictions one can identify A_1 as a normal (but non-self-adjoint) unbounded multiplicative operator. The apparent $(\mu\mu')^{\frac{1}{2}}$ divergence in the symmetric kernel of the integral operator A_2 is only an unfortunate notational consequence of the symmetrizing process defined by (4). It must be remembered that when A_2 operates on functions

$$\psi_\kappa(\mu, v) \equiv [\mu v/M(v)]^{\frac{1}{2}} F(\kappa; \mu, v),$$

the $(\mu)^{\frac{1}{2}}$ terms cancel. Hence, if one considers only solutions to the original eigenvalue problem (3) which are $\mathfrak{L}_2(G)$, then the functions $\psi_\kappa(\mu, v)$ defined by Eq. (4) are automatically restricted to a somewhat smaller class than $\mathfrak{L}_2(G)$, namely the class of functions for which

$$\left(\frac{M(v)}{\mu v}\right)^{\frac{1}{2}} \psi_\kappa(\mu, v) \in \mathfrak{L}_2(G).$$

Hence in this sense, assuming $\tilde{S}$ to be completely continuous is sufficient to imply that A_2 is completely continuous.

We can now prove:

Theorem: The Boltzmann wave operator A for noncrystalline media decomposes the spectral κ plane as follows:

$\sigma_c(A) = C$, where

$$C \equiv \left\{\kappa : \kappa = \frac{i\omega}{\mu v} + \frac{\Sigma_t(v)}{\mu}, \mu \in [-1, +1], v \in [0, \infty)\right\};$$

$\sigma_p(A)$: a point set containing those $\kappa \notin C$ such that $f(\kappa; \mu, v)$

$$= \int_{-1}^{+1} d\mu' \int_0^\infty dv' \left[\frac{v' \Sigma_s(v' \to v, \mu' \to \mu)}{i\omega + v' \Sigma_t(v') - \kappa \mu' v'}\right] f(\kappa; \mu', v') \quad (7)$$

[6] M. M. R. Williams, *The Slowing Down and Thermalization of Neutrons* (North-Holland Publ. Co., Amsterdam, 1966).

[7] I. Kuščer and N. Corngold, Phys. Rev. **139A**, 981 (1965).

possesses nontrivial solutions $f(\kappa; \mu, v)$;

$\sigma_r(A)$: an empty set;

$\rho(A)$: all other points of the κ plane not contained in the spectrum $\sigma(A)$.

Proof: To show that the continuous spectrum $\sigma_c(A)$ is just the set C, one can use a theorem due to Weyl and von Neumann[8] to find that since A_1 is a normal operator and A_2 is a self-adjoint completely continuous operator, $\sigma_c(A_1 + A_2) = \sigma_c(A_1)$. Hence we need only determine the continuous spectrum of A_1.

For any $\kappa \in C$, define μ_κ and v_κ by

$$\kappa = \frac{i\omega}{\mu_\kappa v_\kappa} + \frac{\Sigma_t(v_\kappa)}{\mu_\kappa}.$$

Now consider the sequence of functions

$$\varphi_\delta(\mu, v) = \begin{cases} [M(v)/\delta^2]^{\frac{1}{2}}, & v_\kappa \leq v < v_\kappa + \delta, \\ & \mu_\kappa \leq \mu < \mu_\kappa + \delta, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Notice that $\|\varphi_\delta\| > 0$, $(f, \varphi_\delta) \to 0$ for any $f \in \mathfrak{L}_2(G)$ as $\delta \to 0$, while $\|(A_1 - \kappa)\varphi_\delta\| \to 0$ as $\delta \to 0$. Hence, by the Weyl criterion,[8] we can conclude that $\kappa \in \sigma_c(A_1)$ for all $\kappa \in C$, subject to the assumption that none of the point spectrum $\sigma_p(A)$ is "imbedded" in C. To show that these are the only points $\kappa \in \sigma_c(A_1)$, it is an easy matter to demonstrate that $(A_1 - \kappa)^{-1}$ is a bounded operator for $\kappa \notin C$. Hence we can conclude that

$$C = \sigma_c(A_1) = \sigma_c(A_1 + A_2) = \sigma_c(A).$$

If we take the complex conjugate of Eq. (5), we find that $\bar{\kappa} \in \sigma_p(A^\dagger) \Rightarrow \kappa \in \sigma_p(A)$ (since A_2 is real and self-adjoint). This is sufficient to imply that the residual spectrum $\sigma_r(A)$ is empty.[9]

To derive the characteristic equation or "dispersion law" for the point eigenvalues, restrict $\kappa \notin C$, divide (3) by $[i\omega + v\Sigma_t(v) - \kappa\mu v]$, multiply through by $v'\Sigma_s(v' \to v, \mu' \to \mu)$, and integrate over (μ', v') to arrive at the homogeneous Fredholm equation (7) for $f(\kappa; \mu, v)$

$$\equiv \int_0^\infty dv' \int_{-1}^{+1} d\mu'\, v'\Sigma_s(v' \to v, \mu' \to \mu) F(\kappa; \mu', v').$$

The proof is now complete.

A sketch of the spectral κ plane for noncrystalline media is given in Fig. 1. It is of interest to notice the symmetry between the first and third quadrants which corresponds to plane waves propagating in the $x \gtrless 0$ directions. Also notice that the continuous spectrum $\sigma_c(A) = C$ is a two-dimensional set in the complex κ plane.

[8] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Frederick Ungar Publ. Co., New York, 1955).

[9] B. Friedman, *Principles and Techniques of Applied Mathematics* John Wiley & Sons, Inc., New York, 1956).

B. Polycrystalline Media

The wave operator A assumes a somewhat different form for polycrystalline media such as graphite or beryllium: $\Sigma_t(v)$ is no longer smooth and monotonic in v and indeed can exhibit essentially discontinuous behavior; in addition, the scattering operator $\tilde{S}$ is no longer completely continuous. Both of these effects are due to the diffraction of the neutron wavefunction by the ordered crystal lattice (elastic scattering).[6]

We analyze neutron-wave propagation in polycrystalline media by separating the scattering kernel into its inelastic and elastic components:

$$\Sigma_s(v' \to v, \mu' \to \mu) = \Sigma_i(v' \to v, \mu' \to \mu) + \Sigma_e(v' \to v, \mu' \to \mu).$$

We now assume that:

(i) $\Sigma_i(v' \to v, \mu' \to \mu)$ or one of its iterates is square-integrable over G. [Since $\Sigma_i(v' \to v, \mu' \to \mu)$ is similar to the scattering kernel for noncrystalline media, this is a reasonable assumption.]

(ii) We can approximate the elastic component by $\Sigma_e(v' \to v, \mu' \to \mu) = \Sigma_e(v)R(\mu' \to \mu)\delta(v' \to v)$, where $R(\mu' \to \mu)$ or one of its iterates is square-integrable over I_μ. [Since the scattering kernel for elastic coherent scattering contains terms of the form $\delta(\underline{v}' - \underline{v})$, this assumption is open to some challenge. However, summation over crystal-lattice vectors and averaging over the orientation of the polycrystals smooth out this apparent angular singularity and also assure the validity of the detailed balance condition necessary for symmetrizing the scattering kernel.[6]]

(iii) $\Sigma_e(v) = 0$ for $v < v_B$ (elastic scattering vanishes below the Bragg cutoff speed v_B); $\Sigma_t(v) = \Sigma_t^0/v$ for $v < v_B$ (reasonable to within current cross section measurement accuracy); $\Sigma_e(v)$ is bounded for $v > v_B$, while $\Sigma_e(v) \to 0$ as $v \to \infty$.

Previous experience in neutron-transport calculations[10] leads one to believe that these assumptions are essentially valid for the purposes of this analysis.

Our linear operator A is now

$$\begin{aligned} A &= A_1 + A_2 + A_3, \\ A_1 &\equiv \left[\frac{i\omega}{\mu v} + \frac{\Sigma_t(v)}{\mu}\right], \\ A_2 &\equiv \left[-\int_{-1}^{+1} d\mu' \int_0^\infty dv' \frac{\tilde{\Sigma}_i(v' \to v, \mu' \to \mu)}{(\mu'\mu)^{\frac{1}{2}}} \cdot\right], \\ A_3 &\equiv \left[-\Sigma_e(v) \int_{-1}^{+1} d\mu' \frac{\tilde{R}(\mu' \to \mu)}{(\mu'\mu)^{\frac{1}{2}}} \cdot\right]. \end{aligned} \quad (9)$$

[10] N. Corngold, IAEA Symp. Pulsed Neutron Res. 1, 119 (1965).

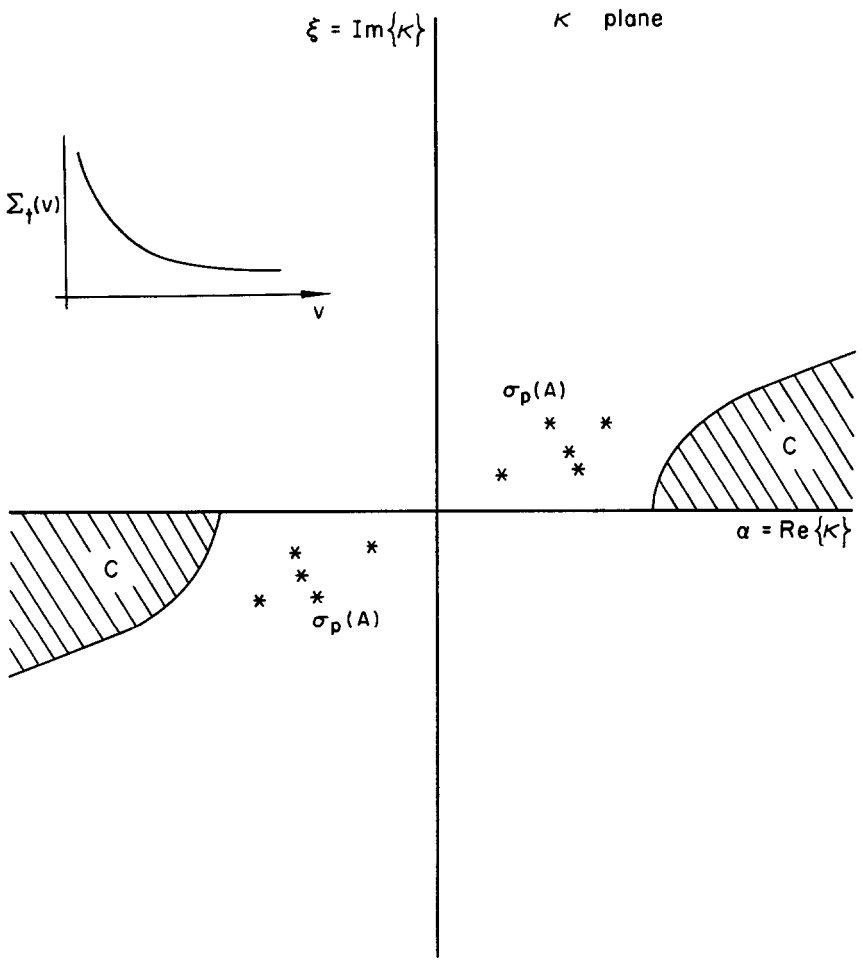


FIG. 1. A sketch of the spectral κ plane for noncrystalline media.

Again A_1 is normal and unbounded, while A_2 is symmetric and, by assumption, completely continuous. The operator A_3 due to elastic scattering is bounded and symmetric, but it is not completely continuous. Hence we expect it to add additional continuous spectra. We now find:

Theorem: The Boltzmann wave operator A for polycrystalline media decomposes the spectral κ plane as follows:

$\sigma_c(A) = C \cup \Gamma$, where

$$C \equiv \left\{\kappa : \kappa = \frac{i\omega}{\mu v} + \frac{\Sigma_t(v)}{\mu}, \mu \in [-1, +1], v \in [0, \infty)\right\},$$

and Γ is the set of all κ such that

$$\mathfrak{L}_e\{h\} \equiv h(\kappa; \mu) - v\Sigma_e(v) \times \int_{-1}^{+1} d\mu' \left[\frac{R(\mu' \to \mu)}{i\omega + v\Sigma_t(v) - \kappa\mu' v}\right] h(\kappa; \mu') = 0, \quad v \in [0, \infty), \quad (10)$$

possesses nontrivial solutions $h(\kappa; \mu)$.

$\sigma_p(A)$: A point set containing those $\kappa \notin C \cup \Gamma$ such that

$$f_i(\kappa; \mu, v) = \int_{-1}^{+1} d\mu' \int_0^{\infty} dv' \left[\frac{v'\Sigma_i(v' \to v, \mu' \to \mu)}{i\omega + v'\Sigma_t(v') - \kappa\mu' v'}\right] \times \mathfrak{L}_e^{-1}\{f_i(\kappa; \mu', v')\} \equiv P(\kappa) f_i, \quad (11)$$

possesses nontrivial solutions $f_i(\kappa; \mu, v)$.

$\sigma_r(A)$: An empty set.

$\rho(A)$: All other points of the κ plane not contained in the spectrum $\sigma(A)$.

Proof: Unfortunately since $A_1 + A_3$ is not normal, we cannot use the Weyl–von Neumann theorem to conclude that $\sigma_c(A) = \sigma_c(A_1 + A_3)$. We must proceed by an alternative route.

Using the sequence defined by (8), we can demonstrate that $\|\varphi_\delta\| > 0$, $\|(A - \kappa)\varphi_\delta\| \to 0$ as $\delta \to 0$ for $\kappa \in C \cup \Gamma$. This is sufficient[11] to imply that $C \subset \sigma(A)$. [Although not necessarily $C \subset \sigma_c(A)$.] By considering the spectrum of $A_1 + A_3$, one can arrive at (10) as the

[11] E. Hille and R. S. Phillips, *Functional Analysis and Semigroups* (American Mathematical Society, Providence, R.I., 1957).

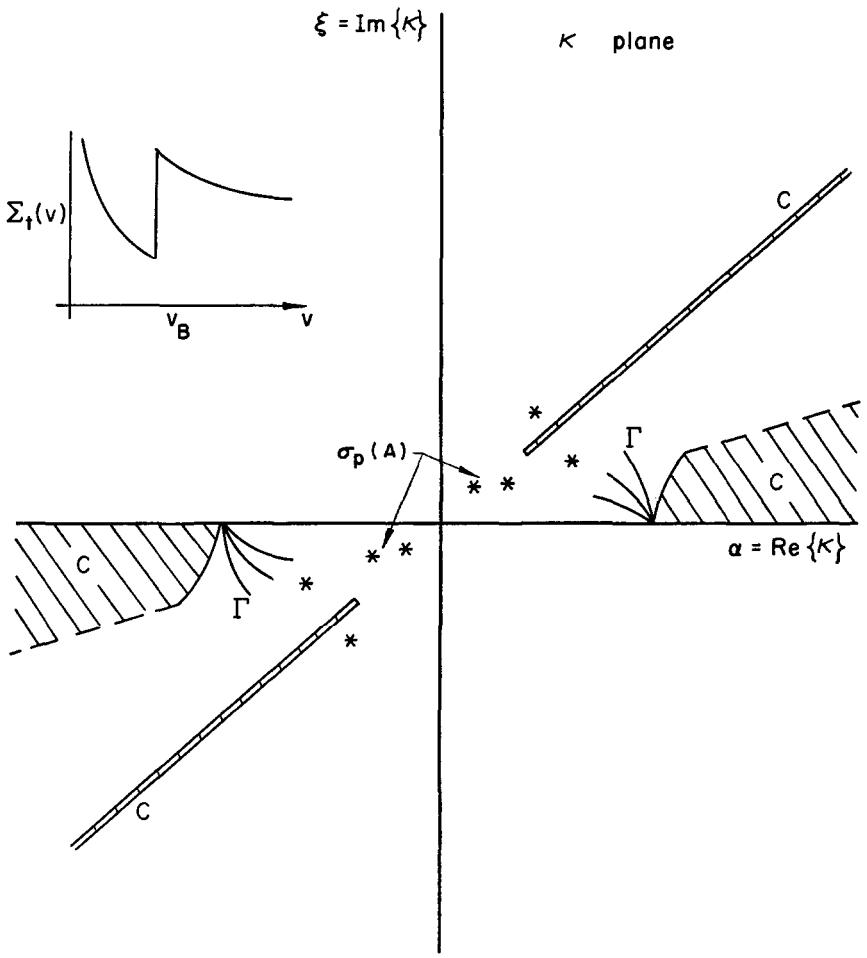


FIG. 2. A sketch of the spectral κ plane for polycrystalline media.

dispersion law determining Γ, and then using the φ_δ, demonstrate that $\Gamma \subset \sigma(A)$.

To demonstrate that $\sigma_c(A)$ must be contained in $C \cup \Gamma$, we first restrict $\kappa \notin C \cup \Gamma$ and manipulate Eq. (3) into the form (11). Now notice that the operator $P(\kappa)$ defined by (11) is analytic in κ for $\kappa \notin C \cup \Gamma$. Furthermore since $\Sigma_e(v) \to 0$ as $v \to \infty$, and $[i\omega + v\Sigma_t(v) - \kappa\mu v]^{-1}$ is bounded for all μ and v, one can demonstrate that $P(\kappa)$ is completely continuous for each $\kappa \notin C \cup \Gamma$. But a theorem due to Šmul'yan[12] states that if an operator $P(\kappa)$ is analytic and completely continuous for κ in some domain D, then the domain D must either be entirely in the spectrum of P, or D can contain no limit points of the spectrum. Applying this theorem to our operator $P(\kappa)$ indicates that since the first alternative is clearly impossible, there can be no limit points of the spectrum of $P(\kappa)$ in the region $\kappa \notin C \cup \Gamma$. Now, since (11) is related to (3) by straightforward manipulation, the spectrum of $P(\kappa)$ is identical to the spectrum of A for $\kappa \notin C \cup \Gamma$. Hence we can conclude that there can be no limit points of the spectrum of A (and hence continuous spectra of A) in the region $\kappa \notin C \cup \Gamma$.

Since the symmetry of A is not disturbed by the addition of A_3, we can again conclude that $\sigma_r(A)$ is empty.

Now we have seen that $\sigma_p(A)$ is determined by those values of κ such that (11) possesses nontrivial solutions. Since C and Γ are continuous sets, $\sigma_p(A)$ cannot be contained in $C \cup \Gamma$ [provided we again rule out the possibility of point eigenvalues "imbedded" in $\sigma_c(A)$.] Since σ_r is empty, the only remaining possibility is that $\sigma_c(A) = C \cup \Gamma$.

The κ-plane structure for polycrystalline media has been sketched in Fig. 2. The fragmentation of C has been caused by the discontinuity at $v = v_B$, while the $1/v$ behavior for $v < v_B$ causes a portion of C to degenerate into a line continuum. The location of the Γ curves are also indicated.

[12] Yu. Šmul'yan, Am. Math. Soc. Translations (2) **10**, 34 (1958); see also I. Kuščer, *IAEA Symposium on Neutron Thermalization and Reactor Spectra, Ann Arbor, 1967* (IAEA, Vienna, 1968), SM-96/103.

C. Multidimensional Transport Effects

Thus far the analysis has been restricted to the propagation of plane neutron waves in an infinite medium. However, most neutron-wave experiments involve the excitation of such waves in parallelepiped geometries. While it is reasonable to treat these geometries as infinite in length (neglecting wave reflections from the far boundary), the effects of finite transverse dimension cannot be so easily ignored.

We shall now proceed to indicate the necessary modifications in the analysis of $\sigma(A)$ following the work of Williams.[13] The general Boltzmann equation now becomes

$$\frac{\partial f}{\partial t} + \mu v \frac{\partial f}{\partial x} + v(1-\mu^2)^{\frac{1}{2}}\left(\cos\psi \frac{\partial f}{\partial y} + \sin\psi \frac{\partial f}{\partial z}\right) + v\Sigma_t(v) f = \int_0^\infty dv' \int d\hat{\Omega}' v' \Sigma_s(v' \to v, \hat{\Omega}' \cdot \hat{\Omega}) f(\mathbf{r}, \mu', \psi', v', t), \tag{12}$$

where ψ is the azimuthal angle and $\hat{\Omega}$ is the unit vector (ψ, μ). To treat the spatial dependence in the transverse y–z dimensions, it is convenient to use "asymptotic reactor theory." [13] That is, we assume a plane-wave ansatz of the form

$$f(\mathbf{r}, \mu, \psi, v, t) = F(\rho; \mu, \psi, v) e^{-\rho x} e^{i(B_y y + B_z z)} e^{i\omega t}, \tag{13}$$

where B_y, B_z, and ω are real and fixed, while ρ and $F(\rho; \mu, \psi, v)$ are as yet unspecified. Substituting this ansatz into (12), we arrive at the eigenvalue problem for ρ, the spatial eigenvalue for finite transverse dimensions,

$$[i\omega + v\Sigma_t(v) + iv(1-\mu^2)^{\frac{1}{2}}(B_y \cos\psi + B_z \sin\psi) - \rho\mu v] F(\rho; \mu, \psi, v) = SF. \tag{14}$$

Of course for infinite transverse dimensions, B_y and $B_z \to 0$ and (14) reduces to equation (3) (and hence $\rho \to \kappa$). The Boltzmann wave operator A_T for finite transverse dimensions becomes

$$A_T \equiv \left[\frac{i\omega}{\mu v} + \frac{\Sigma_t(v)}{\mu} + i\frac{(1-\mu^2)^{\frac{1}{2}}}{\mu}(B_y \cos\psi + B_z \sin\psi)\right] + \left[-\int d\hat{\Omega}' \int_0^\infty dv' \frac{\tilde{\Sigma}_s(v' \to v, \hat{\Omega} \cdot \hat{\Omega}')}{(\mu'\mu)^{\frac{1}{2}}} \cdot\right]. \tag{15}$$

The analysis of the eigenvalue spectrum of A_T is quite similar to that of A. The primary differences in the spectral theorems are in the identification of

$$C \to C_T \equiv \left\{\rho: \rho = \frac{\Sigma_t(v)}{\mu} + i\left[\frac{\omega}{\mu v} + \frac{(1-\mu^2)^{\frac{1}{2}}}{\mu}(B_y \cos\psi + B_z \sin\psi)\right], \quad \mu \in [-1, +1], \psi \in [0, 2\pi], v \in [0, \infty]\right\}.$$

$\Gamma \to \Gamma_T \equiv$ the set of all κ such that

$$h(\rho; \mu, \psi) - v\Sigma_e(v) \int d\hat{\Omega}' \left[\frac{R(\hat{\Omega}' \to \hat{\Omega})}{i\omega + v\Sigma_t(v) + iv(1-\mu'^2)^{\frac{1}{2}}(B_y \cos\psi' + B_z \sin\psi') - \rho\mu' v}\right] h(\rho; \mu', \psi') = 0 \quad v \in [0, \infty).$$

For isotropic scattering kernels, one can demonstrate[13] that the point spectrum $\sigma_p(A)$ is shifted from κ to $\rho = (\kappa^2 + B_\perp^2)^{\frac{1}{2}}$ where $B_\perp^2 \equiv B_y^2 + B_z^2$.

Hence the adaptation of the spectral theory to systems of finite transverse dimension can be accomplished by modifying $\sigma_c(A)$ and $\sigma_p(A)$ directly.

D. The Point Spectrum $\sigma_p(A)$

If we associate a plane-wave mode (or eigenfunction) to each point κ in the spectrum $\sigma(A)$, we can consider the general oscillatory disturbance in the particle distribution to be composed of a superposition of these plane-wave modes. The neutron-wave experiment makes the implicit assumption that sufficiently far from the modulated neutron source, one of these modes will dominate. Stated mathematically then, the experimental goal is to measure the discrete eigenvalue $\kappa_0 \in \sigma_p(A)$ with least real part (damping) since this should correspond to the dominant mode at large distances from the source. Thus an analysis of the point spectrum is usually sufficient to allow quantitative comparisons with experimental results.

Unfortunately the dispersion laws (7) and (11) for general scattering kernels represent formidable mathematical problems since they are not only implicit eigenvalue problems (i.e., cannot be factored into the form $Hf_\kappa = \kappa f_\kappa$) but non-self-adjoint as well. One must resort to numerical calculation schemes to study the detailed location of the point eigenvalues. However, it is possible to establish various *necessary* conditions for the existence of the point spectrum. Consider, then, the following theorem:

Theorem: There exist certain critical bounds on the source frequency ω, the absorption $\Sigma_a(v)$, and the

[13] M. M. R. Williams, Nukleonik 9, 305 (1967).

transverse buckling $B_\perp^2$, such that if

$$\omega > \omega^*, \quad \text{or} \quad \Sigma_a > \Sigma_a^*, \quad \text{or} \quad B_\perp^2 > B_\perp^{*2},$$

the point eigenvalue spectrum is empty.

Of course each of these bounds is dependent on the other two parameters, e.g., $\omega^* = \omega^*(\Sigma_a, B^2)$, etc. One can show in fact that increasing any two of the parameters lowers the bound on the third parameter *in most cases*. This fact is of particular interest in that it indicates the interrelations between these three independent experimental parameters in the neutron wave experiment.

Such theorems are well known in neutron transport theory,[14] and the proof of the theorem above contains no essential variations. This proof is somewhat involved, however, and will be deferred to Appendix A.

III. THE SOLUTION OF BOUNDARY-VALUE PROBLEMS

Since wave-propagation experiments perform asymptotic measurements far from the source, a knowledge of the eigenvalue spectrum of A is sufficient in most cases to allow the analysis of these experiments—and certainly suffices to determine which portions of the spectrum will dominate at large distances from the source. However, it is sometimes desirable to study the particle distribution function $f(x, \mu, v, t)$ in regimes in which a single plane-wave mode does not dominate the solution. For this reason, it is necessary to study how one might actually use the eigenfunctions corresponding to $\sigma(A)$ to solve various boundary-value problems which serve as mathematical models for the experiments.

Due to their asymptotic nature, it usually suffices to model these propagation experiments by using so-called *full-range* boundary-value problems. By definition, such boundary-value problems involve only boundary conditions which are given at a specific position in x for *all* μ and v. In particular such problems will involve eigenfunction expansions over the full range of the eigenvalue spectrum $\sigma(A)$. Our task is first to derive the eigenfunctions corresponding to those $\kappa \in \sigma(A)$ and then to indicate how these eigenfunctions may be used to solve such boundary-value problems. The treatment will be confined to the case of noncrystalline media.

It will be convenient to introduce a new notation. Notice that since C is an area in the κ plane, the continuous spectrum can only be specified completely by two parameters. That such area spectra arise in time-dependent transport problems has been recognized for some time[10,15]; however, only recently have methods been developed to yield the corresponding spectral representation. The principal work is due to Cercignani[15] and relies upon the theory of generalized analytic functions.[16] Much of the analysis in this section will be based upon Cercignani's work, although some extensions and modifications will be necessary.

It is useful to define a new independent variable

$$z \equiv [i\omega + v\Sigma_t(v)]/\mu v \equiv \vartheta + i\varphi. \tag{16}$$

This defines a one-to-one mapping of (μ, v) into the complex z plane [for monotonic $\Sigma_t(v)$ at least]. It allows (3) to be rewritten as

$$(z - \kappa)F(\kappa; z) = \iint_C K(z', z)F(\kappa; z')\, dz' \equiv f(\kappa; z), \tag{17}$$

where we have defined

$$\begin{aligned} F(\kappa; z) &\leftrightarrow F(\kappa; \mu(\vartheta, \varphi), v(\vartheta, \varphi)), \\ K(z', z) &\leftrightarrow \frac{v\Sigma_s(v' \to v, \mu' \to \mu)}{\mu v\, |\partial(\vartheta, \varphi)/\partial(\mu, v)|}(\vartheta, \varphi), \\ dz' &\leftrightarrow d\vartheta'\, d\varphi'. \end{aligned} \tag{18}$$

Notice that the spectrum of A in the κ plane remains unchanged under this transformation, while G maps onto C in the z plane.

It is now possible to construct the eigenfunctions $F(\kappa; z)$ corresponding to $\kappa \in \sigma(A)$. The *discrete eigenfunctions* corresponding to those $\kappa_l \in \sigma_p(A)$ can be found directly from (17) as

$$F(\kappa_l; z) = f(\kappa_l; z)/(z - \kappa_l), \quad \kappa_l \in \sigma_p(A), \tag{19}$$

while the dispersion law (7) for $\sigma_p(A)$ becomes, in z notation,

$$f(\kappa; z) = \iint_C \left[\frac{K(\zeta, z)}{\zeta - \kappa}\right] f(\kappa; \zeta)\, d\zeta, \quad \kappa \notin C. \tag{20}$$

We recognize that the eigenfunctions corresponding to the continuous spectrum $\sigma_c(A)$ will not be contained in $\mathfrak{L}_2(G)$. Hence our Hilbert space must be extended to include more general functions—and, in particular, distributions in the sense of Schwartz[9]—when discussing the *continuum eigenfunctions*. Generalizing the work of Case,[17] Bednarz and Mika[18] suggests the

[14] N. Corngold, Nucl. Sci. Eng. **19**, 80 (1964).

[15] C. Cercignani, Ann. Phys. **40**, 454 (1966).

[16] I. N. Vekua, *Generalized Analytic Functions* (Pergamon Press, Ltd., Oxford, 1962).

[17] K. M. Case and P. Zweifel, *Linear Transport Theory* (Addison-Wesley Publ. Co., Inc., Reading, Mass., 1967).

[18] R. Bednarz and J. Mika, J. Math. Phys. **4**, 1285 (1963).

form of these eigenfunctions will be

$$F(\kappa; z) = \frac{f(\kappa; z)}{(z - \kappa)} + \lambda(\kappa)\delta(\kappa - z), \quad \kappa \in C. \quad (21)$$

To obtain an expression for $\lambda(\kappa)$, multiply (21) by $K(z, u)$ and integrate over z to find

$$\lambda(\kappa) = [K(\kappa, u)]^{-1}\left[f(\kappa; u) - \iint_C \left[\frac{K(z, u)}{z - \kappa}\right] f(\kappa; z)\, dz\right], \quad \kappa \in C. \quad (22)$$

Notice that $f(\kappa; z)$ is still unspecified for the continuum eigenfunctions. It is found to be more convenient to use unnormalized eigenfunctions[18] when one is concerned with problems involving a general scattering kernel $\Sigma_s(v' \to v, \mu' \to \mu)$ in velocity-dependent transport theory. This added degree of freedom will be used to satisfy the boundary conditions for the specific problem of interest.

If one considers the equation adjoint to (17),

$$(\bar{z} - \bar{\kappa})F^\dagger(\kappa; z) = \iint \overline{K(z, z')}\, F^\dagger(\kappa; z')\, dz', \quad (23)$$

it is possible to demonstrate the usual biorthogonality relation for these eigenfunctions:

$$(F^\dagger_{\kappa_l}, F_{\kappa_m}) = \iint_C \overline{F^\dagger(\kappa_l; z)} F(\kappa_m; z)\, dz = \delta_{lm}(F^\dagger_{\kappa_l}, F_{\kappa_l}), \quad \kappa_l \in \sigma_p(A), \quad (24)$$

and a similar relation for $\kappa \in C$.

It is also possible to demonstrate the following full-range completeness property for the class of *integrable* functions:

Theorem: The set of functions $\{F(\kappa; z): \kappa \in \sigma(A)\}$ is complete for the class of all functions $\psi(z) \in \mathfrak{L}_1(\bar{C})$. (Here $\bar{C}$ denotes the closure of C.)

The proof of this theorem is deferred to Appendix B. For our purposes it suffices to note that if we wish to expand

$$\psi(z) = \sum_{\kappa_l \in \sigma_p(A)} a_l F(\kappa_l; z) + \iint_C F(\kappa; z)\, dz, \quad (25)$$

then the discrete expansion coefficients a_l can readily be evaluated by biorthogonality as

$$a_l = (F^\dagger_{\kappa_l}, \psi)/(F^\dagger_{\kappa_l}, F_{\kappa_l}),$$

while the continuum expansion coefficient $F(\kappa; z)$ is given by Eq. (B10) in Appendix B.

As an example of how these results may be applied, consider the problem of an oscillating plane source at the origin of an infinite medium. Then the Boltzmann equation of interest is (1) with a source term $S(x, \mu, v) = S(\mu, v)\delta(x)$. We shall take the physically significant boundary conditions at infinity:

$$\lim_{|x| \to \infty} |f(x, \mu, v, t)| = 0.$$

To obtain the boundary condition at the source, integrate (1) over x from $0 - \epsilon$ to $0 + \epsilon$ and then let $\epsilon \to 0$ to find

$$\mu v[f(0^+, \mu, v, t) - f(0^-, \mu, v, t)] = S(\mu, v)e^{i\omega t}.$$

To solve this problem by the method of spectral representation, we seek the solution in the form of expansions in the plane-wave eigenfunctions. Using the boundary conditions at infinity demands the expansions

$$f(x, \mu, v, t) = \begin{cases} \sum_{l^+=1}^{L/2} a_{l^+} F(\kappa_{l^+}; \mu, v) e^{-\kappa_{l^+} x + i\omega t} \\ \qquad + \iint_{\Delta C^+} e^{-\kappa x + i\omega t} F(\kappa; \mu, v)\, d\kappa, \quad x > 0, \\ \sum_{l^-=1}^{L/2} a_{l^-} F(\kappa_{l^-}; \mu, v) e^{-\kappa_{l^-} x + i\omega t} \\ \qquad + \iint_{\Delta C^-} e^{-\kappa x + i\omega t} F(\kappa; \mu, v)\, d\kappa, \quad x < 0. \end{cases} \quad (26)$$

Then applying the source boundary condition yields

$$\frac{S(\mu, v)}{\mu v} = \sum_{\kappa_l \in \sigma_p(A)} a_l F(\kappa_l; \mu, v) + \iint_C F(\kappa; \mu, v)\, d\kappa, \quad (27)$$

as the condition from which the expansion coefficients are to be determined. But since $S(\mu, v)$ is considered to be defined over all $\mu \in [-1, +1]$ and $v \in [0, \infty)$, (27) is just the full-range expansion (25), and by setting $\psi(z) = S(\mu, v)/\mu v(z)$, we can adapt (B9) and (B10) directly to yield the solution:

$$f(x, \mu, v, t) = f_a + f_c + f_s$$

$$f_d \equiv \sum_{\kappa_l \in \sigma_p} a_l \frac{f(\kappa_l; \mu, v) e^{-\kappa_l x + i\omega t}}{i\omega + v\Sigma_t(v) - \kappa_l \mu v},$$

$$f_c \equiv \iint_C \frac{f(\kappa; \mu, v) e^{-\kappa x + i\omega t}}{i\omega + v\Sigma_t(v) - \kappa\mu v}\, d\kappa, \quad (28)$$

$$f_s \equiv \frac{S(\mu, v)}{\mu v} e^{-[\Sigma_t(v)/\mu]x} e^{i\omega(t - x/\mu v)}.$$

It should be noted that each of these terms has a definite physical significance. The f_s term represents uncollided particles of speed v and angle μ emitted by

the source. These particles correspond to a plane-wave contribution moving along the x axis at phase velocity $v_{ph} = \mu v$ and hence phase shift $\omega x/v_{ph}$. Since this plane wave is due only to uncollided particles of μ and v, it suffers an attenuation in the x direction of $\exp[-\Sigma_t(v)x/\mu]$.

The individual continuum modes (eigenfunctions) of f_c also correspond to free streaming particles since $\kappa \in C$ implies that the integrand of f_c contains plane-wave terms of the form

$$\exp[-\Sigma_t(v)x/\mu] \exp[i\omega(t - x/\mu v)]$$

for all μ and v. However, the total continuum contribution f_c represents collided particles as well. One can distinguish these continuum mode particles from the discrete mode particles since whenever the former type suffers a collision, it is transferred to a different free-streaming mode. As a consequence, it never remains in the same mode for more than one collision. (By way of contrast, a neutron propagating in a discrete mode will remain in that mode for several collisions.) This intermode coupling is represented by the factor

$$\left[\frac{f(\kappa;\mu,v)}{i\omega + v\Sigma_t(v) - \kappa\mu v}\right]$$

in the integrand. The total contribution of continuum particles consists of the sum of these free-streaming modes, each weighted with the appropriate intermode coupling factor. This total contribution f_c, although composed of individual plane-wave modes, usually is not a plane wave itself, since each of the individual continuum modes in the integral interfere coherently with each other by phase mixing.

The terms in f_d are the discrete or "collective" plane-wave modes. It is evident that these modes must involve particles which have suffered collisions. If one considers the wave as moving at a certain phase velocity v_{ph}, those particles moving with v far from v_{ph} will tend to be more easily scattered from the wave phase, hence causing wave attenuation; in addition, the discrete waves can also be damped directly by the loss of particles due to absorption.

IV. CONCLUSIONS

We have attempted to provide an exact transport theory based upon the linearized Boltzmann equation by which one can analyze various boundary-value problems that arise in describing plane-wave propagation in kinetic theory. Although our particular interest has been in the study of the propagation of such waves in a distribution of neutrons, the application of these ideas to different types of "gases" should be evident.

It was possible to give a rather general analysis of the eigenvalue spectrum of the Boltzmann transport operator for plane-wave propagation. A procedure by which the corresponding spectral representation could be used to solve full-range boundary-value problems was also indicated, and a physical interpretation of this representation was given.

Let it be remarked that there have been recent wave propagation experiments which cannot be modeled by full-range boundary-value problems. Such experiments involve the reflection of the wavelike disturbance from an interface between two dissimilar materials—and hence necessitate modeling by half-range boundary conditions.[19] Unfortunately the theory developed in its most *general* form cannot be extended to these problems, since a proof of half-range completeness for a general scattering kernel does not exist at this time (as evidenced by the plethora of investigations and subsequent approximations concerning this subject[20,21]). The only models of the scattering kernel which presently allow the treatment of half-range problems are the simple degenerate kernels of the form

$$\Sigma_s(v' \to v, \mu' \to \mu) = f(v', \mu')g(v, \mu).$$

Since the theory for such kernels has been developed extensively by Cercignani[15] and others,[19] we shall refer the reader to his work for further information.

However, it should again be stressed that the theory as developed in this paper is sufficient for the analysis of the customary neutron wave propagation experiments (since they are asymptotic experiments and are not particularly sensitive to the type of source boundary condition chosen). Indeed, this theory has been applied[22] in some detail to analyze neutron wave experiments performed in graphite and D_2O moderators. In this latter treatment, the cross sections and scattering kernels were modeled to allow a more detailed examination (and numerical calculation) of the point eigenvalue spectrum, and a more transparent interpretation of the general solution to full-range boundary-value problems was obtained. Agreement with experimental data was quite good and tended to substantiate many of the assumptions made earlier in this analysis.

[19] H. G. Kaper, J. H. Ferziger, and S. K. Loyalka, *IAEA Symposium on Neutron Thermalization and Reactor Spectra, Ann Arbor, 1967* (IAEA, Vienna, 1968), SM-96/60.

[20] I. Kuščer in *Developments in Transport Theory*, E. Inönü and P. F. Zweifel, Eds. (Academic Press Inc., New York, 1967).

[21] A. Leonard and J. H. Ferziger, Nucl. Sci. Eng. **26**, 170, 181 (1966).

[22] J. J. Duderstadt, Nucl. Sci. Eng. **33**, 119 (1968).

ACKNOWLEDGMENTS

The author gratefully acknowledges the guidance given this work by his thesis advisor, Dr. Harold Lurie, and the assistance given through numerous discussions with Dr. Noel Corngold, Dr. Robert Conn, and Dr. Anthony Leonard.

APPENDIX A

We wish to prove the theorem of Sec. II.D. It is necessary to consider several preliminary results. First we demonstrate that for isotropic, noncrystalline media, there exists a frequency ω such that for $\omega > \omega^*$ the point eigenvalue spectrum $\sigma_p(A)$ is empty. We shall do this by demonstrating that for sufficiently large ω, the dispersion law has no nontrivial solutions for any $\kappa \notin C$. It is possible to show that the dispersion law (7) reduces to

$$\psi(\kappa) = \int_0^\infty dv' \Sigma_s(v', v) g(\kappa, \omega; v') \psi_\kappa(v') \equiv K_\kappa \psi_\kappa, \quad \kappa \notin C,$$

where

$$g(\kappa, \omega; v) \equiv 1/(2\kappa) \ln \left[\frac{i\omega + v\Sigma_t(v) + \kappa v}{i\omega + v\Sigma_t(v) - \kappa v} \right]. \quad \text{(A1)}$$

A necessary condition for (A1) to possess a nontrivial solution is for

$$\|K_\kappa\| \equiv \left[\int_0^\infty dv' \int_0^\infty dv \Sigma_s^2(v', v) \, |g(\kappa, \omega; v')|^2 \right]^{\frac{1}{2}} \geq 1, \quad \text{for some } \kappa \notin C.$$

We will show that this condition can never be satisfied for sufficiently large ω. To do this, we merely note that regardless of the behavior of $\kappa(\omega)$, $g(\kappa, \omega; v) \to 0$ as $\omega \to \infty$ for bounded v. Hence for sufficiently well behaved $\Sigma_s(v', v)$ (which we have for noncrystalline media),

$$\lim_{\omega \to \infty} \int_0^{v^*} dv' \int_0^\infty dv \Sigma_s^2(v', v) \, |g(\kappa, \omega; v')|^2 \to 0, \quad v^* < \infty.$$

If we consider the remainder R, we can use the slowing-down form of the scattering kernel

$$\Sigma_s(v', v) = \begin{cases} C_1 v/v'^2, & v \in [(\alpha)^{\frac{1}{2}} v', v'], \\ 0, & \text{otherwise}, \end{cases}$$

to demonstrate that $R \to 0$ as $v^* \to \infty$. Hence we can conclude that

$$\lim_{\omega \to \infty} \|K_\kappa\| = 0.$$

In particular for some $\omega^* < \infty$, $\omega > \omega^*$ will imply $\|K_\kappa\| < 1$ for any $\kappa \notin C$.

It is possible to give a similar proof for the bounds on Σ_a and $B_\perp^2$. Since variations of these results have been given elsewhere for the $\omega = 0$ case,[14,23] a detailed discussion will be omitted.

Now to extend these results to more general kernels, we would like to show that for sufficiently high $\omega(\Sigma_a$ or $B_\perp^2)$,

$$\|G_\kappa\| \equiv \left[\int_{-1}^{+1} d\mu \int_0^\infty dv \int_{-1}^{+1} d\mu' \int_0^\infty dv' \times \frac{\Sigma_s^2(v' \to v, \mu' \to \mu)}{|i\omega/v + \Sigma_t(v) - \kappa\mu| \, |i\omega/v' + \Sigma_t(v') - \kappa\mu'|} \right]^{\frac{1}{2}} < 1, \quad \text{for all } \kappa \notin C,$$

since this corresponds to the more general dispersion law (7). A rather crude argument would be to realize that since the probability to scatter into any arbitrary angle is always greater than the probability to scatter from a specific μ' to a specific μ, one should be able to bound

$$|\Sigma_s(v' \to v, \mu' \to \mu)| < M \, |\Sigma_s(v', v)|.$$

But, then, note that

$$\|G_\kappa\| < M' \left[\int_{-1}^{+1} d\mu \int_0^\infty dv \int_{-1}^{+1} d\mu' \int_0^\infty dv' \times \frac{\Sigma_s^2(v', v)}{|i\omega/v + \Sigma_t(v) - \kappa\mu| \, |i\omega/v' + \Sigma_t(v') - \kappa\mu'|} \right]^{\frac{1}{2}} = M' \, \|K_\kappa\|.$$

Hence using our earlier results that $\lim_{\omega \to \infty} \|K_\kappa\| = 0$, we can conclude that $\|G_\kappa\| \to 0$ as $\omega \to \infty$. Q.E.D.

By using the fact that $\Sigma_e(v)$ is bounded and behaves as c/v^2 for large v, one can show that in the limit of high ω, the norms of the analogous operators for polycrystalline materials approach the expressions for noncrystalline materials and hence conclude that the theorem of Sec. II.D is quite general.

APPENDIX B

The proof of the completeness theorem is constructive in nature and consists of actually evaluating the expansion coefficients in the expansion

$$\psi(z) = \sum_{\kappa_l \in \sigma_p(A)} a_l F(\kappa_l; z) + \iint_C F(\kappa; z) \, dz, \quad \text{(B1)}$$

and then demonstrating that such an evaluation is unique. First define

$$\psi'(z) = \psi(z) - \sum_l a_l F(\kappa_l; z),$$

[23] M. M. R. Williams, *IAEA Symposium on Neutron Thermalization and Reactor Spectra, Ann Arbor, 1967* (IAEA, Vienna, 1968), SM-96/3.

and consider (B1) as

$$\psi'(z) = \iint_C F(\kappa; z)\, d\kappa = \iint_C \frac{f(\kappa; z)}{z - \kappa}\, d\kappa + \lambda(z), \quad \text{(B2)}$$

where the explicit form of the continuum eigenfunctions (21) has been used. Now using equation (22) in (B2), one obtains

$$K(z, u)\psi'(z) = K(z, u)\iint_C \frac{f(\kappa; z)}{z - \kappa}\, d\kappa + f(z; u) - \iint_C \left[\frac{K(\zeta, u)}{\zeta - z}\right] f(z; \zeta)\, d\zeta. \quad \text{(B3)}$$

This is a rather complicated two-dimensional integral equation for $f(z; u)$. Notice how both of the variables z and u are involved in the equation.

To complete the proof of the theorem, we must demonstrate the existence and uniqueness of a solution to (B3) for any $\psi'(z) \in \mathfrak{L}_1(\bar{C})$. Fortunately, techniques exist which will allow the construction of a formal solution to (B3). These techniques rely upon the theory of generalized analytic functions in which one defines an operator

$$T_C[f]_z^\zeta \equiv -\frac{1}{\pi}\iint_G \frac{f(\zeta)}{\zeta - z}\, d\zeta.$$

Now rewrite Eq. (B3) in terms of this operator:

$$K(z, u)\psi'(z) = K(z, u)\pi T_C[f(\kappa; z)]_z^\kappa + f(\kappa; u) + \pi T_C[K(\zeta, u)f(z; \zeta)]_z^\zeta, \quad z \in C.$$

But this can be rearranged as

$$\frac{\partial}{\partial \bar{z}}\left[\{1 + \pi T_C[K(\zeta, u)\cdot]_z^\zeta\}\{T_C[f(\kappa; u)]_z^\kappa\}\right] = K(z, u)\psi'(z), \quad z \in C,$$

$$\frac{\partial}{\partial \bar{z}}\left[\{1 + \pi T_C[K(\zeta, u)\cdot]_z^\zeta\}\{T_C[f(\kappa; u)]_z^\kappa\}\right] = 0, \quad z \notin \bar{C}, \quad \text{(B4)}$$

where we have defined

$$\left\{\iint_C \frac{K(\zeta, u)\cdot d\zeta}{\zeta - z}\right\}\phi(z; u) \equiv \iint_C \frac{K(\zeta, u)\phi(z; \zeta)\, d\zeta}{\zeta - z}.$$

Hence we can apply a lemma due to Vekua[15,16] to find

$$\{1 + \pi T_C[K(\zeta, u)\cdot]_z^\zeta\}\{T_C[f(\kappa; u)]_z^\kappa\} = T_C[K(\zeta, u)\psi'(\zeta)]_z^\zeta. \quad \text{(B5)}$$

Now we must solve (B5) for $f(z; u)$. Clearly this will involve using the inverse operator

$$\{1 + \pi T_C[K(\zeta, u)\cdot]_z^\zeta\}^{-1} \equiv [1 - \mathfrak{L}_{(u,z)}]^{-1},$$

where we have defined

$$\mathfrak{L}_{(u,z)}\varphi(z; \zeta) \equiv \iint_C \left[\frac{K(\zeta, u)}{\zeta - z}\right]\varphi(z; \zeta)\, d\zeta. \quad \text{(B6)}$$

Of course, one must account for those values of z for which this inverse fails to exist. Now recognize from Eq. (20) that these points $z = \kappa_l$ are just the discrete spectrum $\sigma_p(A)$, and the usual application of the Fredholm alternative[9] requires

$$(\varphi^\dagger(z; u), T_C[K(\zeta, u)\psi'(\zeta)]_z^\zeta) = 0, \quad z \in \sigma_p(A), \quad \text{(B7)}$$

where $\varphi^\dagger(z; u)$ is the solution to the homogeneous adjoint problem

$$\varphi^\dagger(z; u) = \iint_C \left[\frac{\overline{K(u, \zeta)}}{\bar{u} - \bar{z}}\right]\varphi^\dagger(z; \zeta)\, d\zeta. \quad \text{(B8)}$$

But notice that we can identify (B8) as just (23), which implies that $\varphi^\dagger(z; u) = F^\dagger(z; u)$. We can now use this fact in equation (B7) to find, after some manipulation,

$$a_l = (F_{\kappa_l}^\dagger, \psi)/(F_{\kappa_l}^\dagger, F_{\kappa_l}), \quad \text{(B9)}$$

which is just the expression for the discrete expansion coefficients one would have obtained by applying biorthogonality directly to (B1). Thus subject to (B9), we can solve (B3) for

$$T_C[f(\kappa; u)]_z^\kappa = [1 - \mathfrak{L}_{(u,z)}]^{-1}T_C[K(\zeta, u)\psi'(\zeta)]_z^\zeta$$

or

$$f(z; u) = \frac{\partial}{\partial \bar{z}}\{[1 - \mathfrak{L}_{(u,z)}]^{-1}T_C[K(\zeta, u)\psi'(\zeta)]_z^\zeta\}. \quad \text{(B10)}$$

Hence by formal construction we have demonstrated the existence of a solution $f(z; u)$ to (B3) for arbitrary $\psi(z) \in \mathfrak{L}_1(\bar{C})$. The converse of the lemma yields the uniqueness of $f(z; u)$. Q.E.D.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Symmetry of the Two-Dimensional Hydrogen Atom

ARTURO CISNEROS* AND HAROLD V. MCINTOSH
Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional
México 14, D.F., México

(Received 11 March 1968)

A paradox has arisen from some recent treatments of accidental degeneracy which claim that, for three degrees of freedom, $SU(3)$ should be a universal symmetry group. Such conclusions are in disagreement both with experimentally observed spectra and with the generally accepted solutions of Schrödinger's equation. The discrepancy occurs in the transition between classical and quantum-mechanical formulations of the problems, and illustrates the care necessary in forming quantum-mechanical operators from classical expressions. The hydrogen atom in parabolic coordinates in two dimensions, for which the traditional treatment of Fock, extended by Alliluev, requires the symmetry group $O(3)$, is a case for which the newer methods of Fradkin, Mukunda, Dulock, and others require $SU(2)$. Although these groups are only slightly different, $SU(2)$ fails to be the "universal" symmetry group on account of the multiple-valuedness of the parabolic representation. This conclusion extends a result of Han and Stehle: that, for rather similar reasons, $SU(2)$ cannot be the classical symmetry group for the two-dimensional hydrogen atom.

The symmetry of the quantum-mechanical hydrogen atom was already given a definitive treatment in Fock's[1] widely quoted paper of 1935, in which it was shown that the Schrödinger equation in momentum space could be transformed by stereographic projection into Laplace's equation for a hypersphere, in the case of the bound states, or a suitable hyperboloid in the case of the free states. Since that time similar methods have been used to explain the "accidental" degeneracy of other systems, among which the one of principal interest has been the harmonic oscillator[2]; this is necessarily so because of considerations which indicate that only the hydrogen atom and the harmonic oscillator, of all non-velocity-dependent central-force problems, are likely to present a degree of degeneracy in excess of that pertinent to their spherical symmetry. This result is sometimes cited as "Bertrand's theorem,"[3] and depends upon the observation that those problems showing accidental degeneracy are those for which all the bounded orbits are closed. With this supposition, one may apply perturbation theory to those circular orbits which exist for spherically symmetric potentials to determine those for which the neighboring orbits are also closed, with the results mentioned. Winternitz[4] has recently investigated the possibilities arising when more general forms of the potentials or of the kinetic energy are admitted.

Among the variant problems, which have attracted some degree of interest, are those of the harmonic oscillator and hydrogen atom where the number of dimensions is different from three. In particular, Alliluev[5] has treated the n-dimensional hydrogen atom, and seen that the same method of stereographic projection from momentum space used by Fock is applicable, and that one expects to find $O(n+1)$, the orthogonal group in $(n+1)$ dimensions, as the symmetry group. There are constants of motion associated with this symmetry group; in addition to the components of the angular-momentum vector which are expected on account of the spherical symmetry, there is another vector constant of motion which is generally known as the Runge, or Runge–Lenz, vector. Its components, together with those of the angular-momentum vector, generate the Lie algebra of the orthogonal group $O(n+1)$. Loudon[6] has even obtained the rather curious result that this description is valid when $n=1$, so that the levels of the 1-dimensional hydrogen atom may be doubly degenerate according to the symmetry group $O(2)$.

There has recently been a sharp revival of interest in the whole realm of accidental degeneracy, due to the hope for its possible applicability to the theory of elementary particles. There have also arisen some new principles according to which a symmetry group might be constructed, over and beyond the method of stereographic projection, but of the very disturbing nature that they tend to indicate that both $SU(3)$ and $O(4)$ should be universal symmetry groups, especially for three-dimensional central-force potentials.

* Based in part upon a Professional Thesis submitted to the Escuela Superior de Física y Matemáticas of the Instituto Politécnico Nacional in partial fulfilment of the requirements for the title "Licenciado en Física y Matemáticas."

[1] V. Fock, Z. Phys. **98**, 145–154 (1935).

[2] J. M. Jauch, Ph.D. thesis, Univ. of Minnesota, 1939; J. M. Jauch, Phys. Rev. **55A**, 1132 (1939); V. A. Dulock and H. V. McIntosh, Am. J. Phys. **33**, 109 (1965).

[3] J. Bertrand, Compt. Rend. **77**, 849 (1873).

[4] P. Winternitz, Ya. A. Smorodinskii, M. Ulhir, and I. Fris, Sov. J. Nucl. Phys. **4**, 444 (1967).

[5] S. P. Alliluev, Zh. Eksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys.—JETP **6**, 156 (1958)].

[6] R. Loudon, Am. J. Phys. **27**, 649 (1959).

These expectations are at variance with known facts, both the experimental ones regarding the actual spectra, and the theoretical ones arising from the degeneracies found in the quite familiar solutions of the relevant Schrödinger equations. Nevertheless there are unequivocal demonstrations that these are universal symmetry groups for classical mechanics. It is therefore apparent that considerable caution must be exercised in extending classical results to quantum mechanical problems, particularly in expecting Poisson-bracket relations to remain valid for commutator brackets. Clarification of the transition is of not inconsiderable interest, if only because one would like to know to what extent he may rely on classical results in trying to understand a quantum-mechanical situation. Particularly in the search for symmetries, constants of motion, and accidental degeneracies, such generalization has been a constant source of inspiration. Our recent investigations have shown that the discrepancy, which allows the intrusion of universal symmetry groups, arises in the failure of certain commutation relations when operators belonging to supposedly independent coordinates fail to commute when applied to wavefunctions bearing extreme quantum numbers. The details are properly the subject of another discussion, but our present motivation is that there is one interesting case, much cited and used as a basis for discussion in the literature, in which this failure does not occur. Moreover, the supposed universal group fails, both classically and quantum mechanically. The problem is, therefore, of some illustrative value in pointing up some of the possible pitfalls. At first sight one expects to see both a unitary group and an orthogonal group as the symmetry group responsible for accidental degeneracy. This is the two-dimensional case of the hydrogen atom in which the concepts and constructions leading to a universal symmetry group lead both to $SU(2)$ and $O(3)$ as admissible symmetry groups. Admittedly the difference between these two groups is not very great, as one is the covering group of the other.

Han and Stehle[7] have discussed the problem from a classical point of view and a somewhat negative aspect—they have shown that only the harmonic-oscillator potential, among all two-dimensional problems of a reasonable class which includes the hydrogen atom, may have the full $SU(2)$ as its symmetry group. By default, then, the two-dimensional hydrogen atom must have $O(3)$ and not $SU(2)$ symmetry if it has the Lie algebra leading to either. Since the constants of motion do in fact form this Lie algebra, the symmetry $O(3)$ is indicated. Failure of the hydrogen-atom orbits to be symmetric with respect to inversion in the origin, a characteristic of the harmonic-oscillator orbits, is the reason attributed for this failure to achieve the full symmetry.

Fradkin,[8] Mukunda,[9] Dulock,[10] and others have exhibited explicit classical generators for the supposed universal symmetry groups. Dulock has, moreover, given the generators of his $SU(2)$ group for the two-dimensional hydrogen atom in parabolic coordinates, where the operators were entirely analogous to those producing the $SU(2)$ symmetry of the two-dimensional harmonic oscillator. Re-examination of this solution shows that the mapping from parabolic coordinates to Cartesian coordinates is double-valued, and that consequently only those solutions in parabolic coordinates which are unaffected by this double-valuedness are admissible solutions of Schrödinger's equation. With this resolution, both arguments, classical and quantum mechanical, are in agreement that $O(3)$ and not $SU(2)$ is the symmetry group of this system.

The solution in parabolic coordinates is not without its intrinsic interest, however, since the procedures of these recent authors do in fact result in a degeneracy-producing calculus of ladder operators. Its only flaw, if such the error in its possible interpretation is to be called, is in its not generating a group the same as the classical group would have been. In the two-dimensional hydrogen atom this discrepancy takes the innocuous form of yielding the correct Lie algebra but with a difference in the Lie groups which is finally lost because of the two-valuedness of the parabolic transformation. This is seen clearly from examining the actual wavefunction.

There have been other solutions for the two-dimensional hydrogen atom; for example, Shibuya[11] has worked out the solution in momentum space and even constructed a mechanical model of the wavefunctions. Jauch[2] likewise considered the two-dimensional atom in his thesis, and very recently Zaslow and Zandler[12] have obtained a solution in polar coordinates. However, there does not seem to have been any particular interest in treating the problem in parabolic coordinates in spite of their yielding a useful separation of the equations of motion in three dimensions. Moreover, the parabolic coordinates which are most convenient in three dimensions are different from those which we use. Ours are squares

[7] M. Y. Han and P. Stehle, Nuovo Cimento **48**, 180 (1967).

[8] D. M. Fradkin, Progr. Theoret. Phys. (Kyoto) **37**, 798 (1967).

[9] N. Mukunda, Phys. Rev. **155**, 1383 (1967); J. Math. Phys. **8**, 1069 (1967).

[10] V. A. Dulock, Ph.D. thesis, University of Florida, 1964; V. A. Dulock and H. V. McIntosh, Pacific J. Math. **19**, 39 (1966).

[11] T. I. Shibuya and C. F. Wulfman, Am. J. Phys. **33**, 570 (1965).

[12] B. Zaslow and M. F. Zandler, Am. J. Phys. **35**, 1118 (1967).

of the former, making the two-valuedness more apparent. The lack of interest in parabolic coordinates is undoubtedly due to the superiority of polar coordinates when one is interested in describing Fock's stereographic projection.

We are interested in a different approach from that of Fock to the problem of constructing constants of the motion and symmetry groups. We depend upon noticing that whenever the Hamilton–Jacobi equation can be solved by the separation of variables and the introduction of action-angle variables, and the Hamiltonian written as a linear combination of the action variables with equal coefficients, the motion is degenerate with three equal frequencies. Furthermore, a set of functions can be written down which satisfy the Poisson-bracket relations for a set of canonical coordinates and momenta. Of course this is already true for the action-angle variables themselves. The particular utility of the collection which we introduce,

$$a_k^{\pm} = \mp i(J_k/2\pi)^{\frac{1}{2}} e^{\pm 2\pi i w_k}, \tag{1}$$

is that they are additionally eigenfunctions of the Hamiltonian, taken with respect to the Poisson-bracket operation. This relation depends on the Hamiltonian being a function only of the sum of the action variables and therefore degenerate.

Writing

$$H = H\left(\sum_j J_j\right), \tag{2}$$

we find

$$[H, a_k^{\pm}] = \left(\mp 2\pi i H'\left(\sum_j J_j\right)\right) a_k^{\pm}. \tag{3}$$

Since these "eigenfunctions" belong to sets of negative pairs of eigenvalues, and the eigenvalues themselves are triply degenerate, the nine possible products of eigenfunctions belonging to negative pairs of eigenvalues are constants of the motion. The procedure was already described by Dulock and McIntosh,[10] and produces the following constants:

$$\begin{array}{lll} a_1^+a_1^-, & a_1^+a_2^-, & a_1^+a_3^-, \\ a_2^+a_1^-, & a_2^+a_2^-, & a_2^+a_3^-, \\ a_3^+a_1^-, & a_3^+a_2^-, & a_3^+a_3^-. \end{array} \tag{4}$$

Consequently, these constants of motion can be identified as the generators of the group $SU(3)$.[10] A similar line of reasoning applies for an arbitrary number of degrees of freedom, and substantiates the claim that $SU(n)$ should be a symmetry group for any problem which is classically n-fold degenerate. It is another matter to extend this claim to any spherically symmetric potential whatsoever, regardless of whether it is classically degenerate or not.[8,9] However, the foregoing analysis suffices for the case at hand and, in particular, leads us to the expectation of $SU(2)$ as the symmetry group of the two-dimensional hydrogen atom, both in polar and parabolic coordinates. These are two of the coordinate systems in which its Hamilton–Jacobi equation is separable. For the moment it is the parabolic coordinates which interest us.

Letting x and y be the Cartesian coordinates of the problem, we introduce parabolic coordinates, defined through the equations

$$\begin{aligned} x &= \mu\nu, \\ y &= \tfrac{1}{2}(\mu^2 - \nu^2). \end{aligned} \tag{5}$$

In such a coordinate system, the Hamiltonian becomes

$$H = \frac{p_\mu^2 + p_\nu^2 - 4}{2(\mu^2 + \nu^2)}, \tag{6}$$

in which m, the mass of the electron, and Z_e, its electric charge, have both been placed equal to 1. The resulting Hamilton–Jacobi equation is

$$2(\mu^2 + \nu^2)E = \left(\frac{\partial S}{\partial \mu}\right)^2 + \left(\frac{\partial S}{\partial \nu}\right)^2 - 4. \tag{7}$$

The constant E represents the total energy. Writing S in the form $S = S_\mu + S_\nu$, and introducing two separation constants α_1 and α_2, Eq. (7) separates into the two equations

$$\left(\frac{dS_\mu}{d\mu}\right)^2 - 2E\mu^2 - 2\alpha_1 = 0, \tag{8a}$$

$$\left(\frac{dS_\nu}{d\nu}\right)^2 - 2E\nu^2 - 2\alpha_2 = 0, \quad \text{with} \quad \alpha_1 + \alpha_2 = 2. \tag{8b}$$

Since $p_\mu = \partial S/\partial\mu$, $p_\nu = \partial S/\partial\nu$, we find expressions for the momenta

$$p_\mu = (2E\mu^2 + 2\alpha_1)^{\frac{1}{2}}, \tag{9a}$$

$$p_\nu = (2E\nu^2 + 2\alpha_2)^{\frac{1}{2}}. \tag{9b}$$

The action variables, which are defined by the integrals $J_i = \oint p_i\, dq_i$ become

$$J_\mu = \oint (2E\mu^2 + 2\alpha_1)^{\frac{1}{2}}\, d\mu, \tag{10a}$$

$$J_\nu = \oint (2E\nu^2 + 2\alpha_2)^{\frac{1}{2}}\, d\nu. \tag{10b}$$

These can be evaluated by a contour integration,[13] yielding

$$J_\mu = 2\pi\alpha_1(-2E)^{-\frac{1}{2}}, \tag{11a}$$

$$J_\nu = 2\pi\alpha_2(-2E)^{-\frac{1}{2}}. \tag{11b}$$

[12] M. Born, *The Mechanics of the Atom* (Frederick Ungar Publ. Co. Inc., New York, 1959).

The above equations may be solved for the energy, yielding the result

$$E = -8\pi^2(J_\mu + J_\nu)^{-2}, \tag{12}$$

from which the degeneracy of the frequencies of the angle variables is evident. Consequently, Dulock's construction is applicable, which requires that we now compute the angle variables. They are derivatives of the principal function with respect to the action variables, so that for w_μ we obtain

$$w_\mu = \frac{\partial S}{\partial J_\mu} = \frac{\pi}{J^2}\int \frac{4\mu^2 + J_\nu J/2\pi^2}{(JJ_\mu/4\pi^2 - \mu^2)^{\frac{1}{2}}}\,d\mu + \frac{\pi}{J^2}\int \frac{4\nu^2 - J_\nu J/2\pi^2}{(JJ_\nu/4\pi^2 - \nu^2)^{\frac{1}{2}}}\,d\nu, \tag{13}$$

with the substitution $J = J_\mu + J_\nu$. Routine evaluation of the integrals and substitution of the momenta according to Eqs. (9) results in

$$2\pi w_\mu = \arcsin\frac{2\pi\mu}{(JJ_\mu)^{\frac{1}{2}}} - \frac{\pi}{J}(\mu p_\mu + \nu p_\nu). \tag{14}$$

Dulock's operator in the μ coordinate is then

$$a_\mu^{\pm} = [2(-2E)^{-\frac{1}{2}}]^{\frac{1}{2}} \exp\,[\mp i\tfrac{1}{4}(-2E)^{\frac{1}{2}}(\mu p_\mu + \nu p_\nu)] \times [(-2E)^{\frac{1}{2}}\mu \mp i p_\mu]. \tag{15a}$$

Similarly for the ν coordinate

$$a_\nu^{\pm} = [2(-2E)^{-\frac{1}{2}}]^{\frac{1}{2}} \exp\,[\mp i\tfrac{1}{4}(-2E)^{\frac{1}{2}}(\mu p_\mu + \nu p_\nu)] \times [(-2E)^{\frac{1}{2}}\nu \mp i p_\nu). \tag{15b}$$

In order to discuss the properties of these ladder operators it will be convenient to have the wavefunctions for the problem explicitly available. The Schrödinger equation for the two-dimensional hydrogen atom, which has the general form

$$\mathcal{H}f = (-\tfrac{1}{2}\nabla^2 - r^{-1})f = Ef, \tag{16}$$

takes the particular form

$$\frac{1}{2(\mu^2+\nu^2)}\left(\frac{\partial^2 f}{\partial\mu^2} + \frac{\partial^2 f}{\partial\nu^2}\right) + \frac{2f}{\mu^2+\nu^2} + Ef = 0 \tag{17}$$

in parabolic coordinates defined according to Eqs. (5). The assumption that f has the form $f(\mu, \nu) = f_1(\mu)f_2(\nu)$ permits separation into two ordinary differential equations

$$\frac{d^2 f_1}{d\mu^2} + (2E\mu^2 + 2k_1)f_1 = 0, \tag{18}$$

$$\frac{d^2 f_2}{d\nu^2} + (2E\nu^2 + 2k_2)f_2 = 0. \tag{19}$$

Analogous to the procedure for the classical Hamilton–Jacobi equation, we introduce two separation constants k_1 and k_2 such that $k_1 + k_2 = 2$.

These separated equations are quite similar to the differential equation for the one-dimensional harmonic oscillator whose quantum-mechanically acceptable solutions are known to be expressible in terms of Hermite polynomials multiplying a Gaussian exponential.[14] However, the energy constant E appears as a factor in precisely the place where the classical frequency of the oscillator usually appears, and the separation constants appear in the position usually occupied by the energy eigenvalue.

The requirement of square-integrability of the wavefunctions leads to the following restrictions on the values of the separation constants:

$$2k_1(-2E)^{-\frac{1}{2}} = 2n_1 + 1, \tag{20a}$$

$$2k_2(-2E)^{-\frac{1}{2}} = 2n_2 + 1, \quad n_1, n_2 = 0, 1, 2, \cdots. \tag{20b}$$

As a result, the energy levels are forced to present the form

$$E = -\tfrac{1}{2}[(n_1 + n_2 + 1)/2]^{-2} \tag{21}$$

when the separation constants are eliminated from Eqs. (20). Inasmuch as the energy depends only on the sum of two nonnegative integers, and not on their individual values, there is a degeneracy. This degeneracy is known to correspond to irreducible representations of the group $SU(2)$, which is the symmetry group of the two-dimensional harmonic oscillator.

Explicitly, the functions f_1 and f_2 have the form

$$f_1 = c_1 e^{-\frac{1}{2}u^2} H_{n_1}(u), \tag{22a}$$

$$f_2 = c_2 e^{-\frac{1}{2}v^2} H_{n_2}(v). \tag{22b}$$

The product of the two constants c_1 and c_2 has to be determined by the normalization of the entire wavefunction. The argument variables are defined by

$$u = \left(\frac{2}{n_1 + n_2 + 1}\right)^{\frac{1}{2}}\mu, \tag{23a}$$

$$v = \left(\frac{2}{n_1 + n_2 + 1}\right)^{\frac{1}{2}}\nu. \tag{23b}$$

A word about the origin of these multiplying factors is in order. In the Schrödinger equation for the harmonic oscillator, the frequency is the multiplier of r^2, while the quantized energy is the constant term in a differential equation similar to Eq. (18). At present it is E, which multiplies r^2, and the separation constant which is quantized. Therefore, the energy constant E enters into the scaling factor for the wavefunction

[14] H. Margenau and G. M. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand, Inc., Princeton, N.J., 1956).

argument, in the place where the classical frequency ordinarily appears. The result is that the energy eigenvalues enter into the argument of the wavefunction, as well as into the indices of the Hermite polynomials which comprise part of the wavefunctions.

The final formula for the wavefunctions, so adjusted in the space of parabolic coordinates as to be properly normalized in the original Cartesian space, is

$$\Psi_{n_1,n_2} = [2^{n_1+n_2-3}\pi(n_1+n_2+1)^3 n_1!\, n_2!]^{-\frac{1}{2}} \times e^{-\frac{1}{2}u^2}H_{n_1}(u)e^{-\frac{1}{2}v^2}H_{n_2}(v). \quad (24)$$

Now that we have determined the wavefunctions, we must confront the problem of transcribing the classical ladder operators into an appropriate quantum-mechanical form. It is not directly obvious how to do this, since there occurs an exponential of non-commuting operators, along with other algebraic operators whose quantum-mechanical form presents no special problem. If we begin with the transcription of these latter terms we obtain

$$[2/(n_1+n_2+1)]\mu \mp \partial/\partial\mu \quad (25)$$

by the replacements $p_\mu \to (1/i)\partial/\partial\mu$, $\mu \to \mu$ for the coordinate and momentum operators, and the eigenvalues $-\frac{1}{2}[(n_1+n_2+1)/2]^2$ for the energy operator E. This accounts for the term $[(-2E)^{\frac{1}{2}}\mu \mp p_\mu]$, and in fact these operators are already adequate to factorize the Schrödinger equation in the sense of Infeld and Hull, as was noticed by Dulock.[10]

However, at this stage we have ladder operators only for the functional form of the wavefunctions

$$\{[2/(n_1+n_2+1)]\mu \mp \partial/\partial\mu\}\Psi_{n_1,n_2}(u,v) \sim \Psi_{n_1\pm1,n_2}(u,v). \quad (26)$$

Hence, to get a complete ladder operator, an extra operator is needed to change the scale factor appearing in the argument from

$$\left(\frac{2}{n_1+n_2+1}\right)^{\frac{1}{2}} \text{ to } \left(\frac{2}{n_1\pm1+n_2+1}\right)^{\frac{1}{2}}.$$

A scaling operator is required, and to this end we may recall the operator identity

$$e^{ax\,d/dx}f(x) = f(e^a x). \quad (27)$$

In the present instance it appears that what is required are the operators

$$\exp\left[\ln\left(\frac{n_1+n_2+1}{n_1\pm1+n_2+1}\right)^{\frac{1}{2}}\right](\mu p_\mu + \nu p_\nu). \quad (28)$$

Therefore, if we construct the operator

$$\exp\left[\mp i\tfrac{1}{4}(-2E)^{\frac{1}{2}}(\mu p_\mu + \nu p_\nu)\right], \quad (29)$$

we may verify that in the limit of large n_1 and n_2 we obtain the proper limiting value, since

$$\ln\left(\frac{n_1+n_2+1}{n_1+n_2+2}\right)^{\frac{1}{2}} = -\frac{1}{2}\ln\left(1+\frac{1}{n_1+n_2+1}\right) \xrightarrow[n_1,n_2\to\infty]{} -\frac{1}{2}\frac{1}{n_1+n_2+1} = -\frac{(-2E)^{\frac{1}{2}}}{4},$$

$$\ln\left(\frac{n_1+n_2+1}{n_1+n_2}\right)^{\frac{1}{2}} = \frac{1}{2}\ln\left(1+\frac{1}{n_1+n_2}\right) \xrightarrow[n_1,n_2\to\infty]{} \frac{1}{2}\frac{1}{n_1+n_2+1} = \frac{(-2E)^{\frac{1}{2}}}{4}.$$

Since the operator (29) is exactly the classical expression that we have already obtained, it would seem that we have found the appropriate ladder operators for the entire wavefunction. Normalized, they are

$$\mathcal{N}_\mu^+ = \left(\frac{n_1+n_2+1}{4(n_1+1)}\right)^{\frac{1}{2}}\left(\frac{n_1+n_2+1}{n_1+n_2+2}\right)^{\frac{3}{2}} \times \exp\left\{\left[\ln\left(\frac{n_1+n_2+1}{n_1+n_2+2}\right)^{\frac{1}{2}}\right]\left(\mu\frac{\partial}{\partial\mu}+\nu\frac{\partial}{\partial\nu}\right)\right\} \times \left(\frac{2}{n_1+n_2+1}\mu - \frac{\partial}{\partial\mu}\right), \quad (30a)$$

$$\mathcal{N}_\mu^- = \left(\frac{n_1+n_2+1}{4n_1}\right)^{\frac{1}{2}}\left(\frac{n_1+n_2+1}{n_1+n_2}\right)^{\frac{3}{2}} \times \exp\left\{\left[\ln\left(\frac{n_1+n_2+1}{n_1+n_2}\right)^{\frac{1}{2}}\right]\left(\mu\frac{\partial}{\partial\mu}+\nu\frac{\partial}{\partial\nu}\right)\right\} \times \left(\frac{2}{n_1+n_2+1}\mu + \frac{\partial}{\partial\mu}\right), \quad (30b)$$

$$\mathcal{N}_\nu^+ = \left(\frac{n_1+n_2+1}{4(n_2+1)}\right)^{\frac{1}{2}}\left(\frac{n_1+n_2+1}{n_1+n_2+2}\right)^{\frac{3}{2}} \times \exp\left\{\left[\ln\left(\frac{n_1+n_2+1}{n_1+n_2+2}\right)^{\frac{1}{2}}\right]\left(\mu\frac{\partial}{\partial\mu}+\nu\frac{\partial}{\partial\nu}\right)\right\} \times \left(\frac{2}{n_1+n_2+1}\nu - \frac{\partial}{\partial\nu}\right), \quad (30c)$$

$$\mathcal{N}_\nu^- = \left(\frac{n_1+n_2+1}{4n_2}\right)^{\frac{1}{2}}\left(\frac{n_1+n_2+1}{n_1+n_2}\right)^{\frac{3}{2}} \times \exp\left\{\left[\ln\left(\frac{n_1+n_2+1}{n_1+n_2}\right)^{\frac{1}{2}}\right]\left(\mu\frac{\partial}{\partial\mu}+\nu\frac{\partial}{\partial\nu}\right)\right\} \times \left(\frac{2}{n_1+n_2+1}\nu + \frac{\partial}{\partial\nu}\right). \quad (30d)$$

Applied to the normalized wavefunctions $\Psi_{n_1,n_2}(\mu, \nu)$, they satisfy the identities

$$\mathcal{N}_\mu^{\pm}\Psi_{n_1,n_2}(\mu, \nu) = \Psi_{n_1\pm1,n_2}(\mu, \nu), \tag{31a}$$

$$\mathcal{N}_\nu^{\pm}\Psi_{n_1,n_2}(\mu, \nu) = \Psi_{n_1,n_2\pm1}(\mu, \nu). \tag{31b}$$

Checking the passage to the classical limit, we find

$$\mathcal{N}_\mu^{\pm} \to \mp ie^{\pm 2\pi i w_\mu},$$

$$\mathcal{A}_\mu^{+} \equiv (n_1 + 1)^{\frac{1}{2}}\mathcal{N}_\mu^{+} \to -i(J_\mu/2\pi)^{\frac{1}{2}}e^{2\pi i w_\mu},$$

$$\mathcal{A}_\mu^{-} \equiv (n_1)^{\frac{1}{2}}\mathcal{N}_\mu^{-} \to i(J_\mu/2\pi)^{\frac{1}{2}}e^{-2\pi i w_\mu}.$$

Finally, it is important to verify the commutation relations satisfied among the operators $\mathcal{A}$. The nonzero commutators are

$$[\mathcal{A}_\mu^{-}, \mathcal{A}_\mu^{+}] = [\mathcal{A}_\nu^{-}, \mathcal{A}_\nu^{+}] = 1. \tag{32}$$

Since these are the commutation rules of harmonic-oscillator ladder operators, we know that an $SU(2)$ Lie algebra may be constructed from bilinear products of the $\mathcal{A}$'s. The importance of verifying these commutation rules arises from the fact that in other instances of the application of Dulock's procedure, the corresponding commutators are modified slightly. Generally this is no more than to the extent that when a product is applied in one order to a wavefunction with extreme quantum numbers, the product is zero, but in the other order it is a multiple of the transformed wavefunction. Such a result can come about when one operator raises an eigenvalue beyond its limit, but the other lowers it. If the lowering occurs first, the wavefunction survives, but if it occurs later the wavefunction has already been annihilated.

Having obtained the ladder operators, we may proceed to the formation of constants of the motion from their bilinear functions. The procedure was discussed in detail by Dulock,[10] and earlier by McIntosh,[15] but essentially depends upon the fact that the product of a raising and a lowering operator either annihilates a wavefunction, or produces another of the same energy, in those cases where the energy is a sum of two quantum numbers, and the increase in one is just equal to the decrease in the other.

Since the Hamiltonian is a function of one bilinear combination $a_\mu^{+}a_\mu^{-} + a_\nu^{+}a_\nu^{-}$, there are three other independent combinations. It is convenient to select them to be real, and the most frequent choice is the following:

$$K = [a_\mu^{+}a_\nu^{-} + a_\mu^{-}a_\nu^{+}], \tag{33a}$$

$$L = i[a_\mu^{+}a_\nu^{-} - a_\mu^{-}a_\nu^{+}], \tag{33b}$$

$$D = [a_\mu^{+}a_\mu^{-} - a_\nu^{+}a_\nu^{-}]. \tag{33c}$$

[15] H. V. McIntosh, Am. J. Phys. 27, 620 (1959).

When they are written for the present problem in terms of the parabolic coordinates and momenta, we obtain

$$K = (p_\mu p_\nu - 2E\mu\nu)/(-2E)^{\frac{1}{2}}, \tag{34a}$$

$$L = \tfrac{1}{2}(\nu p_\mu - \mu p_\nu), \tag{34b}$$

$$D = (p_\mu^2 - p_\nu^2 - 2E(\mu^2 - \nu^2))/2(-2E)^{\frac{1}{2}}. \tag{34c}$$

They could also be written in terms of the Cartesian coordinates and momenta, whereupon one would discover that K and D are the x and y components of the vector $2\mathbf{R}/(-2E)^{\frac{1}{2}}$ where $\mathbf{R}$ is the Runge vector, and that L was the angular momentum. Dulock[10] reports a similar result, but in his derivation he used ladder operators which did not contain the scaling operator which we have incorporated. It cancels out from the calculation, however.

The quantum-mechanical operators are directly analogous to the classical operators. One would define, by analogy to the classical construction,

$$\mathcal{K} = [\mathcal{A}_\mu^{+}\mathcal{A}_\nu^{-} + \mathcal{A}_\mu^{-}\mathcal{A}_\nu^{+}], \tag{35a}$$

$$\mathcal{L} = i[\mathcal{A}_\mu^{+}\mathcal{A}_\nu^{-} - \mathcal{A}_\mu^{-}\mathcal{A}_\nu^{+}], \tag{35b}$$

$$\mathcal{D} = [\mathcal{A}_\mu^{+}\mathcal{A}_\mu^{-} - \mathcal{A}_\nu^{+}\mathcal{A}_\nu^{-}]. \tag{35c}$$

These are all Hermitian operators and commute with the Hamiltonian. Consequently their eigenvalues are constants of the motion. To write them all as differential operators it is not convenient to use the definitions of the ladder operators directly, because of complications arising from the presence of the scaling operator. Rather, we apply them to the wavefunctions and then note which differential operators applied to the wavefunctions would produce the same results. For example, applying $\mathcal{K}$ we find

$$\begin{aligned}\frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{K}F &= \frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{K}\sum_{n_1,n_2} b_{n_1,n_2}\Psi_{n_1,n_2} \\ &= \sum_{n_1n_2} b_{n_1,n_2}\left(\frac{1}{n_1 + n_2 + 1}\right)[(n_2(n_1 + 1))^{\frac{1}{2}}\Psi_{n_1+1,n_2-1} \\ &\quad + [(n_2 + 1)n_1]^{\frac{1}{2}}\Psi_{n_1-1,n_2+1}],\end{aligned} \tag{36}$$

where F is any function expandible in terms of the wavefunctions. But, we see that

$$\frac{(-2E)^{\frac{1}{2}}}{4}\left[\left(v + \frac{\partial}{\partial v}\right)\left(u - \frac{\partial}{\partial u}\right) + \left(u + \frac{\partial}{\partial u}\right)\left(v - \frac{\partial}{\partial v}\right)\right] = \left(-\mu\nu\mathcal{K} - \frac{1}{2}\frac{\partial}{\partial\mu}\frac{\partial}{\partial\nu}\right) \tag{37}$$

produces the same result; we conclude that

$$\frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{K} = -\mu\nu\mathcal{N} - \frac{1}{2}\frac{\partial}{\partial\mu}\frac{\partial}{\partial\nu}. \tag{38}$$

Similar considerations apply to the other operators, and in summary we find

$$\frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{K} = \mu\nu\left(\frac{1}{2}\nabla^2 + \frac{2}{\mu^2+\nu^2}\right) - \frac{1}{2}\frac{\partial}{\partial\mu}\frac{\partial}{\partial\nu}, \tag{39a}$$

$$\mathcal{L} = \frac{1}{2i}\left(\nu\frac{\partial}{\partial\mu} - \mu\frac{\partial}{\partial\nu}\right), \tag{39b}$$

$$\frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{D} = \frac{\mu^2-\nu^2}{2}\left(\frac{1}{2}\nabla^2 + \frac{2}{\mu^2+\nu^2}\right) - \frac{1}{4}\left(\frac{\partial^2}{\partial\mu^2} - \frac{\partial^2}{\partial\nu^2}\right). \tag{39c}$$

Transforming these operators to Cartesian coordinates, one obtains

$$\frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{K} = \frac{x}{(x^2+y^2)^{\frac{1}{2}}} + x\frac{\partial^2}{\partial y^2} - y\frac{\partial}{\partial x}\frac{\partial}{\partial y} - \frac{1}{2}\frac{\partial}{\partial x}, \tag{40a}$$

$$\frac{(-2E)^{\frac{1}{2}}}{2}\mathcal{D} = \frac{y}{(x^2+y^2)^{\frac{1}{2}}} + y\frac{\partial^2}{\partial x^2} - x\frac{\partial}{\partial y}\frac{\partial}{\partial x} - \frac{1}{2}\frac{\partial}{\partial y}, \tag{40b}$$

$$\mathcal{L} = \frac{1}{i}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right). \tag{40c}$$

However, these operators coincide with the quantum-mechanical operator versions of the components of the Runge vector and angular momentum,[2] so we see that we have successfully obtained the operators generating Fock's symmetry group for the two-dimensional hydrogen atom.

The principal point of interest of our presentation, however, is to examine the behavior of the wavefunctions themselves when we return to Cartesian coordinates.

Examining the wavefunctions as expressed by Eq. (24), we see that the Gaussian term is insensitive to any replacement of μ or ν by its negative, while the Hermite polynomials themselves are of either even or odd parity depending upon their degree. The coordinate functions themselves are such that x maps into $-x$ when either μ or ν is replaced by its negative, but not both simultaneously. However, a reflection in the 45° line, which interchanges μ and ν, maps y into $-y$. An inversion in the origin of parabolic coordinates, however, leaves both x and y unchanged, and on this account both (μ, ν) and $(-\mu, -\nu)$ represent the same point (x, y). The effect of all the symmetries of the square applied to the parabolic coordinates and the corresponding effect on Cartesian coordinates is summarized in Table I.

TABLE I. Rotational and reflective symmetries in parabolic coordinates, and their corresponding effect on the Cartesian coordinates.

Operation	Effect on parabolic coordinates		Effect on cartesian coordinates	
Identity	$\mu \to \mu$	$\nu \to \nu$	$x \to x$	$y \to y$
+90° rotation	$\mu \to -\nu$	$\nu \to \mu$	$x \to -x$	$y \to -y$
+180° rotation	$\mu \to -\mu$	$\nu \to -\nu$	$x \to x$	$y \to y$
−90° rotation	$\mu \to \nu$	$\nu \to -\mu$	$x \to -x$	$y \to -y$
μ axis reflection	$\mu \to \mu$	$\nu \to -\nu$	$x \to -x$	$y \to y$
ν axis reflection	$\mu \to -\mu$	$\nu \to \nu$	$x \to -x$	$y \to y$
+45° line reflection	$\mu \to \nu$	$\nu \to \mu$	$x \to x$	$y \to -y$
−45° line reflection	$\mu \to -\nu$	$\nu \to -\mu$	$x \to x$	$y \to -y$

$\Psi(-\mu, -\nu) = \Psi(\mu, \nu)$ is a condition which has to be imposed on an acceptable wavefunction, if it is to be single-valued. This requirement is met for quantum numbers n_1 and n_2 such that the sum $(n_1 + n_2)$ is even. We must abandon one-half of the parabolically acceptable wavefunctions, for which the only criterion was square-integrability. The various constants of motion are unaffected by this restriction, since they are composed of products of one raising and one lowering operator. Such composites always leave the parity of the sum of the quantum numbers unaltered.

One might be tempted to divide the μ–ν plane in half, and make use of the resulting single-valued wavefunctions in Cartesian coordinates. However, one must join points near one edge of the half-plane to their mirror images in the origin if he is not to have a discontinuity in the Cartesian space, with the further condition that the partial derivatives must also be continuous across the cut. For definiteness and illustration purposes we choose the $\mu \geq 0$ half-plane for the analysis. The same results would be obtained regardless of the particular cut chosen. Figure 1(a) shows the mapping back to Cartesian coordinates of a wavefunction even both in μ and ν; the values of the function match across the cut as well as the partial derivatives. Similarly, an acceptable matching is obtained for the mapping of a wavefunction odd both in μ and ν [Fig. 1(b).] However, if we try to map a wavefunction even in μ and odd in ν, the values of the function change sign for mirror-image points along the cut in the μ–ν plane, since the cut line is mapped onto a half-line in the Cartesian plane; the function we get is discontinuous along such a cut [Fig. 1(c)]. For a wavefunction odd in μ and even in ν we do get a match in the values of the function along the cut;

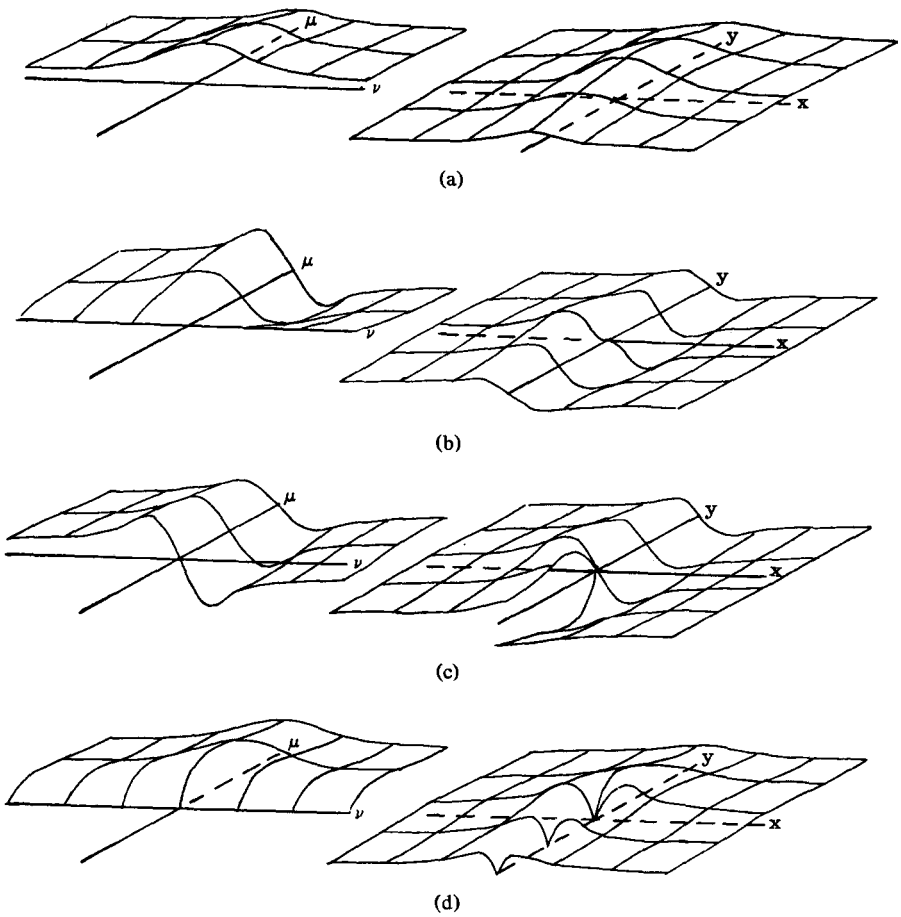


FIG. 1. Symmetry of the wave functions in parabolic coordinates. The half-plane shown on the left maps into the Cartesian space shown on the right with the cut line along the negative y axis. (a) Even in μ and ν. (b) Odd in u and ν. (c) Even in μ and odd in ν. (d) Odd in μ and even in ν.

nevertheless, the partial derivatives are discontinuous at the cut [Fig. 1(d)].

The result of this analysis then is to show that among the admissible solutions of the Schrödinger equation for the two-dimensional hydrogen atom in Cartesian coordinates there never occur any wavefunctions which transform according to representations of $SU(2)$ which are not simultaneously representations of $O(3)$, when the various constants of motion are applied to them. Consequently, any thought that $SU(2)$ and not $O(3)$ is the symmetry group is purely academic. If we choose to regard $SU(2)$ as the symmetry group, we will have to concede that all those representations which would distinguish it from $O(3)$ are missing. The situation would have been different if all the square-integrable functions acceptable as solutions of the Schrödinger equation in parabolic coordinates had also been admissible in terms of Cartesian coordinates, for then certain irreducible representations would have been present which simply were not representations of $O(3)$. On the other hand, we probably should not push the identification of the symmetry group further since we have operators which are operating on the wavefunctions and not on the configuration or phase space itself.

The manner in which classical symmetry would be interpreted is somewhat different, since one could apply the infinitesimal generators to the coordinates and momenta and determine the actual change of the phase space. The infinitesimal transformations could be integrated, to determine the global symmetry group. Quantum mechanically, the generators could be applied to the coordinate or momentum operators, rather than to the wavefunctions, in order to try to interpret them as transformation operators on the configuration or momentum space. However, since the generators take precisely the form of the constants of motion used in Fock's discussion of the accidental degeneracy, the only possible consistent interpretation would be to obtain the same symmetry group, namely, $O(3)$.

We have only discussed the technique proposed by

Dulock for constructing general symmetry operators and constants of the motion, but the operators obtained are substantially the same as those obtained by the other authors whom we have mentioned, and consequently the same considerations would apply. Likewise, we have not discussed Hamiltonians of the most general form, since it appears that the literal preservation of the Poisson bracket relations as commutation relations does not occur in general. The details of this failure are quite important to resolve the seeming paradox of the "universal symmetry group" but will be given a more extensive treatment elsewhere. Even though it is then somewhat exceptional, the two-dimensional hydrogen atom merits interest if only because it seemed at first sight to be such a clear case of a problem to which the universal symmetry concept would apply and give results in disagreement with the known symmetry. The lesson learned, that multiple-valued canonical mappings may lead to spurious wavefunctions, should provide a useful caution to one who is tempted to forget the precise requirements of quantization in other contexts as well. Not only do spurious wavefunctions occur, but Eq. (21) written in terms of $n = n_1 + n_2$ yields

$$E = -2(n/2 - \tfrac{1}{2})^{-2}, \tag{41}$$

which shows that they belong to spurious energy levels. Only an even n, corresponding to acceptable wavefunctions, produces the proper energy levels.

Although this paper has been addressed to a very specific point, it is worth considering its relation to some other recent works. First, Stehle and Han[16] have made a quite general analysis of the effect of single-valuedness or its lack in canonical transformations, from the point of view of the Bohr–Sommerfeld quantization conditions of the old quantum mechanics. Only if the canonical mapping is finite valued, and more especially single valued, can the quantification in terms of the phase integral be preserved, since otherwise one cycle of a closed orbit does not map into a complete image cycle. Such a result gives a very intuitive idea of why the attempt to form a universal symmetry group by passage to a set of canonical coordinates in which the system would be equivalent to force-free motion or to a harmonic oscillator might fail. Suggestive as it is, though, their paper still leaves untouched the mechanism by which such mappings could fail in quantum mechanics.

The same authors, in a companion paper which we have already cited, have shown that $SU(2)$ cannot be the classical symmetry group of the two-dimensional hydrogen atom, a conclusion which we have confirmed quantum mechanically. The example is anomalous, however, because the failure is due to a technicality with respect to the covering group. Such is not the case in three dimensions, in which $SU(3)$ can be exhibited as an explicit symmetry group for the Keplerian bound orbits, and in which it is not the covering group of the traditional $O(4)$ symmetry group. Moreover, the generators of this $SU(3)$ group can be formed quantum mechanically, can be shown to account for the degeneracy of the hydrogen atom, and in fact only fail to form the $SU(3)$ group itself because of some minor but immutable defects in their commutation relations. These results will be elaborated elsewhere, but meanwhile should serve notice that the classical arguments, although suggestive, are not conclusive for understanding the quantum-mechanical symmetry.

There has been a great upsurge of interest in the de Sitter group and its relation to the hydrogen atom. This group arises when the energy ladder operators are taken into account in addition to the symmetry operators, and produces what has come to be called the noninvariance group of the hydrogen atom. Its attraction lies in the possibility of subsuming the entire bound-state hydrogen spectrum in one irreducible representation of the de Sitter group, to the extent that one might not only speculate about universal symmetry groups for a large class of systems, but universal noninvariance groups as well. Inevitably, such investigations have brought to light ladder operators for the hydrogen atom, which have invariably been found to be related to the ladder operators originally proposed by Schrödinger,[17] and subsequently refined and perfected by Infeld and Hull.[18]

Noteworthy among the recent investigations of the de Sitter group are the papers of Bander and Itzykson,[19] Pratt and Jordan,[20] Musto,[21] Bacry,[22] Han,[23] and a related paper of Hwa and Nuyts.[24] Although very similar, our results must be distinguished from these others on two accounts. First of all, the "parabolic" coordinates suitable in two dimensions are different from those suitable in several dimensions, and contribute in great measure to the apparent unitary symmetry of the two-dimensional hydrogen atom. Secondly, the attempt to construct a

[16] P. Stehle and M. Y. Han, Phys. Rev. **159**, 1076 (1967).

[17] E. Schrödinger, Proc. Roy. Irish Acad. **46A**, 9 (1940); **46A** 183 (1941).

[18] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).

[19] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330, 346 (1966).

[20] R. H. Pratt and T. F. Jordan, Phys. Rev. **148**, 1276 (1966).

[21] R. Musto, Phys. Rev. **148**, 1274 (1966).

[22] H. Bacry, Nuovo Cimento **41A**, 222 (1966).

[23] M. Y. Han, Nuovo Cimento **42B**, 367 (1966).

[24] R. C. Hwa and T. Nuyts, Phys. Rev. **145**, 1188 (1966).

de Sitter group generally presumes that the imbedded symmetry group is an *orthogonal* group, while we have explicitly attempted to construct a *unitary* symmetry group. Consequently we have more nearly implemented the theory of Hwa and Nuyts than that of the other authors, because we have effectively embedded the symmetry group $SU(2)$ in the symplectic group $sp(4)$. The effect of the double-valuedness of some of the parabolic wavefunctions is that we must select one of the two families, that of even parity, into which the symplectic representation splits. However, the relation of the de Sitter group to the symplectic group in four dimensions is a traditional one.

Finally the exhaustive study of the possible accidental degeneracies in two dimensions made by Winternitz, Smorodinskii, Uhlir, and Fris[4] should be mentioned. They found four types of potentials, which showed accidental degeneracy, and not surprisingly the applicable symmetry groups were either $O(3)$ or its covering group $SU(2)$. In particular, they found that the separation of the hydrogen atom in parabolic coordinates led to double-valued as well as single-valued wavefunctions, whose exclusion produced the symmetry group $O(3)$ in place of $SU(2)$, which is exactly the result which we have found from having tried to formulate the $SU(2)$ group initially.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Elementary Solutions of the Reduced Three-Dimensional Transport Equation

HANS G. KAPER*
Stanford University, Stanford, California

(Received 1 August 1967)

By treating one of the space dimensions exactly and approximating the other two by the exp $(-i\mathbf{B}\cdot\mathbf{r})$ assumption, which is suggested by asymptotic transport theory, it is possible to reduce the three-dimensional transport equation to an equation that is of one-dimensional form and that still contains details of the complete three-dimensional angular distribution. In this paper we develop the method of elementary solutions for the reduced transport equation in the case of time-independent, monoenergetic neutron transport in homogeneous media with isotropic scattering. The spectrum of the transport operator consists of a pair of discrete points if B^2 is sufficiently small and a continuum which occupies a two-dimensional region in the complex spectral plane. The eigenfunctions possess full-range and half-range orthogonality and completeness properties, which are proved via the solution of two-dimensional integral equations using the theory of boundary-value problems for generalized analytic functions. As applications we solve the Green's function for an infinite homogeneous prism and the albedo operator for a semi-infinite homogeneous prism. Also discussed are possible generalizations of the method to more complicated forms of the reduced transport equation.

1. INTRODUCTION

In the past few years considerable progress has been made in the study of the one-dimensional transport equation. The introduction of the method of elementary solutions (singular eigenfunctions) by Case[1] has proved to be very efficient in obtaining exact solutions for one-speed and energy-dependent problems in systems with plane symmetry. Attempts have been made to generalize Case's approach to systems with other symmetry properties[2] and to arbitrary geometrical configurations.[3] However, these attempts do not seem to have been very successful.

In systems with rectangular symmetry the neutron distribution can often be described adequately by a "reduced transport equation" of one-dimensional form, which is derived from the three-dimensional transport equation through the application of ideas of asymptotic transport theory. Treating one of the space dimensions exactly and approximating the solution to the transport equation in the other, transverse, space dimensions by a solution of the type exp $(-i\mathbf{B}\cdot\mathbf{r})$, which is suggested by asymptotic transport theory, one obtains a one-dimensional equation which contains information on the solution of the complete three-dimensional problem. This approach has recently been used by Williams[4] to solve the one-speed Milne problem in a finite prism via the Wiener–Hopf technique.

It is the purpose of this paper to develop the method of elementary solutions for the reduced

* Present address: Department of Mathematics, Groningen University, Groningen, The Netherlands.

[1] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).

[2] G. J. Mitsis, "Transport Solutions to the Monoenergetic Critical Problem," Argonne National Laboratory Report-6787, 1963.

[3] E. H. Bareiss, "Isotropic Transport Operators in Three Independent Spatial Variables," in *Proc. Symp. Appl. Math., Vol. 20* (American Mathematical Society, Providence, R.I.) (to be published 1968).

[4] M. M. R. Williams, Nukleonik **9**, 305 (1967).

de Sitter group generally presumes that the imbedded symmetry group is an *orthogonal* group, while we have explicitly attempted to construct a *unitary* symmetry group. Consequently we have more nearly implemented the theory of Hwa and Nuyts than that of the other authors, because we have effectively embedded the symmetry group $SU(2)$ in the symplectic group $sp(4)$. The effect of the double-valuedness of some of the parabolic wavefunctions is that we must select one of the two families, that of even parity, into which the symplectic representation splits. However, the relation of the de Sitter group to the symplectic group in four dimensions is a traditional one.

Finally the exhaustive study of the possible accidental degeneracies in two dimensions made by Winternitz, Smorodinskii, Uhlir, and Fris[4] should be mentioned. They found four types of potentials, which showed accidental degeneracy, and not surprisingly the applicable symmetry groups were either $O(3)$ or its covering group $SU(2)$. In particular, they found that the separation of the hydrogen atom in parabolic coordinates led to double-valued as well as single-valued wavefunctions, whose exclusion produced the symmetry group $O(3)$ in place of $SU(2)$, which is exactly the result which we have found from having tried to formulate the $SU(2)$ group initially.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Elementary Solutions of the Reduced Three-Dimensional Transport Equation

HANS G. KAPER*
Stanford University, Stanford, California

(Received 1 August 1967)

By treating one of the space dimensions exactly and approximating the other two by the exp $(-i\mathbf{B}\cdot\mathbf{r})$ assumption, which is suggested by asymptotic transport theory, it is possible to reduce the three-dimensional transport equation to an equation that is of one-dimensional form and that still contains details of the complete three-dimensional angular distribution. In this paper we develop the method of elementary solutions for the reduced transport equation in the case of time-independent, monoenergetic neutron transport in homogeneous media with isotropic scattering. The spectrum of the transport operator consists of a pair of discrete points if B^2 is sufficiently small and a continuum which occupies a two-dimensional region in the complex spectral plane. The eigenfunctions possess full-range and half-range orthogonality and completeness properties, which are proved via the solution of two-dimensional integral equations using the theory of boundary-value problems for generalized analytic functions. As applications we solve the Green's function for an infinite homogeneous prism and the albedo operator for a semi-infinite homogeneous prism. Also discussed are possible generalizations of the method to more complicated forms of the reduced transport equation.

1. INTRODUCTION

In the past few years considerable progress has been made in the study of the one-dimensional transport equation. The introduction of the method of elementary solutions (singular eigenfunctions) by Case[1] has proved to be very efficient in obtaining exact solutions for one-speed and energy-dependent problems in systems with plane symmetry. Attempts have been made to generalize Case's approach to systems with other symmetry properties[2] and to arbitrary geometrical configurations.[3] However, these attempts do not seem to have been very successful.

In systems with rectangular symmetry the neutron distribution can often be described adequately by a "reduced transport equation" of one-dimensional form, which is derived from the three-dimensional transport equation through the application of ideas of asymptotic transport theory. Treating one of the space dimensions exactly and approximating the solution to the transport equation in the other, transverse, space dimensions by a solution of the type exp $(-i\mathbf{B}\cdot\mathbf{r})$, which is suggested by asymptotic transport theory, one obtains a one-dimensional equation which contains information on the solution of the complete three-dimensional problem. This approach has recently been used by Williams[4] to solve the one-speed Milne problem in a finite prism via the Wiener–Hopf technique.

It is the purpose of this paper to develop the method of elementary solutions for the reduced

* Present address: Department of Mathematics, Groningen University, Groningen, The Netherlands.

[1] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).

[2] G. J. Mitsis, "Transport Solutions to the Monoenergetic Critical Problem," Argonne National Laboratory Report-6787, 1963.

[3] E. H. Bareiss, "Isotropic Transport Operators in Three Independent Spatial Variables," in *Proc. Symp. Appl. Math., Vol. 20* (American Mathematical Society, Providence, R.I.) (to be published 1968).

[4] M. M. R. Williams, Nukleonik **9**, 305 (1967).

three-dimensional transport equation. For illustration, attention here is restricted to the case of monoenergetic neutron transport in homogeneous media with isotropic scattering. The spectrum of the transport operator associated with the time-independent reduced three-dimensional transport equation consists of a pair of discrete points if the system is not too small in the transverse directions and a continuum which occupies a two-dimensional region in the complex spectral plane. The eigenfunctions possess orthogonality and completeness properties of both "full-range" and "half-range" type. The proofs of these properties are based upon the solution of two-dimensional integral equations that degenerate to the singular integral equations of the analogous one-dimensional case if $B^2 \to 0$. These two-dimensional integral equations can be reduced to boundary-value problems for generalized analytic functions and solved via a method described by Vekua[5] and Gakhov.[6] In order to illustrate the analysis we apply the results to the solution of the Green's function for an infinite homogeneous prism, and to the albedo operator for a semi-infinite homogeneous prism. From these two examples it is apparent that the method of singular eigenfunctions given in this paper can be applied to any transport problem in two- and three-dimensional systems, whenever the approximation of asymptotic transport theory in the transverse space dimensions is justified. At the end of the paper we discuss possible generalizations to more general situations.

2. THE TIME-INDEPENDENT REDUCED TRANSPORT EQUATION

The time-independent, monoenergetic transport equation in homogeneous systems with rectangular symmetry can be written as[7]

$$\mu \frac{\partial N}{\partial x} + (1-\mu^2)^{\frac{1}{2}}\left[\cos\psi \frac{\partial N}{\partial y} + \sin\psi \frac{\partial N}{\partial z}\right] + N(x, y, z, \mu, \psi) = \frac{c}{4\pi}\int_{-1}^{1} d\mu' \int_0^{2\pi} d\psi' N(x, y, z, \mu', \psi'), \quad (2.1)$$

where we have assumed isotropic scattering; distances are measured in units of mean free path. N is the neutron density in phase space $R^3 \times \Omega^3$, R^3 is the three-dimensional coordinate space (x, y, z), and Ω^3 is the unit sphere in velocity space. In Ω^3, coordinates

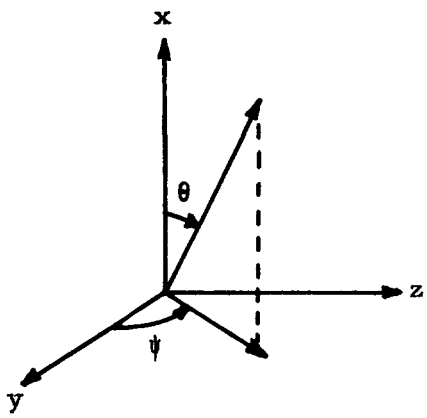


FIG. 1. Coordinate system in phase space.

are the Eulerian angles ($\theta = \cos^{-1}\mu$, ψ); see Fig. 1. c is the mean number of secondary neutrons per collision.

In accordance with asymptotic transport theory we assume a solution of the form $\exp[-i(B_y y + B_z z)]$ in the transverse directions and associate B_y^2 and B_z^2 with the corresponding bucklings. Then the reduced three-dimensional transport equation is

$$\mu\left(\frac{\partial \tilde{N}}{\partial x}\right) + \{1 - i(1-\mu^2)^{\frac{1}{2}}[B_y \cos\psi + B_z \sin\psi]\} \times \tilde{N}(x, B_y, B_z, \mu, \psi) = \frac{c}{4\pi}\int_{-1}^{1} d\mu' \int_0^{2\pi} d\psi' \tilde{N}(x, B_y, B_z, \mu', \psi'). \quad (2.2)$$

With the definitions

$$B^2 = B_y^2 + B_z^2, \quad \Delta = \tan^{-1}(B_z/B_y), \quad (2.3)$$

and

$$\Phi(x, \mu, \psi) = \{1 - i(1-\mu^2)^{\frac{1}{2}}[B_y \cos\psi + B_z \sin\psi]\} \times \tilde{N}(x, B_y, B_z, \mu, \psi), \quad (2.4)$$

Eq. (2.2) can be written as

$$\left[1 + \frac{\mu}{1 - i(1-\mu^2)^{\frac{1}{2}} B \cos(\psi - \Delta)} \frac{\partial}{\partial x}\right]\Phi(x, \mu, \psi) = \frac{c}{4\pi}\int_{-1}^{1} d\mu' \int_0^{2\pi} d\psi' \frac{\Phi(x, \mu', \psi')}{1 - i(1-\mu'^2)^{\frac{1}{2}} B \cos(\psi' - \Delta)}. \quad (2.5)$$

Equation (2.5) constitutes the complex form of a system of two real equations. It is useful to define a complex variable ζ as

$$\zeta = \frac{\mu}{1 - i(1-\mu^2)^{\frac{1}{2}} B \cos(\psi - \Delta)}. \quad (2.6)$$

[5] I. N. Vekua, *Generalized Analytic Functions* (Pergamon Press, Inc., New York, 1962).

[6] F. D. Gakhov, *Boundary Value Problems* (Pergamon Press, Inc., New York, 1966).

[7] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1958).

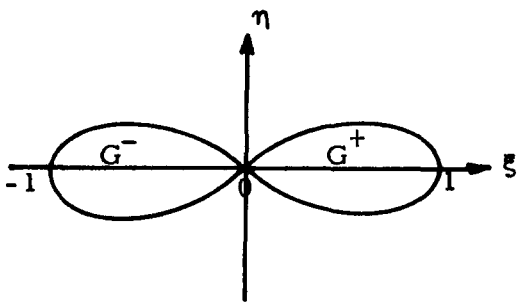


FIG. 2. The domain G in the complex ζ plane.

If we write $\zeta = \xi + i\eta$, Eq. (2.6) defines a transformation of variables $(\mu, \psi) \Rightarrow (\xi, \eta)$:

$$\xi = \frac{\mu}{1 + (1 - \mu^2)B^2 \cos^2(\psi - \Delta)},$$
$$\eta = \frac{\mu(1 - \mu^2)^{\frac{1}{2}} B \cos(\psi - \Delta)}{1 + (1 - \mu^2)B^2 \cos^2(\psi - \Delta)}. \quad (2.7)$$

The inverse transformation is

$$\mu = (\xi^2 + \eta^2)/\xi,$$
$$\psi = \Delta + \cos^{-1}\{\eta/B[\xi^2 - (\xi^2 + \eta^2)^2]^{\frac{1}{2}}\}. \quad (2.8)$$

The Jacobians for these transformations are

$$\frac{\partial(\xi, \eta)}{\partial(\mu, \psi)} = -\frac{\mu(1 - \mu^2)^{\frac{1}{2}} B \sin(\psi - \Delta)}{[1 + (1 - \mu^2)B^2 \cos^2(\psi - \Delta)]^2} \quad (2.9)$$

and

$$\frac{\partial(\mu, \psi)}{\partial(\xi, \eta)} = -\frac{\xi^2 + \eta^2}{\xi^2\{B^2[\xi^2 - (\xi^2 + \eta^2)^2] - \eta^2\}^{\frac{1}{2}}}. \quad (2.10)$$

The transformation (2.7) is a two-to-one mapping of the domain $(\Omega^3)' = (\Omega^3)^+ \cup (\Omega^3)^-$, where $(\Omega^3)^\pm = \{(\mu, \psi) \mid \mu \gtrless 0,\ 0 < |\mu| \leq 1,\ 0 \leq \psi < 2\pi\}$, onto a region $G = G^+ \cup G^-$ of the complex ζ plane. It is easily seen that the boundary $\partial G^\pm$ of $G^\pm$ is the image of the lines $\psi = \Delta$ and $\psi = \pi + \Delta$ under the mapping (2.7),

$$\partial G \equiv \eta^2 - B^2[\xi^2 - (\xi^2 + \eta^2)^2] = 0; \quad (2.11)$$

see Fig. 2. We remark that, if $B^2 \to 0$, the region G collapses into the interval $-1 \leq \xi \leq 1$ of the real axis.

The transformation of variables (2.6), together with the definition[8]

$$\Psi(x, \zeta) \equiv \Psi(x, \xi, \eta) = \Phi(x, \mu, \psi), \quad (2.12)$$

gives the following form for the reduced transport equation:

$$\Psi(x, \zeta) + \zeta \frac{\partial \Psi}{\partial x} = \iint_G g(\zeta')\Psi(x, \zeta')\, d\xi'\, d\eta' \quad (2.13)$$

with

$$g(\zeta) \equiv g(\xi, \eta) = \frac{c}{2\pi}\left|\frac{\partial(\mu, \psi)}{\partial(\xi, \eta)}\right| \times \left[\frac{1}{1 - i(1 - \mu^2)^{\frac{1}{2}} B \cos(\psi - \Delta)}\right]_{(\mu,\psi) \Rightarrow (\xi,\eta)}. \quad (2.14)$$

Essentially, Eq. (2.13) is of the conventional one-dimensional type. In the following we develop the method of elementary solutions for Eq. (2.13).

3. THE ELEMENTARY SOLUTIONS OF THE REDUCED TRANSPORT EQUATION

If Eq. (2.13) is formally written as $\partial\Psi/\partial x = A\Psi$, the reduced transport operator A is defined by

$$(A\varphi)(\zeta) = -\zeta^{-1}\left[\varphi(\zeta) - \iint_G g(\zeta')\varphi(\zeta')\, d\xi'\, d\eta'\right], \quad (3.1)$$

for all φ belonging to the domain of A. Thus one has the eigenvalue problem $A\varphi = -\nu^{-1}\varphi$, or

$$(\nu - \zeta)\varphi(\nu, \zeta) = \nu \iint_G g(\zeta')\varphi(\nu, \zeta')\, d\xi'\, d\eta'. \quad (3.2)$$

It is convenient to choose the normalization

$$\iint_G g(\zeta)\varphi(\nu, \zeta)\, d\xi\, d\eta = g(\nu). \quad (3.3)$$

In the space $\mathcal{E}'$ of distributions with support $\bar{G}$ on the (ξ, η) plane, the eigenvalue problem (3.2) admits the following solution:

$$\varphi(\nu, \zeta) = \nu g(\nu)/(\nu - \zeta) + \lambda(\nu)\delta(\nu - \zeta), \quad (3.4)$$

where the distributions $(\nu - \zeta)^{-1}$ and $\delta(\nu - \zeta)$ are defined by the functionals

$$\langle (\nu - \zeta)^{-1}, \phi(\zeta)\rangle = \iint_G \frac{\phi(\zeta)}{\nu - \zeta}\, d\xi\, d\eta, \quad (3.5)$$

$$\langle \delta(\nu - \zeta), \phi(\zeta)\rangle = \begin{cases} \phi(\nu), & \text{if } \nu \in G, \\ 0, & \text{if } \nu \notin \bar{G}, \end{cases} \quad (3.6)$$

for all functions ϕ that belong to the space of test functions $\mathcal{E}$. In Eq. (3.4), $\lambda(\nu)$ is an arbitrary complex-valued function. Since $\zeta \in G$, we distinguish two cases:

(1) $\nu \notin \bar{G}$. From Eqs. (3.3)–(3.6) one obtains the dispersion law

$$1 - \iint_G \frac{\nu g(\zeta)}{\nu - \zeta}\, d\xi\, d\eta = 0, \quad (3.7)$$

[8] Here and in the sequel, the notation $\Psi(x, \zeta)$ does not mean that Ψ is a holomorphic function of the complex variable ζ; it is only a shorthand notation for $\Psi(x, \zeta, \eta)$.

which determines the discrete spectrum of the operator A.

(2) $\nu \in G$. Equations (3.3)–(3.6) lead to the following expression for $\lambda(\nu)$:

$$\lambda(\nu) = 1 - \iint_G \frac{\nu g(\zeta)}{\nu - \zeta}\, d\xi\, d\eta. \tag{3.8}$$

Therefore, if we take $\lambda(\nu)$ as in Eq. (3.8), the domain G belongs to the spectrum of the operator A.

Let $L_p(\bar{G})$ be the set of functions $f(\zeta)$, defined on G, which satisfy the inequality

$$\left[\iint_G |f(\zeta)|^p\, d\xi\, d\eta\right]^{1/p} < \infty.$$

Following Vekua,[5] we define, for $f \in L_1(\bar{G})$,

$$T_G f(\zeta) = -\frac{1}{\pi}\iint_G \frac{f(\zeta')}{\zeta' - \zeta}\, d\xi'\, d\eta'. \tag{3.9}$$

The operator T plays a fundamental role in the theory of generalized analytic functions; its properties are discussed by Vekua in his monograph.[5] With reference to this work (especially Chap. I, Secs. 5.1–5.4) we state, without proof, the following theorems:

(A) If $f \in L_1(\bar{G})$, then $T_G f$ exists for all points ζ outside $\bar{G}$, is holomorphic outside $\bar{G}$, and vanishes at infinity; $T_G f$, regarded as a function of a point ζ of the domain G, exists almost everywhere and belongs to an arbitrary class $L_p(\bar{G}_*)$, where p is any number satisfying the inequality $1 \leq p < 2$ and G_* is any bounded domain of the plane.

(B) If $f \in L_1(\bar{G})$, then $T_G f \in D_{\bar{\zeta}}(G)$, i.e., the class of functions having a generalized derivative with respect to $\bar{\zeta}$ (for the definition of a generalized derivative, see Ref. 5, Chap. I, Sec. 5.2), and

$$\begin{aligned} \frac{\partial}{\partial \bar{\zeta}} T_G f(\zeta) &= f(\zeta), \quad \text{if} \quad \zeta \in G, \\ &= 0, \quad \text{if} \quad \zeta \notin \bar{G}. \end{aligned} \tag{3.10}$$

(C) If $f \in L_1(\bar{G})$ and $\partial g/\partial \bar{\zeta} = f$, then

$$g(\zeta) = \Phi(\zeta) + T_G f(\zeta), \tag{3.11}$$

where Φ is a function holomorphic inside G.

Now, it is easily verified that the function g, defined in Eq. (2.14), belongs to the class $L_1(\bar{G})$. Thus if we introduce the function

$$\Lambda(\zeta) = 1 - \pi\zeta T_G[g(\zeta)], \tag{3.12}$$

it follows that Λ exists for all $\zeta \notin \bar{G}$ and is holomorphic outside $\bar{G}$, with

$$\begin{aligned} \Lambda(\infty) &= \lim_{\zeta\to\infty} \Lambda(\zeta) = 1 - \iint_G g(\zeta)\, d\xi\, d\eta \\ &= 1 - c\,\frac{\tan^{-1} B}{B}. \end{aligned} \tag{3.13}$$

Inside G, Λ exists almost everywhere. Finally, $\Lambda \in L_p(\bar{G}_*)$ with $1 \leq p < 2$. $\Lambda(\zeta)$ possesses a generalized derivative with respect to $\bar{\zeta}$ which satisfies

$$\begin{aligned} \partial\Lambda/\partial\bar{\zeta} &= 0, \quad \text{if} \quad \zeta \notin \bar{G}, \\ &= -\pi\zeta g(\zeta), \quad \text{if} \quad \zeta \in G. \end{aligned} \tag{3.14}$$

From (3.7) and (3.12) we conclude that the discrete spectrum of A is determined by the equation

$$\Lambda(\zeta) = 0 \quad \text{for} \quad \zeta \notin \bar{G}. \tag{3.15}$$

On the other hand, in the case of the continuum part of the spectrum, it follows from (3.8) and (3.12) that

$$\lambda(\zeta) = \Lambda(\zeta) \quad \text{for} \quad \zeta \in G. \tag{3.16}$$

In the next section we discuss the existence of solutions of Eq. (3.15).

4. THE DISPERSION LAW

In terms of the original variables, the dispersion law (3.15) can be written as

$$\begin{aligned} \Lambda(\zeta) \equiv 1 - \frac{c}{4\pi}\int_{-1}^{1} d\mu \\ \times \int_0^{2\pi} \frac{d\psi}{1 - i(1-\mu^2)^{\frac{1}{2}} B\cos(\psi - \Delta) - \mu/\zeta} = 0 \\ \text{for} \quad \zeta \notin \bar{G}. \end{aligned} \tag{4.1}$$

The repeated integral can be evaluated. Writing z for ζ^{-1} we have

$$\begin{aligned} &\int_{-1}^{1} d\mu \int_0^{2\pi} \frac{d\psi}{1 - i(1-\mu^2)^{\frac{1}{2}} B\cos(\psi - \Delta) - z\mu} \\ &\quad = 2\int_{-1}^{1} \frac{1}{[(1-z\mu)^2 + (1-\mu^2)B^2]^{\frac{1}{2}}} \\ &\qquad \times \left[\tan^{-1}\left\{\frac{(1-z\mu)\tan(\psi - \Delta)}{[(1-z\mu)^2 + (1-\mu^2)B^2]^{\frac{1}{2}}}\right\}\right]_{\psi=0}^{\psi=\pi} d\mu \\ &\quad = 2\pi\int_{-1}^{1} \frac{d\mu}{[(1-z\mu)^2 + (1-\mu^2)B^2]^{\frac{1}{2}}} \\ &\quad = \frac{2\pi}{(z^2 - B^2)^{\frac{1}{2}}} \log\frac{1 + (z^2 - B^2)^{\frac{1}{2}}}{1 - (z^2 - B^2)^{\frac{1}{2}}}. \end{aligned} \tag{4.2}$$

Therefore, the dispersion law associated with the reduced three-dimensional transport equation has the

form

$$\Lambda(\zeta, B^2) \equiv 1 - (c/2\kappa) \log [(1 + \kappa)/(1 - \kappa)] = 0,$$

where

$$\kappa = (\zeta^{-2} - B^2)^{\frac{1}{2}}, \quad \zeta \notin \bar{G}. \tag{4.3}$$

In terms of the variable κ, Eq. (4.3) is formally identical with the dispersion law for one-dimensional transport problems, which has been discussed extensively in the literature, e.g., by Case *et al.*[9] If we denote the solutions of the one-dimensional dispersion law $\Lambda(\zeta, 0) = 0$, $\zeta \notin [-1, 1]$ by $\pm L$, i.e., if L is implicitly given by the equation

$$1 - \tfrac{1}{2}cL \log [(L + 1)/(L - 1)] = 0, \quad L \notin [-1, 1], \tag{4.4}$$

then we can state the following conclusion.

For $0 < c \leq 1$ (L real, $1 \leq L < \infty$), the dispersion law (4.3) has two real roots $\zeta = \pm\nu_0$, $1 < \nu_0 < B^{-1}$ if $0 < B^2 < 1 - 1/L^2$, no roots if $B^2 \geq 1 - 1/L^2$.

For $c > 1$ (L purely imaginary, $L = i|L|$), the dispersion relation (4.3) has two purely imaginary roots $\zeta = \pm\nu_0 = \pm i|\nu_0|$ if $0 < B^2 \leq -1/L^2$, two real roots $\zeta = \pm\nu_0$, $1 \leq \nu_0 < \infty$ if $-1/L^2 \leq B^2 < 1 - 1/L^2$, no roots if $B^2 \geq 1 - 1/L^2$.

For fixed values of c, the roots $\pm\nu_0$ of Eq. (4.3), if they exist, can be found from the roots $\pm L$ of Eq. (4.4) through the expression

$$\nu_0^{-2} = L^{-2} + B^2. \tag{4.5}$$

The domain of existence and the character of the solutions $\pm\nu_0$ of the dispersion law (4.3) are illustrated in Fig. 3.

We define a characteristic function $\chi_c(B^2)$,

$$\begin{aligned} \chi_c(B^2) &= 1, \quad \text{if} \quad 0 < B^2 < 1 - 1/L^2, \\ &= 0, \quad \text{if} \quad B^2 \geq 1 - 1/L^2, \end{aligned} \tag{4.6}$$

so that $\chi_c(B^2) = 1$ if the dispersion law has one (and only one) pair of solutions, $\chi_c(B^2) = 0$ if the dispersion law has no solutions.

The expression for $\Lambda(\zeta, B^2)$, given in Eq. (4.3), has been derived from Eq. (4.1) under the assumption that $\zeta \notin \bar{G}$. Of course, Eq. (4.3) can be used to define a function $\Lambda^*(\zeta, B^2)$ that is analytic in the ζ plane cut along the real axis from $\zeta = -(1 + B^2)^{-\frac{1}{2}}$ to $\zeta = (1 + B^2)^{-\frac{1}{2}}$ which is the analytic continuation of Λ in the interior of the domain G. However, it must be realized that, for $\zeta \in G$, this analytic continuation $\Lambda^*(\zeta, B^2)$ does not correspond to the function

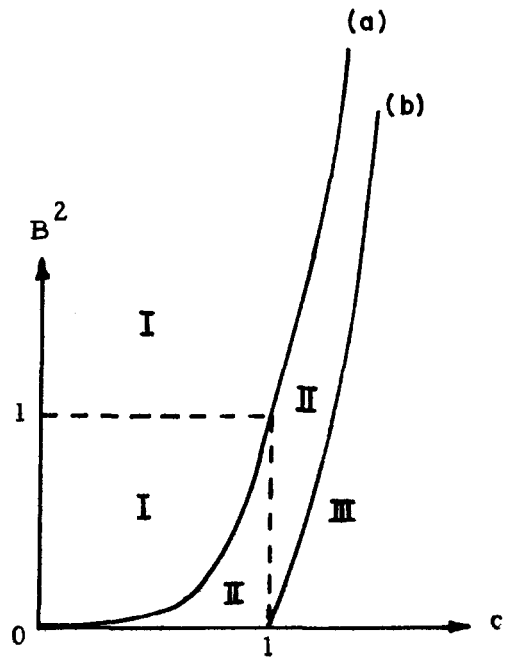


FIG. 3. Existence of solutions to Eq. (4.3). (a) $B^2 = 1 - 1/L^2$, (b) $B^2 = -1/L^2$; I, no solutions; II, two real solutions; III, two purely imaginary solutions.

$\Lambda(\zeta, B^2)$ that has been defined originally in Eq. (3.12). As a matter of fact, if $\zeta \in G$, it is not possible to derive an expression for Λ in terms of elementary functions of ζ only, as was done for $\zeta \notin \bar{G}$ in Eq. (4.2).

The analysis of the foregoing section can now be summarized as follows. The spectrum of the reduced transport operator A, defined by the expression (3.1) on the space $\mathcal{E}'$ of distributions with support $\bar{G}$ on the complex ζ plane, is composed of:

(1) a discrete part if $\chi_c(B^2) = 1$, which consists of one pair of eigenvalues $\nu = \pm\nu_0$; these eigenvalues are found as the solutions of Eq. (4.3); the elementary solutions of Eq. (2.13) are

$$\varphi_\pm(x, \zeta) = \varphi_\pm(\zeta) \exp(\mp x/\nu_0), \tag{4.7}$$

with

$$\varphi_\pm(\zeta) = \pm\nu_0 g(\pm\nu_0)/(\pm\nu_0 - \zeta). \tag{4.8}$$

(2) a continuum part, which consists of all $\nu \in G$; the elementary solutions of Eq. (2.13) are

$$\varphi_\nu(x, \zeta) = \varphi(\nu, \zeta) \exp(-x/\nu), \tag{4.9}$$

with

$$\varphi(\nu, \zeta) = \nu g(\nu)/(\nu - \zeta) + \Lambda(\nu)\delta(\nu - \zeta), \tag{4.10}$$

where $\Lambda(\nu)$, $\nu \in G$, is defined in Eq. (3.12).

In the following sections we study the full-range and half-range properties of the set of eigendistributions (4.8) and (4.10).

[9] K. M. Case, F. de Hoffmann, and G. Placzek, *Introduction to the Theory of Neutron Diffusion, Vol. I* (U.S. Government Printing Office, Washington D.C., 1953).

5. FULL-RANGE ORTHOGONALITY AND COMPLETENESS

(A) From Eq. (3.2) one derives in the usual way[1] that the eigendistributions $\varphi_\pm(\zeta)$ and $\varphi(\nu, \zeta)$ are mutually orthogonal in the full domain G with weight function $\zeta g(\zeta)$, i.e.,

$$\iint_G \zeta g(\zeta)\varphi_\pm(\zeta)\varphi_\mp(\zeta)\, d\xi\, d\eta = 0, \tag{5.1}$$

$$\iint_G \zeta g(\zeta)\varphi_\pm(\zeta)\varphi(\nu, \zeta)\, d\xi\, d\eta = 0, \quad \nu \in G, \tag{5.2}$$

$$\iint_G \zeta g(\zeta)\varphi(\nu, \zeta)\varphi(\nu', \zeta)\, d\xi\, d\eta = 0, \quad \nu, \nu' \in G, \quad \nu \neq \nu'. \tag{5.3}$$

Since the explicit forms of the functions are known, the normalization integrals can be evaluated. We write

$$\iint_G \zeta g(\zeta)\varphi_\pm^2(\zeta)\, d\xi\, d\eta = N_\pm, \tag{5.4}$$

$$\iint_G \zeta g(\zeta)\varphi(\nu, \zeta)\varphi(\nu', \zeta)\, d\xi\, d\eta = N(\nu)\delta(\nu - \nu'), \quad \nu, \nu' \in G. \tag{5.5}$$

Before we may continue, a remark is due with respect to the left-hand side of Eq. (5.5). In the literature on the one-dimensional transport equation there has been some confusion about the interpretation of a similar integral of a direct product of two generalized functions; see Kaper.[10] As far as Eq. (5.5) is concerned, the situation is essentially easier since the functional (3.5), which defines the generalized function $(\nu - \zeta)^{-1}$, exists as an ordinary integral, whereas in one-dimensional transport theory the analogous functional exists only as the principal value of a Cauchy integral. It can be proved that the following decomposition in partial fractions is valid in the sense of the theory of generalized functions:

$$(\nu - \zeta)^{-1}(\nu' - \zeta)^{-1} = (\nu' - \nu)^{-1}[(\nu - \zeta)^{-1} - (\nu' - \zeta)^{-1}]. \tag{5.6}$$

Now, in order to evaluate the integral (5.5) we use the decomposition in partial fractions:

$$\zeta(\nu - \zeta)^{-1}(\nu' - \zeta)^{-1} = (\nu' - \nu)^{-1}[\nu(\nu - \zeta)^{-1} - \nu'(\nu' - \zeta)^{-1}]. \tag{5.7}$$

[10] H. G. Kaper, Nucl. Sci. Eng. **24**, 423 (1966).

Then we have, for ν and $\nu' \in G$,

$$\iint_G \zeta g(\zeta)\varphi(\nu, \zeta)\varphi(\nu', \zeta)\, d\xi\, d\eta = \frac{\nu\nu' g(\nu)g(\nu')}{\nu' - \nu}\left\{\left[\Lambda(\nu) - \nu\iint_G \frac{g(\zeta)}{\zeta - \nu}\, d\xi\, d\eta\right] - \left[\Lambda(\nu') - \nu'\iint_G \frac{g(\zeta)}{\zeta - \nu'}\, d\xi\, d\eta\right]\right\} + \nu g(\nu)\Lambda^2(\nu)\delta(\nu - \nu'). \tag{5.8}$$

One verifies easily that the expression inside curly braces vanishes. Thus we find that the normalization constant in Eq. (5.5) is given by

$$N(\nu) = \nu g(\nu)\Lambda^2(\nu), \quad \nu \in G. \tag{5.9}$$

Finally, with the relation

$$\iint_G \zeta g(\zeta)\varphi(\nu, \zeta)\varphi(\nu', \zeta)\, d\xi\, d\eta = \nu\nu' g(\nu)g(\nu')\frac{\Lambda(\nu) - \Lambda(\nu')}{\nu - \nu'}, \tag{5.10}$$

which is valid for arbitrary ν and $\nu' \notin \bar{G}$, we find the normalization constant $N_\pm$ by taking the limits $\nu \to \pm\nu_0$, $\nu' \to \pm\nu_0$,

$$N_\pm = \nu_0^2 g^2(\pm\nu_0)\Lambda'(\pm\nu_0). \tag{5.11}$$

Since

$$\Lambda'(\pm\nu_0) = \pm\frac{L^2}{\nu_0^3}\left(\frac{c}{1 - 1/L^2} - 1\right), \tag{5.12}$$

where L and ν_0 are related through Eq. (4.5), this can also be written as

$$N_\pm = \pm g^2(\pm\nu_0)\frac{L^2}{\nu_0}\left(\frac{c}{1 - 1/L^2} - 1\right). \tag{5.13}$$

(B) Next we prove the completeness of the set of generalized functions $[\varphi(\nu, \zeta)]$, $\nu \in G$, defined in Eq. (4.10), supplemented by the functions $\varphi_\pm(\zeta)$, defined in Eq. (4.8) if $\chi_c(B^2) = 1$, for the class $\mathcal{E}'$ of generalized functions with support $\bar{G}$. Although we have not been able to verify this assumption, we shall suppose that Λ does not vanish inside G. If, however, the assumption is violated, the theory can be modified without great difficulty by supplementing the above set of functions with the functions

$$\varphi_{\pm ij}(\zeta) = \{(\partial/\partial\nu)^j[\nu g(\nu)/(\nu - \zeta)]\}_{\nu=\nu_i},$$

$j = 0, 1, \cdots, m_i - 1$, where we have supposed that $\zeta = \nu_i$ is a zero of Λ inside G of multiplicity m_i.

The proof is based upon a "full-range closure

relation." Thus we verify the identity

$$[a_+\varphi_+(\zeta) + a_-\varphi_-(\zeta)]\chi_c(B^2) + \iint_G A(\zeta')\varphi(\zeta', \zeta)\, d\xi'\, d\eta' = \psi(\zeta), \quad (5.14)$$

for any generalized function ψ with support $\bar{G}$, where the coefficients in the expansion have been calculated according to the orthogonality relations (5.1)–(5.5):

$$a_\pm = \frac{1}{N_\pm}\iint_G \zeta g(\zeta)\psi(\zeta)\varphi_\pm(\zeta)\, d\xi\, d\eta, \quad (5.15)$$

$$A(\zeta) = \frac{1}{N(\zeta)}\iint_G \zeta' g(\zeta')\psi(\zeta')\varphi(\zeta, \zeta')\, d\xi'\, d\eta'. \quad (5.16)$$

Since the set of generalized functions $\{\delta(\nu - \zeta)\}$, $\nu \in G$, is complete for the class $\mathcal{E}'$ of generalized functions with support $\bar{G}$ (see Ref. 11) it is sufficient to verify the closure relation (5.14) for $\psi(\zeta) = \delta(\nu - \zeta)$, ν fixed, $\nu \in G$. In this case the expansion is

$$\nu g(\nu)\Big\{\iint_G \frac{\varphi(\zeta', \nu)\varphi(\zeta', \zeta)}{\zeta' g(\zeta')\Lambda^2(\zeta')}\, d\xi'\, d\eta' + \left[\frac{\varphi_+(\nu)\varphi_+(\zeta)}{\nu_0^2 g^2(\nu_0)\Lambda'(\nu_0)} + \frac{\varphi_-(\nu)\varphi_-(\zeta)}{\nu_0^2 g^2(-\nu_0)\Lambda'(-\nu_0)}\right]\chi_c(B^2)\Big\} = \delta(\nu - \zeta) \quad (5.17)$$

for $\nu, \zeta \in G$.

We start with the first term of the expression between curly braces, which will be denoted by $F(\nu, \zeta)$. Substitution of φ from Eq. (4.10) and use of the partial-fraction decomposition (5.6) gives, after some trivial algebra,

$$F(\nu, \zeta) = \frac{\delta(\nu - \zeta)}{\nu g(\nu)} + \frac{1}{\nu - \zeta}[f(\nu) - f(\zeta)], \quad (5.18)$$

where we have introduced the function f:

$$f(z) = \frac{1}{\Lambda(z)} + \iint_G \frac{\zeta' g(\zeta')}{\Lambda^2(\zeta')}\frac{d\xi'\, d\eta'}{\zeta' - z}, \quad z \in G. \quad (5.19)$$

Now we recall Eq. (3.14) and notice that Eq. (5.19) can also be written as

$$f(z) = \frac{1}{\Lambda(z)} + \frac{1}{\pi}\iint_G \frac{\partial}{\partial\bar{\zeta}'}\left[\frac{1}{\Lambda(\zeta')}\right]\frac{d\xi'\, d\eta'}{\zeta' - z}, \quad z \in G. \quad (5.20)$$

[11] I. M. Gel'fand and G. E. Schilow, *Verallgemeinerte Funktionen (Distributionen)* (VEB-Deutscher Verlag der Wissenschaften, Berlin, 1964), Vol. III.

According to Vekua[5] (Chap. I, Sec. 4.1), this is then equivalent to

$$f(z) = \frac{1}{2\pi i}\int_{\partial G}\frac{1}{\Lambda(\zeta')}\frac{d\zeta'}{\zeta' - z}, \quad z \in G. \quad (5.21)$$

The contour integral can be evaluated by aid of Cauchy's theorem of residues applied to the exterior of the domain G, where Λ^{-1} is known to be analytic except for two poles at $\pm\nu_0$. The result is

$$f(z) = \frac{1}{\Lambda(\infty)} - \left\{\sum_{\zeta'=\pm\nu_0}\frac{1}{\zeta' - z}\operatorname{res}\left[\frac{1}{\Lambda(\zeta')}\right]\right\}\chi_c(B^2). \quad (5.22)$$

Together with Eq. (5.18) this gives

$$F(\nu, \zeta) = \frac{\delta(\nu - \zeta)}{\nu g(\nu)} - \left[\frac{1}{(\nu_0 - \zeta)(\nu_0 - \nu)\Lambda'(\nu_0)} + \frac{1}{(-\nu_0 - \zeta)(-\nu_0 - \nu)\Lambda'(-\nu_0)}\right]\chi_c(B^2). \quad (5.23)$$

It is easily verified that the expressions between brackets in Eqs. (5.17) and (5.23) are identical, so the conjectured identity (5.17) is indeed true and therefore the full set of eigendistributions (4.8) and (4.10) has the full-range completeness property stated above.

6. HALF-RANGE ORTHOGONALITY AND COMPLETENESS

(A) In this section we prove that the eigendistributions $\varphi(\nu, \zeta)$, with $\nu \in G^+(\nu \in G^-)$, and the eigendistribution $\varphi_+(\varphi_-)$ are mutually orthogonal in the half-range $G^+(G^-)$ with a certain weight function. We shall give the arguments for the case G^+ only, since the case G^- can be treated in exactly the same way. The proof actually consists of the construction of the unknown function H in the relations

$$\iint_{G^+}\zeta g(\zeta)H(\zeta)\varphi_+(\zeta)\varphi(\nu, \zeta)\, d\xi\, d\eta = 0, \quad \nu \in G^+, \quad (6.1)$$

$$\iint_{G^+}\zeta g(\zeta)H(\zeta)\varphi(\nu, \zeta)\varphi(\nu', \zeta)\, d\xi\, d\eta = 0, \quad \nu, \nu' \in G^+, \quad \nu \neq \nu'. \quad (6.2)$$

When the decomposition (5.6) is applied we find the identities

$$\varphi_+(\zeta)\varphi(\nu, \zeta) = (\nu - \nu_0)^{-1} \times [\nu g(\nu)\varphi_+(\zeta) - \nu_0 g(\nu_0)\varphi(\nu, \zeta)], \quad (6.3)$$

$$\varphi(\nu, \zeta)\varphi(\nu', \zeta) = (\nu' - \nu)^{-1} \times [\nu' g(\nu')\varphi(\nu, \zeta) - \nu g(\nu)\varphi(\nu', \zeta)] + \Lambda(\nu)\Lambda(\nu')\delta(\nu - \zeta)\delta(\nu' - \zeta). \quad (6.4)$$

Thus it is obvious that the orthogonality relations (6.1) and (6.2) will hold if and only if

$$\iint_{G^+} \zeta g(\zeta) H(\zeta) \varphi_+(\zeta)\, d\xi\, d\eta = \alpha \nu_0 g(\nu_0), \tag{6.5}$$

$$\iint_{G^+} \zeta g(\zeta) H(\zeta) \varphi(\nu, \zeta)\, d\xi\, d\eta = \alpha \nu g(\nu), \quad \nu \in G^+, \tag{6.6}$$

where α is a constant. With expressions (4.8) and (4.10) substituted, Eqs. (6.5) and (6.6) can be written as

$$\frac{1}{\pi} \iint_{G^+} \frac{\partial \Lambda}{\partial \bar{\zeta}'} H(\zeta') \frac{d\xi'\, d\eta'}{\zeta' - \nu_0} = \alpha, \tag{6.7}$$

$$\frac{1}{\pi} \iint_{G^+} \frac{\partial \Lambda}{\partial \bar{\zeta}'} H(\zeta') \frac{d\xi'\, d\eta'}{\zeta' - \zeta} + \Lambda(\zeta) H(\zeta) = \alpha, \quad \zeta \in G^+. \tag{6.8}$$

These equations are two-dimensional integral equations for the unknown function H. Similar equations have been encountered recently in a study of neutron wave propagation.[12] It turns out to be possible to relate the solution of Eqs. (6.7) and (6.8) to the solution of a boundary-value problem for a generalized analytic function, which can be found explicitly via a method described by Vekua in his monograph.[5]

Let X be a function from the class $D_\zeta(G^+)$, which satisfies the differential equation

$$\Lambda(\zeta) \frac{\partial X}{\partial \bar{\zeta}} - \frac{\partial \Lambda}{\partial \bar{\zeta}} X(\zeta) = 0, \quad \text{for} \quad \zeta \in G^+. \tag{6.9}$$

Then, the ratio X/Λ is holomorphic in G^+. So, if we define

$$H(\zeta) = X(\zeta)/\Lambda(\zeta) \tag{6.10}$$

it follows that

$$\frac{\partial X}{\partial \bar{\zeta}} = H(\zeta) \frac{\partial \Lambda}{\partial \bar{\zeta}} \quad \text{for} \quad \zeta \in G^+. \tag{6.11}$$

Equations (6.10) and (6.11) suffice to prove that Eqs. (6.7) and (6.8) can be written respectively as

$$\frac{1}{\pi} \iint_{G^+} \frac{\partial X}{\partial \bar{\zeta}'} \frac{d\xi'\, d\eta'}{\zeta' - \nu_0} = \alpha, \tag{6.12}$$

and

$$\frac{1}{\pi} \iint_{G^+} \frac{\partial X}{\partial \bar{\zeta}'} \frac{d\xi'\, d\eta'}{\zeta' - \zeta} + X(\zeta) = \alpha, \quad \text{for} \quad \zeta \in G^+, \tag{6.13}$$

and these expressions are, in turn, equivalent with

$$\frac{1}{2\pi i} \int_{\partial G^+} \frac{X(\zeta')}{\zeta' - \nu_0} d\zeta' = \alpha \tag{6.14}$$

and

$$\frac{1}{2\pi i} \int_{\partial G^+} \frac{X(\zeta')}{\zeta' - \zeta} d\zeta' = \alpha \quad \text{for} \quad \zeta \in G^+. \tag{6.15}$$

Of course, if $\chi_c(B^2) = 0$, i.e., if the discrete spectrum is empty, the condition (6.14) must be omitted. We recall that, in the conditions (6.14) and (6.15), α is any constant independent of ζ.

Thus, the problem of finding the weight function H in the orthogonality relations (6.1) and (6.2) has been reduced to the problem of finding a function X that is a solution of the boundary value problem (6.9), which satisfies the conditions (6.14) and (6.15).

The general solution to the homogeneous equation (6.9) can be found with the aid of Theorem C quoted in Sec. III. It is given by

$$X(\zeta) = \Phi(\zeta) \exp \Gamma(\zeta), \tag{6.16}$$

with

$$\Gamma(\zeta) = T_{G^+}[\Lambda^{-1} \partial \Lambda / \partial \zeta], \tag{6.17}$$

where Φ is an arbitrary function which is holomorphic inside G^+. Now, the definition (6.16) of the X function can be extended to the exterior of G^+, where $\exp \Gamma(\zeta)$ is known to exist everywhere as a holomorphic function. Thus, if Φ is an arbitrary function that is holomorphic in the whole complex plane except possibly at ∂G^+, X defined by Eq. (6.16) exists at every point ζ outside G^+ and is holomorphic outside G^+.

The expression for Γ can be simplified:

$$\Gamma(\zeta) = \log \Lambda(\zeta) - \frac{1}{2\pi i} \int_{\partial G^+} \log \Lambda(\zeta') \frac{d\zeta'}{\zeta' - \zeta}, \quad \text{for} \quad \zeta \in G^+,$$

$$\Gamma(\zeta) = -\frac{1}{2\pi i} \int_{\partial G^+} \log \Lambda(\zeta') \frac{d\zeta'}{\zeta' - \zeta}, \quad \text{for} \quad \zeta \notin \bar{G}^+. \tag{6.18}$$

It is easily verified that $\Lambda(\bar{\zeta}) = \overline{\Lambda(\zeta)}$. Therefore, since ∂G^+ is symmetric with respect to the real axis, Eqs. (6.18) can also be written as

$$\Gamma(\zeta) = \log \Lambda(\zeta) + \frac{1}{\pi} \int_0^1 \theta(\zeta') \frac{d\zeta'}{\zeta' - \zeta}, \quad \text{for} \quad \zeta \in G^+,$$

$$\Gamma(\zeta) = \frac{1}{\pi} \int_0^1 \theta(\zeta') \frac{d\zeta'}{\zeta' - \zeta}, \quad \text{for} \quad \zeta \notin \bar{G}^+, \tag{6.19}$$

with

$$\theta(\zeta) = \arg \Lambda(\zeta). \tag{6.20}$$

[12] H. G. Kaper, J. H. Ferziger, and S. K. Loyalka, "Neutron Wave Propagation with a One-Term Degenerate Thermalization Kernel," in *Proceedings of the Symposium on Neutron Thermalization and Reactor Spectra, Ann Arbor, Michigan, 1967* (International Atomic Energy Agency, Vienna, 1968).

In Eqs. (6.19) the integrals are to be evaluated along the upper half of ∂G^+; $\theta(\zeta)$ has to be changed continuously along the path of integration once a particular branch of $\log \Lambda$ has been chosen. The right-hand sides of Eqs. (6.19) are identical on the boundary ∂G^+.

From the properties of Cauchy integrals (cf. Ref. 13, p. 74), it follows that

$$\exp \Gamma(\zeta) \sim \zeta^{-\theta(0)/\pi}, \quad \text{if} \quad \zeta \to 0,$$
$$\sim (1 - \zeta)^{\theta(1)/\pi}, \quad \text{if} \quad \zeta \to 1. \tag{6.21}$$

Since $\Lambda(0) = 1$, hence $\theta(0) = 0$, one verifies from the argument principle of the theory of functions of a complex variable, taking into account some simple symmetry considerations, that $\theta(1) = \pi$ if $\chi_c(B^2) = 1$, $\theta(1) = 0$ if $\chi_c(B^2) = 0$. Thus, $\exp \Gamma(\zeta)$ is nonvanishing in the whole complex plane except at $\zeta = 1$, where it has a (simple) zero if and only if $\chi_c(B^2) = 1$. On the other hand, if $\zeta \to \infty$, $\exp \Gamma(\zeta)$ approaches the value 1.

Now, consider the conditions (6.14) and (6.15). Obviously, they are fulfilled if and only if X is holomorphic outside $\bar{G}^+$, including the point at infinity, and if X vanishes at $\zeta = \nu_0$ in the case $\chi_c(B^2) = 1$.

The conditions (6.14) and (6.15) must be met by a proper choice of the function Φ in the expression (6.16) for X. One verifies that the function

$$\Phi(\zeta) = \alpha(\nu_0 - \zeta)/(1 - \zeta), \quad \text{if} \quad \chi_c(B^2) = 1, \tag{6.22}$$
$$\Phi(\zeta) = \alpha, \quad \text{if} \quad \chi_c(B^2) = 0, \tag{6.23}$$

satisfies all the requirements. Moreover, the resulting function X has the property that it has neither poles nor zeros in the whole complex plane except at $\zeta = \nu_0$, where it has a simple zero in the case $\chi_c(B^2) = 1$. Thus, if we take the constant α equal to $[\Lambda(\infty)]^{\frac{1}{2}}$, we have proved the existence of a function H:

$$H(\zeta) = \frac{[\Lambda(\infty)]^{\frac{1}{2}}(\nu_0 - \zeta)}{(1 - \zeta)\Lambda(\zeta)} \exp \Gamma(\zeta), \quad \text{if} \quad \chi_c(B^2) = 1, \tag{6.24}$$

$$H(\zeta) = \frac{[\Lambda(\infty)]^{\frac{1}{2}}}{\Lambda(\zeta)} \exp \Gamma(\zeta), \quad \text{if} \quad \chi_c(B^2) = 0, \tag{6.25}$$

such that the orthogonality relations (6.1) and (6.2) hold.

The function H defined in Eqs. (6.24) and (6.25) reduces to Chandrasekhar's H function[14] if $B^2 \to 0$, as is easily shown. The new H function satisfies similar identities as Chandrasekhar's H function. For example, if we consider the function $F(\zeta) = \Lambda(\zeta)/X(\zeta)X(-\zeta)$, which is analytic in the entire complex plane and has the limiting value 1 at infinity, we infer from Liouville's theorem and Eq. (6.18) the identity

$$[H(\zeta)H(-\zeta)]^{-1} = \Lambda(\zeta). \tag{6.26}$$

From Eq. (6.26), a relation between the X and H function can be found:

$$H(\zeta) = [X(-\zeta)]^{-1}. \tag{6.27}$$

We end the discussion of the half-range orthogonality properties with the calculation of the normalization integrals

$$\iint_{G^+} \zeta g(\zeta) H(\zeta) \varphi_+^2(\zeta)\, d\xi\, d\eta = N_+^+, \tag{6.28}$$

$$\iint_{G^+} \zeta g(\zeta) H(\zeta) \varphi(\nu, \zeta) \varphi(\nu', \zeta)\, d\xi\, d\eta = N^+(\nu)\delta(\nu - \nu'), \quad \nu, \nu' \in G^+. \tag{6.29}$$

From Eqs. (6.4) and (6.29), it is immediately clear that we have

$$N^+(\nu) = \nu g(\nu) H(\nu) \Lambda^2(\nu), \quad \nu \in G^+. \tag{6.30}$$

From the formula

$$\iint_{G^+} \zeta g(\zeta) H(\zeta) \varphi(\nu, \zeta) \varphi(\nu', \zeta)\, d\xi\, d\eta = \nu\nu' g(\nu) g(\nu') \frac{X(\nu) - X(\nu')}{\nu - \nu'}, \tag{6.31}$$

which is valid for arbitrary ν and $\nu' \notin \bar{G}^+$, we find the normalization constant N_+^+ by taking the limits $\nu \to \nu_0$, $\nu' \to \nu_0$:

$$N_+^+ = \nu_0^2 g^2(\nu_0) X'(\nu_0), \tag{6.32}$$

which can be written as

$$N_+^+ = \nu_0^2 g^2(\nu_0) [(d/d\nu) H^{-1}(-\nu)]_{\nu_0} = \nu_0^2 g^2(\nu_0) H(\nu_0) \Lambda'(\nu_0). \tag{6.33}$$

(B) Next we prove the completeness of the set of generalized functions $\{\varphi(\nu, \zeta)\}$, $\nu \in G^+$, defined in Eq. (4.10), supplemented by the function $\varphi_+(\zeta)$, defined in Eq. (4.8), if $\chi_c(B^2) = 1$, for the class $\mathcal{E}'$ of generalized functions with support $\bar{G}^+$. As in Sec. 5B we assume that Λ does not vanish inside G^+. The proof is based upon a "half-range closure relation" of the type

$$a_+ \varphi_+(\zeta) \chi_c(B^2) + \iint_{G^+} A(\zeta') \varphi(\zeta', \zeta)\, d\xi'\, d\eta' = \psi(\zeta) \tag{6.34}$$

[13] N. I. Muskhelishvili, *Singular Integral Equations* (P. Noordhoff, Ltd., Groningen The Netherlands, 1953).

[14] S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, London, 1950).

for any generalized function ψ with support $\bar{G}^+$, where the coefficients have been calculated according to the orthogonality relations (6.1), (6.28), and (6.29):

$$a_+ = \frac{1}{N_+^+} \iint_{G^+} \zeta g(\zeta) H(\zeta) \psi(\zeta) \varphi_+(\zeta)\, d\xi\, d\eta, \tag{6.35}$$

$$A(\zeta) = \frac{1}{N(\zeta)} \iint_{G^+} \zeta' g(\zeta') H(\zeta') \psi(\zeta') \varphi(\zeta, \zeta')\, d\xi'\, d\eta'. \tag{6.36}$$

For the same reason as in Sec. 5B, it suffices to verify the closure relation (6.34) for $\psi(\zeta) = \delta(\nu - \zeta)$, ν fixed, $\nu \in G^+$. In this case the expansion is

$$\nu g(\nu) H(\nu) \left[\iint_{G^+} \frac{\varphi(\zeta', \nu)\varphi(\zeta', \zeta)}{\zeta' g(\zeta') H(\zeta') \Lambda^2(\zeta')}\, d\xi'\, d\eta' + \frac{\varphi_+(\nu)\varphi_+(\zeta)}{\nu_0^2 g^2(\nu_0) H(\nu_0) \Lambda'(\nu_0)} \chi_c(B^2) \right] = \delta(\nu - \zeta) \tag{6.37}$$

for ν, $\zeta \in G^+$.

The first term of the expression between square brackets, which is denoted by $F^+(\nu, \zeta)$, can be transformed with the aid of Eq. (5.6). The result is

$$F^+(\nu, \zeta) = \frac{\delta(\nu - \zeta)}{\nu g(\nu) H(\nu)} + \frac{1}{\nu - \zeta} [f^+(\nu) - f^+(\zeta)], \tag{6.38}$$

with

$$f^+(z) = \frac{1}{H(z)\Lambda(z)} + \iint_{G^+} \frac{\zeta' g(\zeta')}{H(\zeta')\Lambda^2(\zeta')} \frac{d\xi'\, d\eta'}{\zeta' - z}, \quad z \in G^+. \tag{6.39}$$

Now, we recall Eq. (3.13) and the fact that H is holomorphic in G^+. This enables us to rewrite Eq. (6.39) in the form

$$f^+(z) = \frac{1}{H(z)\Lambda(z)} + \frac{1}{\pi} \iint_{G^+} \frac{\partial}{\partial \bar{\zeta}'} \left[\frac{1}{H(\zeta')\Lambda(\zeta')} \right] \frac{d\xi'\, d\eta'}{\zeta' - z}, \quad z \in G^+. \tag{6.40}$$

This again reduces to a contour integral

$$f^+(z) = \frac{1}{2\pi i} \int_{\partial G^+} \frac{1}{H(\zeta')\Lambda(\zeta')} \frac{d\xi'\, d\eta'}{\zeta' - z}, \quad z \in G^+, \tag{6.41}$$

which can be evaluated by aid of Cauchy's theorem of residues applied to the exterior of G^+, where $(H\Lambda)^{-1} = X^{-1}$ is known to be analytic except for a pole at ν_0. The result is

$$f^+(z) = \frac{1}{[\Lambda(\infty)]^{\frac{1}{2}}} - \left\{ \frac{1}{\nu_0 - z} \operatorname*{res}_{\zeta' = \nu_0} \left[\frac{1}{H(\zeta')\Lambda(\zeta')} \right] \right\} \chi_c(B^2), \tag{6.42}$$

which gives, in Eq. (6.38),

$$F^+(\nu, \zeta) = \frac{\delta(\nu - \zeta)}{\nu g(\nu) H(\nu)} - \left[\frac{1}{(\nu_0 - \zeta)(\nu_0 - \nu) H(\nu_0) \Lambda'(\nu_0)} \right] \chi_c(B^2). \tag{6.43}$$

It is easily verified that the terms with $\chi_c(B^2)$ in Eqs. (6.37) and (6.43) are identical. Therefore, the conjectured identity (6.37) is indeed true, which proves the half-range completeness property stated above.

(C) As we have stated at the beginning of Sec. 6A, all the properties concerning orthogonality and completeness that have been formulated for the half-range G^+, can be formulated analogously for the half-range G^-. It can be shown that the orthogonality and completeness relations can be generalized to any partial range ΔG which is a simply connected subset of G. This generalization does not present any new difficulty with respect to the half-range case considered in this section and can be done in the same spirit as in one-dimensional transport theory.[15,16]

7. APPLICATIONS

In this section we illustrate the previous theory on some standard problems of neutron transport theory.

(A) We first consider Green's function for a uniform infinite prism. The Green's function N_g satisfies the equation

$$\mu \frac{\partial N_g}{\partial x} + (1 - \mu^2)^{\frac{1}{2}} \times \left[\cos\psi \frac{\partial N_g}{\partial y} + \sin\psi \frac{\partial N_g}{\partial z} \right] + N(x, y, z, \mu, \psi) = \frac{c}{4\pi} \int_{-1}^{1} d\mu' \int_0^{2\pi} d\psi' N_g(x, y, z, \mu', \psi') + \frac{1}{4\pi} \delta(x)\delta(y)\delta(z)\delta(\mu - \mu_0)\delta(\psi - \psi_0). \tag{7.1}$$

Taking Fourier transforms in the y and z directions,

$$\tilde{N}_g(x, B_y, B_z, \mu, \psi) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz N_g(x, y, z, \mu, \psi) \exp[i(B_y y + B_z z)], \tag{7.2}$$

we find that $\tilde{N}_g$ satisfies the homogeneous equation (2.2) for $x \neq 0$ and the jump condition

$$\mu[\tilde{N}_g(x = +0) - \tilde{N}_g(x = -0)] = (1/4\pi)\delta(\mu - \mu_0)\delta(\psi - \psi_0). \tag{7.3}$$

[15] I. Kuščer, N. J. McCormick, and G. C. Summerfield, Ann. Phys. (N.Y.) **30**, 411 (1964).
[16] I. Kuščer and F. Shure, J. Math. Phys. **8**, 823 (1967).

Then, if we use the complex variable $\zeta = \xi + i\eta$, defined in Eq. (2.6), and introduce a new function

$$\Psi_g(x, \zeta) \equiv \Psi_g(x, \xi, \eta) = \{[1 - i(1 - \mu^2)^{\frac{1}{2}}(B_y \cos\psi + B_z \sin\psi)] \times \tilde{N}_g(x, B_y, B_z, \mu, \psi)\}_{(\mu,\psi)\Rightarrow(\xi,\eta)}, \quad (7.4)$$

we find that Ψ_g satisfies Eq. (2.13), for $x \neq 0$, and the jump condition

$$\zeta[\Psi_g(+0, \zeta) - \Psi_g(-0, \zeta)] = (1/4\pi)\delta(\zeta - \zeta_0). \quad (7.5)$$

A solution of Eq. (2.13), which vanishes as $|x| \to \infty$, can be expanded as

$$\Psi_g(x, \zeta) = \pm a_\pm \varphi_\pm(x, \zeta)\chi_c(B^2) \pm \iint_{G^\pm} A(\zeta')\varphi_{\zeta'}(x, \zeta)\, d\xi'\, d\eta', \quad \text{for} \quad x \gtrless 0, \quad (7.6)$$

where $\varphi_\pm(x, \zeta)$ and $\varphi_{\zeta'}(x, \zeta)$ have been defined in Eqs. (4.7) and (4.9). From the condition (7.5) we find an equation for the expansion coefficients:

$$[a_+\varphi_+(\zeta) + a_-\varphi_-(\zeta)]\chi_c(B^2) + \iint_G A(\zeta')\varphi(\zeta', \zeta)\, d\xi'\, d\eta' = \frac{\delta(\zeta - \zeta_0)}{4\pi}, \quad (7.7)$$

where $\varphi_\pm(\zeta)$ and $\varphi(\zeta', \zeta)$ have been defined in Eqs. (4.8) and (4.10). Using the full-range orthogonality relations we obtain the solution

$$a_\pm = g(\zeta_0)\varphi_\pm(\zeta_0)/4\pi N_\pm, \quad (7.8)$$

$$A(\zeta) = g(\zeta_0)\varphi(\zeta, \zeta_0)/4\pi N(\zeta). \quad (7.9)$$

For example, if the source is isotropic, we find the angular density from

$$\Psi_0(x, \zeta) = \iint_G \Psi_g(x, \zeta; \zeta_0)\, d\xi_0\, d\eta_0 = \frac{1}{4\pi}\left[\frac{g(\nu_0)\varphi_+(x, \zeta)}{N_+}\chi_c(B^2) + \iint_{G^+} \frac{g(\zeta')\varphi_{\zeta'}(x, \zeta)}{N(\zeta')}\, d\xi'\, d\eta'\right] \quad (7.10)$$

for $x > 0$, and similarly for $x < 0$. From this expression we derive the Fourier transform of the density at x due to an isotropic source at $x = 0$,

$$\tilde{\rho}_0(x) = \int_{-1}^{1} d\mu \int_0^{2\pi} d\psi \tilde{N}_g(x, B_y, B_z, \mu, \psi) = \frac{e^{-x/\nu_0}}{c\nu_0^2\Lambda'(\nu_0)}\chi_c(B^2) + \iint_{G^+} \frac{g(\zeta)e^{-x/\zeta}}{c\zeta\Lambda^2(\zeta)}\, d\xi\, d\eta, \quad (7.11)$$

for $x > 0$, and similarly for $x < 0$.

(B) Now we consider the albedo operator for a semi-infinite homogeneous prism. Suppose we want to find the solution N_a of Eq. (2.1) in the region $0 \leq x < \infty$, which vanishes as $x \to \infty$, subject to the boundary condition

$$N_a(0, y, z, \mu, \psi) = \delta(y)\delta(z)\delta(\mu - \mu_0)\delta(\psi - \psi_0), \quad \mu, \mu_0 > 0. \quad (7.12)$$

Taking Fourier transforms as in Eq. (7.2), using the variable $\zeta = \xi + i\eta$, Eq. (2.6), and introducing the new function

$$\Psi_a(x, \zeta) \equiv \Psi_a(x, \xi, \eta) = \{[1 - i(1 - \mu^2)^{\frac{1}{2}}(B_y \cos\psi + B_z \sin\psi)] \times \tilde{N}_a(x, B_y, B_z, \mu, \psi)\}_{(\mu,\psi)\Rightarrow(\xi,\eta)}, \quad (7.13)$$

we find that Ψ_a satisfies the homogeneous equation (2.13) for $x > 0$, subject to the boundary condition

$$\Psi_a(0, \zeta) = \delta(\zeta - \zeta_0), \quad \zeta \in G^+. \quad (7.14)$$

A general solution of Eq. (2.13) which vanishes at infinity, is

$$\Psi_a(x, \zeta) = a_+\varphi_+(x, \zeta)\chi_c(B^2) + \iint_{G^+} A(\zeta')\varphi_{\zeta'}(x, \zeta)\, d\xi'\, d\eta'. \quad (7.15)$$

The boundary condition (7.14) then gives the equation

$$a_+\varphi_+(\zeta)\chi_c(B^2) + \iint_{G^+} A(\zeta')\varphi(\zeta', \zeta)\, d\xi'\, d\eta' = \delta(\zeta - \zeta_0) \quad (7.16)$$

for $\zeta \in G^+$. Using the half-range orthogonality relations, we obtain the solution

$$a_+ = \zeta_0 g(\zeta_0)H(\zeta_0)\varphi_+(\zeta_0)/N_+^+, \quad (7.17)$$

$$A(\zeta) = \zeta_0 g(\zeta_0)H(\zeta_0)\varphi(\zeta, \zeta_0)/N^+(\zeta). \quad (7.18)$$

It is possible to express the emerging angular density at the surface $x = 0$ in terms of the ingoing angular density at $x = 0$, if one uses a generalized version of Eq. (6.29):

$$\iint_{G^+} \zeta g(\zeta)H(\zeta)\varphi(\nu, \zeta)\varphi(\nu', \zeta)\, d\xi\, d\eta = N^+(\nu)\delta(\nu - \nu')[1 - \Theta(\nu)] - \nu' g(\nu')[H(-\nu')]^{-1}\varphi(\nu, \nu')\Theta(\nu) - \nu g(\nu)[H(-\nu)]^{-1}\varphi(\nu', \nu)\Theta(\nu'), \quad (7.19)$$

with $\Theta(z) = 0$ if $z \in G^+$, $= 1$ if $z \notin \bar{G}^+$. Then, if at the free surface $x = 0$ the incoming angular density is described by a function $f(\zeta)$ with $\zeta \in G^+$, from Eq.

(7.19), upon multiplication of the half-range expansion of f with $\zeta g(\zeta)H(\zeta)\varphi(-\zeta', \zeta)$, $\zeta' \in G^+$, and integration over the half-range G^+, we find the following identity:

$$f(-\zeta) = \iint_{G^+} f(\zeta')\zeta' g(\zeta') \frac{H(\zeta)H(\zeta')}{\zeta + \zeta'} \, d\xi' \, d\eta', \quad \zeta \in G^+. \tag{7.20}$$

Hence, the right-hand side of this expression describes the albedo operator for a semi-infinite prism, which relates the outgoing angular neutron density to the ingoing one.

8. CONCLUSIONS

In the previous sections we developed the method of elementary solutions for an important class of three-dimensional transport problems, i.e., those problems that are described adequately by the reduced three-dimensional transport equation. Naturally, the approximations that lead to this reduced transport equation prevent us from gaining any knowledge of the exact angular neutron distribution in the transverse space dimensions, e.g., at the lateral surfaces of a prism. However, the angular neutron distribution in the longitudinal direction is treated in a more accurate way than in a one-dimensional theory, in which plane symmetry is assumed. In particular, the effect of transverse leakage on such quantities as the diffusion length is accounted for.

It was shown that a varied set of transport problems can be treated in a unified manner with the present method. Although, at first sight, the formal complexity may seem formidable, it should be emphasized that many formulas become more transparent if one returns from the complex variable ζ to the original variables μ and ψ; with the complex variable ζ, however, it is possible to interpret the equations in terms of the theory of functions of a complex variable and the theory of generalized analytic functions.

With the present method one can formulate closed-form solutions for problems involving infinite or semi-infinite prisms. Problems in finite prisms are reduced to Fredholm integral equations. Generalizations of the present theory are possible. The extensions to time-dependent and to energy-dependent problems are currently under investigation. It is felt that anisotropic scattering can be taken into account. Problems involving two adjoining semi-infinite prisms are more difficult since the continuous parts of the spectrum are different if the media have different mean free paths. The explicit treatment of those problems will be given in future publications.

ACKNOWLEDGMENTS

A NATO Science Fellowship, awarded by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.), enabled me to stay at Stanford University. Part of this work has been supported by National Science Foundation Grant GK-737.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

De Sitter Symmetric Field Theory. I. One-Particle Theory

MAHMOUD M. BAKRI
Applied Mathematics Department, Faculty of Science, Cairo University, Giza, United Arab Republic

(Received 9 February 1968)

The formal structure and the free-particle solutions of the field equations $(S_{\mu\nu}p_\nu + m\gamma_\mu)\psi = p_\mu\psi$, derived recently by the author [Nuovo Cimento **51A**, 864 (1967)] for realizations of the inhomogeneous de Sitter group are discussed. The enveloping algebra of the group is developed, and the covariance of the field equations under the five-dimensional rotations C, P, and T is proved. Bhabha's representation of the matrices γ_μ is completed. Observables, expectation,values and the scalar product are defined, and classical conservation laws are derived. The field equations are derived from a variational principle for the usual Lagrangian density $\bar{\psi}(-i\gamma_\mu\partial_\mu + m)\psi$ under a certain restriction. The free-particle solutions of the field equations are obtained in the canonical and the extreme relativistic representations. The connections between the wavefunctions in these representations, and also in the Foldy–Wouthuysen representation, are derived.

1. INTRODUCTION

Recently the author[1] has derived field equations of the form

$$P_\mu\psi \equiv (S_{\mu\nu}p_\nu + m\gamma_\mu)\psi = p_\mu\psi \qquad (1.1)$$

which describe particles of definite mass m and definite spin s. These equations were derived for the finite-dimensional representations of the inhomogeneous de Sitter group $SO(4, 1)$ with the invariant $p_Ap_A = 0$. p_A is the momentum–energy–mass five-vector, with components p_k = Cartesian components of momentum, $p_4 = ip_0$, p_0 = energy, and $p_5 = m$. The $S_{AB} = (S_{\mu\nu}, S_{\mu5} = \gamma_\mu)$ are the generators of the homogeneous de Sitter group. Their finite-dimensional irreducible representations[2] are those of $SO(5)$, $R_5(\lambda_1, \lambda_2)$, characterized by two nonnegative integers (for bosons) or half-integers (for fermions), such that $\lambda_1 \geq \lambda_2$. γ_μ are Bhabha matrices[3] and $\Sigma_{\mu\nu} = i\lambda_1 S_{\mu\nu}$ is the relativistic spin tensor. Equations (1.1) lead directly to the Bhabha equation[3]

$$(\gamma_\mu p_\mu + m)\psi = 0, \qquad (1.2)$$

to the mass relation

$$p_Ap_A\psi \equiv (p_\mu p_\mu + m^2)\psi = 0, \qquad (1.3)$$

and to the equation

$$W_\mu W_\mu\psi = m^2\lambda_2(\lambda_2 + 1)\psi, \qquad (1.4)$$

where

$$W_\mu = (\lambda_1/2)\epsilon_{\mu\nu\alpha\beta}p_\nu S_{\alpha\beta} \qquad (1.5)$$

is the Pauli–Lubanski pseudovector. ψ decomposes under the Poincaré group (PG) into several components $\psi(s_1, s_2)$, which are hereafter called the Poincaré components. $\psi(s_1, s_2)$ transforms according to the irreducible representation $D(s_1, s_2)$ of the homogeneous Lorentz group (hLG).[4] $SO(4, 1)$ decomposes under hLG into all inequivalent $D(s_1, s_2)$ with[2]

$$\delta \leq |s_1 - s_2| \leq \lambda_2 \leq s_1 + s_2 \leq \lambda_1, \qquad (1.6)$$

where $\delta = 0$ for bosons and $\delta = \frac{1}{2}$ for fermions. Equations (1.3) and (1.4) are satisfied by each Poincaré component separately. Equation (1.4) defines the spin in the rest system.[5] Thus, the theory describes a particle with a definite mass m and a definite spin $s = \lambda_2$. m needs not be real or positive. However, in discussing C and T invariance, and in deriving the conservation laws, m is assumed real.

Several general field equations have been proposed by different authors.[6–22] However, Eqs. (1.1) have the advantage that they follow from a symmetry principle. In fact, Pursey,[23] McKerrel,[24] and Tung[21] have

[1] M. M. Bakri, Nuovo Cimento **51A**, 864 (1967).
[2] E. M. Corson, *Introduction to Tensors, Spinors And Relativistic Wave Equations* (Blackie & Son Ltd., London, 1954), Sec. 38, p. 169.
[3] H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945).
[4] R. Shaw, Nuovo Cimento **33**, 1074 (1964).
[5] E. P. Wigner, Ann. Math. **40**, 149 (1939).
[6] P. A. M. Dirac, Proc. Roy. Soc. (London) **A155**, 447 (1936).
[7] Iu. P. Stepanovskii, Ukr. Fiz. Zh. **9**, 1165 (1964).
[8] M. Fierz, Helv. Phys. Acta **12**, 3 (1939).
[9] M. Fierz and W. Pauli, Proc. Roy. Soc. (London) **A173**, 211 (1939).
[10] W. Rarita and J. Schwinger, Phys. Rev. **60**, 61 (1941).
[11] L. de Broglie, *Théorie générale des particules à spin* (Gauthier-Villars, Paris, 1943).
[12] Harish-Chandra, Phys. Rev. **71**, 793 (1947); Proc. Roy. Soc. (London) **A192**, 195 (1947).
[13] V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **34**, 211 (1948).
[14] F. Cap, Z. Naturforsch. **8a**, 740, 748 (1953); Phys. Rev. **93**, 907 (1954).
[15] H. Donnert, Z. Naturforsch. **8a**, 745 (1953); Acta Phys. Austriaca **7**, 181 (1953).
[16] C. L. Hammer and R. H. Good, Jr., Phys. Rev. **108**, 822 (1957).
[17] D. L. Weaver, C. L. Hammer, and R. H. Good, Jr., Phys. Rev. **135**, B241 (1964).
[18] Confer S. Weinberg, *Lectures on Particles and Field Theory, Delivered at Brandeis Institute in Theoretical Physics, Vol. II, 1964* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1965), p. 407.
[19] D. L. Weaver and D. M. Fradkin, Nuovo Cimento **37**, 400 (1965).
[20] R. A. Berg, Nuovo Cimento **42A**, 148 (1966).
[21] Wu-ki Tung, Phys. Rev. Letters **16**, 763 (1966).
[22] Shau-Jin Chang, Phys. Rev. Letters **17**, 1024 (1966).
[23] D. L. Pursey, Ann. Phys. (N.Y.) **32**, 157 (1965).
[24] A. McKerrel, Ann. Phys. (N.Y.) **40**, 237 (1966).

shown that one can construct infinitely many self-consistent sets of field equations between any number of Poincaré components. The limitation of these equations should follow, then, from a higher symmetry group, which contains PG as a subgroup. The simplest generalization is the de Sitter group, leading to (1.1). Fronsdal, and Roman and Aghassi[25] have proposed the noncompact $SO(5)$ as a dynamical symmetry group, instead of PG. Equations (1.1) have also the advantage of including the supplementary conditions in a most economical way, in contradistinction with the Bargmann–Wigner equations[13] which have been used recently in connection with $S\tilde{U}(12)$ symmetry.[26,27] Equations (1.1) are simply the manifestly covariant generalization of Schrödinger's equation. They are the one-particle version of Schwinger's equations $P_\mu\psi = p_\mu\psi \equiv -i\partial_\mu\psi$. The expectation values of P_μ give the classical energy–momentum expressions. This was demonstrated by the author[28] in the case of the photon ($m = 0$, $s = 1$). Finally, Eqs. (1.1) offer the possibility of describing particles with the same spin $s = \lambda_2$ by different theories, differing in the value of λ_1. In fact the irreducible representation $R_5(\lambda_1, \lambda_2)$ may be obtained from the reduction of the direct product

$$R_5(\tfrac{1}{2}, \tfrac{1}{2}) \otimes R_5(\tfrac{1}{2}, \tfrac{1}{2}) \otimes \cdots \otimes R_5(\tfrac{1}{2}, \tfrac{1}{2}) \ (2\lambda_1 \text{ times}).$$

Thus $2\lambda_1$ may be considered as the number of quarks of which the particle is composed. Dirac's electron theory[29] ($\lambda_1 = \lambda_2 = \frac{1}{2}$) and Kemmer's scalar- and vector-meson theories[30] ($\lambda_1 = 0$, $\lambda_2 = 0, 1$) are special cases of Eqs. (1.1).

The purpose of the present paper is to study the formal structure of the field theory (1.1), and to obtain the free-particle solutions of these equations. We confine ourselves to the classical theory and do not consider the interactions. Also we do not try to discuss particular examples of the theory here.

In Sec. 2, the enveloping algebra of $SO(4, 1)$ is developed, and the operators of the four-dimensional inversions and charge conjugation are derived. The covariance of the field equations under the five-dimensional rotations P, C, and T is proved. In Sec. 3 Bhabha's representation of γ_μ is completed, by calculating the yet unspecified proportionality coefficients of Bhabha's matrix elements.[3] It is shown that Dirac's spinor theory for arbitrary spin[6] belongs to the particular representation $R_5(\lambda_1, \lambda_1)$ of our theory. The spinor and tensorial representations of Eqs. (1.1) will be discussed in a future publication.

In Sec. 4 the formal structure of the theory is discussed. Observables are defined, and classical conservation laws derived. It is shown that Eqs. (1.1) follow from a variational principle for the usual Lagrangian density $\bar{\psi}(\gamma_\mu p_\mu + m)\psi$, under a certain restriction. The expectation value and the scalar product are defined in a manifestly covariant way in configuration space, and are equivalent to that of Bargmann and Wigner.[13]

In Sec. 5 the solutions of Eqs. (1.1) are found in the canonical representation, obtained by the Chakrabarti transformation.[31] In Sec. 6 the plane-wave solutions are found in the extreme-relativistic representation obtained by the Cini–Touschek transformation.[32] The relations between the wavefunctions in these two representations, and also in the Foldy–Wouthuysen representation,[33] are derived. A study of the massless case is implied in the extreme-relativistic representation.

2. THE SYMMETRY GROUP

A. Generators of $SO(4, 1)$

The de Sitter group $SO(4, 1)$ is the group of linear orthogonal transformations in five dimensions,

$$x'_A = a_{AB}x_B, \quad x'_A x'_A = x_A x_A,$$
$$a_{AB}a_{AC} = a_{BA}a_{CA} = \delta_{BC}, \tag{2.1}$$

where the matrix elements a_{kj}, a_{44}, a_{k5}, and a_{5k} are real, while a_{k4}, a_{4k}, a_{45}, and a_{54} are pure imaginary. Here, lower case Latin indices take on the three values 1, 2, 3 [spatial three-dimensional Euclidean space, transforming under the subgroup $O(3)$]. Boldface letters imply three-vectors, as usual. Greek indices take on the four values 1, 2, 3, 4 [Minkowski space, transforming under the extended homogeneous Lorentz group $O(3, 1)$]. The physical timelike component is denoted by the index 0; $x_0 = -ix_4$. Upper case Latin indices take on the five values 1, 2, 3, 4, 5. We prefer to use a pseudo-Euclidean metric, to keep the close connection with the group $SO(5)$, which is characterized by real matrix elements a_{AB}. In this way we need not distinguish between covariant and contravariant components. In what follows we use the completely antisymmetric Levi-Civita pseudotensors ϵ_{kjn}, $\epsilon_{\mu\nu\alpha\beta}$, and ϵ_{ABCDE} with $\epsilon_{123} = \epsilon_{1234} = \epsilon_{12345} = 1$. We denote the transpose, the Hermitian

[25] C. Fronsdal, Rev. Mod. Phys. **37**, 221 (1965); P. Roman and J. J. Aghassi, Nuovo Cimento **42A**, 193 (1966).
[26] A. Salam, R. Delbourgo, and J. Strathdee, Proc. Roy. Soc. (London) **A284**, 146 (1965).
[27] R. J. Rivers, Phys. Rev. **145**, B1306 (1966).
[28] M. M. Bakri, Nucl. Phys. **B 87**, 289 (1966).
[29] Confer M. E. Rose, *Relativistic Electron Theory* (John Wiley & Sons, Inc., New York, 1961).
[30] N. Kemmer, Proc. Roy. Soc. (London) **A173**, 91 (1939).
[31] A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963).
[32] M. Cini and B. Touschek, Nuovo Cimento **7**, 422 (1958).
[33] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).

conjugate and the complex conjugate of a matrix S by S^{T}, $S^{\dagger}$, and S^{*} respectively.

We are interested in the finite-dimensional irreducible representations of the isomorphic group of linear transformations $S = S(a_{AB})$ of the Hilbert space:

$$\psi'(x'_A) = S\psi(x_A). \tag{2.2}$$

The generators, $\Sigma_{AB} = -\Sigma_{BA} = \Sigma^{\dagger}_{AB}$, of the infinitesimal rotation

$$a_{AB} = \delta_{AB} + \epsilon_{AB}, \quad \epsilon_{BA} = -\epsilon_{AB}, \tag{2.3}$$

$$S = I + (i/2)\epsilon_{AB}\Sigma_{AB}, \tag{2.4}$$

are those of $SO(5)$, satisfying Cartan's integrability conditions

$$i[\Sigma_{AB}, \Sigma_{CD}] = \delta_{AD}\Sigma_{BC} + \delta_{BC}\Sigma_{AD} - \delta_{AC}\Sigma_{BD} - \delta_{BD}\Sigma_{AC}. \tag{2.5}$$

For the finite rotations (2.1), S is determined from

$$S^{-1}\Sigma_{AB}S = a_{AC}a_{BD}\Sigma_{CD}. \tag{2.6}$$

For convenience, we denote

$$S_{AB} = -S^{\dagger}_{AB} = -(i/\lambda_1)\Sigma_{AB}, \tag{2.7}$$

$$\begin{aligned} \gamma_k &= S_{k5} = -\gamma^{\dagger}_k, \\ \gamma_0 &= -i\gamma_4 = -iS_{45} = \gamma^{\dagger}_0, \\ \alpha_k &= iS_{k4} = \alpha^{\dagger}_k, \\ \sigma_k &= (i/2)\epsilon_{kjn}S_{jn} = \sigma^{\dagger}_k. \end{aligned} \tag{2.8}$$

From (2.5) it follows that

$$\lambda_1[S_{\mu\nu}, \gamma_\alpha] = \delta_{\mu\alpha}\gamma_\nu - \delta_{\nu\alpha}\gamma_\mu, \tag{2.9}$$

$$S_{\mu\nu} = \lambda_1[\gamma_\mu, \gamma_\nu], \tag{2.10}$$

$$\begin{aligned} [\sigma_k, \sigma_j] &= [\alpha_k, \alpha_j] = -[\gamma_k, \gamma_j] \\ &= (i/\lambda_1)\epsilon_{kjn}\sigma_n, \\ [\alpha_k, \sigma_j] &= (i/\lambda_1)\epsilon_{kjn}\alpha_n, \\ [\gamma_k, \sigma_j] &= (i/\lambda_1)\epsilon_{kjn}\gamma_n, \\ \lambda_1[\gamma_k, \alpha_j] &= \gamma_0\delta_{kj}, \\ [\gamma_0, \sigma_k] &= 0, \\ \lambda_1[\gamma_0, \alpha_k] &= \gamma_k, \\ \lambda_1[\gamma_0, \gamma_k] &= \alpha_k. \end{aligned} \tag{2.11}$$

B. The Pseudovector

For convenience, we introduce the completely antisymmetric pseudotensor of the third rank

$$S_{ABC} = (\tfrac{1}{2})\epsilon_{ABCDE}S_{DE}. \tag{2.12}$$

Its ten independent components are

$$\begin{aligned} S_{5\mu\nu} = \tilde{S}_{\mu\nu} &\equiv (\tfrac{1}{2})\epsilon_{\mu\nu\alpha\beta}S_{\alpha\beta} = \lambda_1\epsilon_{\mu\nu\alpha\beta}\gamma_\alpha\gamma_\beta, \\ S_{\mu\nu\alpha} &= \epsilon_{\mu\nu\alpha\beta}\gamma_\beta. \end{aligned} \tag{2.13}$$

$\tilde{S}_{\mu\nu}$ is the four-dimensional dual of $S_{\mu\nu}$. Its components

$$\tilde{S}_{kj} = -\epsilon_{kjn}\alpha_n, \quad \tilde{S}_{4k} = i\sigma_k \tag{2.14}$$

are obtained from $S_{\mu\nu}$ by interchanging σ_k and α_k. We construct then the five-dimensional pseudovector

$$\Gamma_A = -\lambda_1 S_{ABC}S_{BC}/4(\lambda_1 + 1) = \Gamma^{\dagger}_A. \tag{2.15}$$

From (2.5) and (2.6) it follows that

$$\begin{aligned} \lambda_1[S_{AB}, \Gamma_C] &= \delta_{AC}\Gamma_B - \delta_{AB}\Gamma_C, \\ S^{-1}\Gamma_A S &= a_{AB}\Gamma_B/\det(a), \end{aligned} \tag{2.16}$$

where $\det(a) = \pm 1$ is the determinant of the transformation a_{AB}. In the four-dimensional notation, (2.16) reads

$$[\Gamma_5, S_{\mu\nu}] = 0, \tag{2.17}$$

$$\Gamma_\mu = \lambda_1[\Gamma_5, \gamma_\mu], \tag{2.18}$$

$$\lambda_1[\gamma_\mu, \Gamma_\nu] = \Gamma_5\delta_{\mu\nu}. \tag{2.19}$$

Also, from (2.15) we have

$$\begin{aligned} \Gamma_\mu &= -[\lambda_1/(\lambda_1 + 1)]\tilde{S}_{\mu\nu}\gamma_\nu \\ &= -[\lambda_1^2/(\lambda_1 + 1)]\epsilon_{\mu\nu\alpha\beta}\gamma_\nu\gamma_\alpha\gamma_\beta, \end{aligned} \tag{2.20}$$

and

$$\begin{aligned} \Gamma_5 &= \{\lambda_1/[4(\lambda_1 + 1)]\}\tilde{S}_{\mu\nu}S_{\nu\mu} \\ &= -\{\lambda_1^3/[2(\lambda_1 + 1)]\}\epsilon_{\mu\nu\alpha\beta}\gamma_\mu\gamma_\nu\gamma_\alpha\gamma_\beta \\ &= (\lambda_1/2)\gamma_\mu\Gamma_\mu = -(\lambda_1/2)\Gamma_\mu\gamma_\mu. \end{aligned} \tag{2.21}$$

In the three-dimensional notation

$$\begin{aligned} \Gamma_0 &= -\Gamma^{\dagger}_0 = -i\Gamma_4 = -[\lambda_1/(\lambda_1 + 1)](\boldsymbol{\sigma}\cdot\boldsymbol{\gamma}), \\ \boldsymbol{\Gamma} &= -[\lambda_1/(\lambda_1 + 1)](\boldsymbol{\sigma}\gamma_0 + i\boldsymbol{\alpha}\wedge\boldsymbol{\gamma}), \\ \Gamma_5 &= [\lambda_1/(\lambda_1 + 1)](\boldsymbol{\sigma}\cdot\boldsymbol{\alpha}). \end{aligned} \tag{2.22}$$

Γ_5 is a Lorentz pseudoscalar, which characterizes the irreducible representations of $SO(3, 1)$. It is the generalization of γ_5 of Dirac's[30] and Kemmer's[34] theories.

With the help of Γ_A we construct another antisymmetric tensor

$$\Theta_{AB} = -\Theta^{\dagger}_{AB} = \lambda_1[\Gamma_B, \Gamma_A] = [\lambda_1/(\lambda_1 + 1)]\Gamma_C S_{ABC}. \tag{2.23}$$

In the four-dimensional notation

$$\begin{aligned} g_\mu &= -g^{\dagger}_\mu = \Theta_{\mu 5} \\ &= -[\lambda_1/(\lambda_1 + 1)]\tilde{S}_{\mu\nu}\Gamma_\nu = \lambda_1[\Gamma_5, \Gamma_\mu], \\ \Theta_{\mu\nu} &= [\lambda_1/(\lambda_1 + 1)](\Gamma_5\tilde{S}_{\mu\nu} + \epsilon_{\mu\nu\alpha\beta}\Gamma_\alpha\gamma_\beta). \end{aligned} \tag{2.24}$$

One verifies easily that

$$\Theta_{\mu\nu} = \lambda_1[g_\mu, \gamma_\nu]. \tag{2.25}$$

[34] J. Beckers, Physica **28**, 1019 (1962).

$SO(5)$ possesses two invariants, a quartic operator D_1 and the Casimir operator D_2, given by

$$D_1 = \Gamma_A\Gamma_A, \quad D_2 = (\tfrac{1}{2})S_{AB}S_{BA}. \tag{2.26}$$

For later use we quote the following identities:

$$\lambda_1 S_{ABD}S_{DC} + S_{ABC} = (\lambda_1 + 1)(\Gamma_A\delta_{BC} - \Gamma_B\delta_{AC}),$$
$$S_{AB}\Gamma_B = \Gamma_B S_{BA} = -(2/\lambda_1)\Gamma_A. \tag{2.27}$$

Further,

$$\Theta_{ABC} = (\tfrac{1}{2})\epsilon_{ABCDE}\Theta_{DE} = [\lambda_1/(\lambda_1 + 1)] \times (\Gamma_A S_{BC} + \Gamma_B S_{CA} + \Gamma_C S_{AB})$$
$$= [\lambda_1/(\lambda_1 + 1)](S_{BC}\Gamma_A + S_{CA}\Gamma_B + S_{AB}\Gamma_C)$$

from which we get

$$\Theta_{BA}\Gamma_B = \Gamma_B\Theta_{AB} = -\{1/[2(\lambda_1 + 1)]\}S_{ABC}\Theta_{BC}$$
$$= -\{1/[2(\lambda_1 + 1)]\}S_{BC}\Theta_{ABC}$$
$$= \{[(\lambda_1 + 1)(\lambda_1 + 2) + \lambda_2(\lambda_2 + 1)]/ \lambda_1(\lambda_1 + 1)^2]\}\Gamma_A. \tag{2.28}$$

In the last step, we have made use of the value

$$\lambda_1^2 D_2 = \lambda_1(\lambda_1 + 3) + \lambda_2(\lambda_2 + 1) \tag{2.29}$$

of the invariant D_2 in the irreducible representation $R_5(\lambda_1, \lambda_2)$.[2] For later use, we introduce further the four-pseudovector

$$\Delta_\mu = \lambda_1[\Gamma_5, g_\mu] = [\lambda_1/(\lambda_1 + 1)]g_\nu\tilde{S}_{\nu\mu}. \tag{2.30}$$

C. Inversions and Complex Conjugation

All transformations

$$X_{AB} = a_s \exp(i\pi\Sigma_{AB}) \tag{2.31}$$

are reflection operators, where

$$a_s = \exp(i\pi\lambda_1) = \begin{cases} \pm 1, & \text{for bosons,} \\ \pm i, & \text{for fermions.} \end{cases} \tag{2.32}$$

It was shown by Hepner[35] that the eigenvalues of any generator Σ_{AB} are $\lambda_1, \lambda_1 - 1, \cdots, -\lambda_1$. Hence

$$\exp(2i\pi\Sigma_{AB}) = a_s^2 I,$$

such that

$$X_{AB}^2 = I, \quad X_{AB}^\dagger = X_{AB}^{-1} = X_{AB}. \tag{2.33}$$

However, X_{AB} is not necessarily improper. All transformations considered in what follows satisfy (2.33).

1. Spatial Inversion

Under spatial inversion $x_k \to -x_k$, $x_4 \to x_4$, and $x_5 \to x_5$. The corresponding transformation in the spin space is[2]

$$\beta = \exp[i\pi\lambda_1(\gamma_0 - 1)], \tag{2.34}$$

such that

$$[\beta, \sigma_k] = [\beta, \gamma_0] = 0,$$
$$\beta\alpha_k\beta = -\alpha_k, \quad \beta\gamma_k\beta = -\gamma_k. \tag{2.35}$$

This is verified easily by the use of the expansion theorem

$$e^A B e^{-A} = \sum_{n=0}^{\infty} \frac{1}{n!}[A, B]_n, \tag{2.36}$$

where

$$[A, B]_n = [A, [A, B]_{n-1}], \quad [A, B]_0 = B.$$

2. Space–Time Reflections

The transformation

$$\rho_k = a_s \exp(-i\pi\lambda_1\alpha_k) \tag{2.37}$$

corresponds to $x_k \to -x_k$, $x_j \to x_j$ $(j \neq k)$, $x_4 \to -x_4$, and $x_5 \to x_5$, such that

$$[\rho_k, \gamma_j] = 0 \quad \text{for} \quad k \neq j, \quad [\rho_k, \sigma_k] = 0,$$
$$\rho_k\gamma_k\rho_k = -\gamma_k, \quad \rho_k\gamma_0\rho_k = -\gamma_0,$$
$$[\rho_k, \alpha_k] = 0, \quad \rho_k\alpha_j\rho_k = -\alpha_j \quad \text{for} \quad k \neq j.$$

We notice that

$$\rho_k\rho_j = b_s\rho_j\rho_k \quad (k \neq j), \tag{2.38}$$

where

$$b_s = a_s^2 = \exp(2i\pi\lambda_1) = \begin{cases} +1 & \text{for bosons} \\ -1 & \text{for fermions.} \end{cases} \tag{2.39}$$

3. Space–Time Reversal

The transformation

$$\rho = a_s\rho_1\rho_2\rho_3 \tag{2.40}$$

corresponds to space–time reversal $x_\mu \to -x_\mu$ and $x_5 \to x_5$, such that

$$[S_{\mu\nu}, \rho] = 0, \quad \rho\gamma_\mu\rho = -\gamma_\mu,$$
$$\beta\rho = b_s\rho\beta. \tag{2.41}$$

4. Pauli's Time Reversal

The transformation

$$\tau = a_s\beta\rho \tag{2.42}$$

corresponds to $x_4 \to -x_4$ and $x_A \to x_A$, $A \neq 4$, such that

$$[\tau, \gamma_k] = [\tau, \sigma_k] = 0,$$
$$\tau\gamma_0\tau = -\gamma_0, \quad \tau\alpha_k\tau = -\alpha_k. \tag{2.43}$$

We have also

$$\tau\beta = b_s\beta\tau = a_s b_s\rho,$$
$$\rho\tau = b_s\tau\rho = a_s b_s\beta. \tag{2.44}$$

τ, β, and ρ commute (anticommute) with each other for bosons (fermions).

[35] W. A. Hepner, Phys. Rev. **84**, 744 (1951).

5. Complex Conjugation

As is well known,[36] a complex-conjugation matrix C exists with the properties

$$C^\dagger = C^{-1}, \quad C^{\mathrm{T}} = b_s C, \tag{2.45}$$

such that

$$\begin{aligned} S^*_{AB} &= -S^{\mathrm{T}}_{AB} = C^\dagger S_{AB} C, \\ \beta^* &= \beta^{\mathrm{T}} = b_s C^\dagger \beta C, \\ \rho^* &= \rho^{\mathrm{T}} = C^\dagger \rho C, \end{aligned} \tag{2.46}$$

$$\tau^* = \tau^{\mathrm{T}} = C^\dagger \tau C \tag{2.47}$$

and

$$\begin{aligned} \sigma^*_k &= \sigma^{\mathrm{T}}_k = -C^\dagger \sigma_k C, \\ \alpha^*_k &= \alpha^{\mathrm{T}}_k = -C^\dagger \alpha_k C, \\ \gamma^*_k &= -\gamma^{\mathrm{T}}_k = C^\dagger \gamma_k C, \\ \gamma^*_0 &= \gamma^{\mathrm{T}}_0 = -C^\dagger \gamma_0 C. \end{aligned} \tag{2.48}$$

The choice of C depends on the particular representation. We consider three representations:

(a) γ_2, σ_2, and α_2 are pure imaginary and all other γ_k, σ_k, α_k, and γ_0 are real. Then

$$\beta^* = \beta, \quad \rho^* = \rho, \quad \tau^* = b_s \tau. \tag{2.49}$$

It can be easily verified that

$$C = a_s \rho \exp\{-i\pi\lambda_1\sigma_2\}, \quad C^2 = I. \tag{2.50}$$

Using $\sigma_2^{\mathrm{T}} = \sigma_2^* = -\sigma_2$ and $\rho^{\mathrm{T}} = \rho^* = \rho$, we get $C^{\mathrm{T}} = b_s C$. Hence this representation exists for bosons and fermions.

(b) The Majorana representation: γ_k and γ_0 are pure imaginary, such that α_k is real, σ_k is pure imaginary, and

$$\beta^* = b_s \beta, \quad \rho^* = b_s \rho, \quad \tau^* = b_s \tau. \tag{2.51}$$

Hence,

$$C = \beta. \tag{2.52}$$

Since $\beta^{\mathrm{T}} = \beta^* = b_s \beta$, we get $C^{\mathrm{T}} = b_s C$. Hence, this representation exists for bosons and fermions.

(c) γ_k and γ_0 are real, such that α_k is real, σ_k is pure imaginary, and

$$\beta^* = \beta, \quad \rho^* = b_s \rho, \quad \tau^* = \tau. \tag{2.53}$$

Hence,

$$C = \tau. \tag{2.54}$$

Now, since $\tau^{\mathrm{T}} = \tau^* = \tau$, we have $C^{\mathrm{T}} = C$, i.e., $b_s = 1$. Thus this representation exists only for bosons. In this representation the wavefunction in momentum space is real, apart from an irrelevant phase factor.

[36] Confer L. C. Biedenharn, J. Nuyts, and H. Ruegg, *On the Generalization of Isoparity*, CERN Report No. 65-3, Geneva, 1965, pp. 14 and 15.

D. Covariance of the Field Equations

1. Linear Transformations

Equations (1.1) and (1.2) may be written in the manifestly covariant form

$$P_A \psi \equiv S_{AB} p_B \psi = p_A \psi, \tag{2.55}$$

where $P_5 = -\gamma_\mu p_\mu$, $m \to p_5$. Under the linear transformation (2.1) and (2.6) in five dimensions, we get

$$S^{-1} P_A(p'_C) S = a_{AB} P_B(p_C), \tag{2.56}$$

from which the covariance of (2.55) follows immediately. These linear transformations include four-dimensional rotations and inversions, by fixing $p_5 = m$. The proper Lorentz transformation with velocity $\mathbf{v}$ is

$$L = \exp\{(\lambda_1/v)\boldsymbol{\alpha} \cdot \mathbf{v} \tanh^{-1} v\}. \tag{2.57}$$

The parity

$$P = \beta \Pi_r \tag{2.58}$$

and Pauli's time reversal

$$T_p = \tau \Pi_t \tag{2.59}$$

are good quantum numbers, where

$$\Pi_r \psi(\mathbf{r}, t) = \psi(-\mathbf{r}, t) \tag{2.60}$$

and

$$\Pi_t \psi(\mathbf{r}, t) = \psi(\mathbf{r}, -t). \tag{2.61}$$

Under infinitesimal five-dimensional rotations

$$\begin{aligned} \mathbf{p}' &= \mathbf{p} + \boldsymbol{\theta} \wedge \mathbf{p} + \mathbf{v} p_0 + \mathbf{a} p_5, \\ p'_0 &= p_0 + \mathbf{v} \cdot \mathbf{p} + a_0 p_5, \\ p'_5 &= p_5 - \mathbf{a} \cdot \mathbf{p} + a_0 p_0, \end{aligned} \tag{2.62}$$

$$S = I + \lambda_1(i\boldsymbol{\sigma} \cdot \boldsymbol{\theta} + \boldsymbol{\alpha} \cdot \mathbf{v} - \boldsymbol{\gamma} \cdot \mathbf{a} + \gamma_0 a_0), \tag{2.63}$$

where $\boldsymbol{\theta}$, $\mathbf{v}$, $\mathbf{a}$, and a_0 are all real. Then

$$\tau S^\dagger \tau = S^{-1}. \tag{2.64}$$

We notice that S is invariant under both P and T_P. For the four-dimensional rotations ($a_\mu = 0$) we have, in addition,

$$\beta S^\dagger \beta = S^{-1}. \tag{2.65}$$

We denote in the usual way

$$\bar{\psi} = \psi^\dagger \beta. \tag{2.66}$$

Then, $\bar{\psi}\rho\psi$ and $\bar{\psi}\rho\Gamma_A\psi$ are the only five-dimensional scalars and vectors, respectively. The latter decomposes into a four-dimensional scalar $\bar{\psi}\rho\Gamma_5\psi$ and a four-vector $\bar{\psi}\rho\Gamma_\mu\psi$. However, under spatial inversion they have opposite parities.

In fact, denoting

$$s_A = \bar{\psi}\rho\Gamma_A\psi, \tag{2.67}$$

then, under spatial inversion,

$$\begin{aligned} s_k &\to s_k' = b_s s_k(-\mathbf{r}, t), \\ s_0 &\to s_0' = -b_s s_0(-\mathbf{r}, t), \\ s_5 &\to s_5' = -b_s s_5(-\mathbf{r}, t). \end{aligned} \tag{2.68}$$

Hence for fermions ($b_s = -1$), s_A is a five-vector, while for bosons ($b_s = +1$), it is a pseudovector. Also $\bar{\psi}\rho S_{AB}\psi$ is a five-dimensional antisymmetric tensor for bosons and pseudotensor for fermions. Further, in favor of (2.65), $\bar{\psi}\psi$, $\bar{\psi}\Gamma_5\psi$, $\bar{\psi}\gamma_\mu\psi$ and $\bar{\psi}\Gamma_\mu\psi$ are four-dimensional scalar, vector and pseudovector, respectively. Also $\bar{\psi}S_{\mu\nu}\psi$ is an antisymmetric tensor.

2. Charge Conjugation

The charge conjugation operator is antilinear,

$$\mathcal{C} = C_1\Lambda, \quad \mathcal{C}^\dagger = \mathcal{C}^{-1}, \tag{2.69}$$

where Λ is the complex conjugation operator

$$\Lambda\psi = \psi^* \tag{2.70}$$

and

$$C_1 = \beta C = C_1^{\mathrm{T}} = C_1^\dagger = C_1^{-1}. \tag{2.71}$$

Under charge conjugation, the five-dimensional transformation (2.63) is not invariant, while the four-dimensional transformation ($a_\mu = 0$) is invariant. Hence

$$\psi^{\mathcal{C}} = \mathcal{C}\psi = \beta C\psi^* \tag{2.72}$$

has the same four-dimensional transformation law as ψ. We notice that

$$\mathcal{C}^\dagger P\mathcal{C} = b_s P, \quad \mathcal{C}^\dagger T_P\mathcal{C} = b_s T_P. \tag{2.73}$$

Hence ψ and $\psi^{\mathcal{C}}$ have the same (opposite) parity for bosons (fermions). We notice also that

$$PT_P = b_s T_P P. \tag{2.74}$$

In configuration space,

$$\mathbf{p} \to -i\boldsymbol{\nabla}, \quad p_0 \to i\partial_t, \quad p_5 \to m, \tag{2.75}$$

and

$$\begin{aligned} \mathbf{P} &= \boldsymbol{\alpha}p_0 + i\boldsymbol{\sigma}\wedge\mathbf{p} + m\boldsymbol{\gamma} \to i\boldsymbol{\alpha}\partial_t + \boldsymbol{\sigma}\wedge\boldsymbol{\nabla} + m\boldsymbol{\gamma}, \\ P_0 &= \boldsymbol{\alpha}\cdot\mathbf{p} + m\gamma_0 \to -i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + m\gamma_0, \\ P_5 &= \gamma_0 p_0 - \boldsymbol{\gamma}\cdot\mathbf{p} \to i(\gamma_0\partial_t + \boldsymbol{\gamma}\cdot\boldsymbol{\nabla}). \end{aligned} \tag{2.76}$$

Then,

$$\begin{aligned} &\mathcal{C}^\dagger\mathbf{P}\mathcal{C} = -\mathbf{P}, \quad \mathcal{C}^\dagger\mathbf{p}\mathcal{C} = -\mathbf{p}, \\ &\mathcal{C}^\dagger P_0\mathcal{C} = -P_0, \quad \mathcal{C}^\dagger p_0\mathcal{C} = -p_0, \\ &\mathcal{C}^\dagger P_5\mathcal{C} = P_5, \quad \mathcal{C}^\dagger p_5\mathcal{C} = p_5, \end{aligned} \tag{2.77}$$

where m is assumed real. This proves the invariance of the field equations under charge conjugation. Obviously the equations are also invariant under Wigner's time reversal

$$T_W = \mathcal{C}T_P = \rho C\Lambda\Pi_t \tag{2.78}$$

and

$$\mathcal{C}PT_W = \rho\Pi_r\Pi_t. \tag{2.79}$$

3. BHABHA REPRESENTATION

A. Spinor Formulation

We follow Bhabha's formalism[3] of the generators of $SO(4, 1)$ with the necessary modification due to our use of a pseudo-Euclidean metric. Consider Pauli's matrices

$$\begin{aligned} s_1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad s_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ s_3 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad s_4 = \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}. \end{aligned} \tag{3.1}$$

Let $s_\mu^{a\dot{b}}$ be the element of the matrix s_μ in the row a and the column b, and let

$$s_{k\dot{a}b} = -s_k^{a\dot{b}}, \quad s_{4\dot{a}b} = s_4^{a\dot{b}}. \tag{3.2}$$

Then,

$$s_\mu^{a\dot{b}}s_\mu^{c\dot{d}} = -2\epsilon^{ac}\epsilon^{\dot{b}\dot{d}}, \quad s_\mu^{a\dot{b}}s_{\nu\dot{b}a} = -2\delta_{\mu\nu}, \tag{3.3}$$

where $\epsilon^{ab} = \epsilon_{\dot{a}\dot{b}} = -\epsilon_{\dot{b}\dot{a}} = -\epsilon^{\dot{b}\dot{a}}$, $\epsilon^{12} = 1$. In this section the indices a, b, c, $\cdots$ are spinor indices, running over 1, 2; while k, l, m, $\cdots$ are Cartesian indices running over 1, 2, 3. Further let

$$\begin{aligned} \Sigma_b^{(1)a} &= -(\lambda_1/4)s_\mu^{a\dot{c}}s_{\nu\dot{c}b}S_{\mu\nu} = -(\lambda_1/4)s_\mu^{a\dot{c}}s_{\nu\dot{c}b}\tilde{S}_{\mu\nu}, \\ \Sigma_{\dot{b}}^{(2)\dot{a}} &= -(\lambda_1/4)s_\mu^{c\dot{a}}s_{\nu\dot{b}c}S_{\mu\nu} = (\lambda_1/4)s_\mu^{c\dot{a}}s_{\nu\dot{b}c}\tilde{S}_{\mu\nu}, \end{aligned} \tag{3.4}$$

such that

$$\begin{aligned} S_{\mu\nu} &= -(1/2\lambda_1)s_\mu^{a\dot{b}}s_\nu^{c\dot{d}}(\epsilon_{\dot{b}\dot{d}}\Sigma_{ac}^{(1)} + \epsilon_{ac}\Sigma_{\dot{b}\dot{d}}^{(2)}), \\ \tilde{S}_{\mu\nu} &= -(1/2\lambda_1)s_\mu^{a\dot{b}}s_\nu^{c\dot{d}}(\epsilon_{\dot{b}\dot{d}}\Sigma_{ac}^{(1)} - \epsilon_{ac}\Sigma_{\dot{b}\dot{d}}^{(2)}). \end{aligned} \tag{3.5}$$

Here $\Sigma_{cb}^{(1)}$ and $\Sigma_{\dot{a}\dot{b}}^{(2)}$ are symmetric spinors obtained by lowering the indices:

$$\begin{aligned} \Sigma_{ac}^{(1)} &= \epsilon_{ab}\Sigma_c^{(1)b} = \Sigma_{ca}^{(1)}, \\ \Sigma^{(1)ac} &= \epsilon^{ba}\Sigma_b^{(1)c} = \Sigma^{(1)ca}. \end{aligned} \tag{3.6}$$

Further we denote

$$\begin{aligned} &\Sigma_1^{(1)} + i\Sigma_2^{(1)} = \Sigma_1^{(1)2}, \quad \Sigma_1^{(1)} - i\Sigma_2^{(1)} = \Sigma_2^{(1)1}, \\ &\Sigma_1^{(2)} + i\Sigma_2^{(2)} = -\Sigma_2^{(2)1}, \quad \Sigma_1^{(2)} - i\Sigma_2^{(2)} = -\Sigma_1^{(2)2}, \\ &\Sigma_3^{(1)} = \Sigma_1^{(1)1} = -\Sigma_2^{(1)2}, \quad \Sigma_3^{(2)} = -\Sigma_1^{(2)1} = \Sigma_2^{(2)2}. \end{aligned} \tag{3.7}$$

$\boldsymbol{\Sigma}^{(1)} = (\Sigma_1^{(1)}, \Sigma_2^{(1)}, \Sigma_3^{(1)})$ and $\boldsymbol{\Sigma}^{(2)} = (\Sigma_1^{(2)}, \Sigma_2^{(2)}, \Sigma_3^{(2)})$ are the generators of two $SO(3)$ subgroups,

$$[\Sigma_k^{(s)}, \Sigma_l^{(s')}] = i\delta_{ss'}\epsilon_{kln}\Sigma_n^{(s)}, \quad s, s' = 1, 2. \tag{3.8}$$

$SO(3, 1)$ is isomorphic to the direct product of these two groups. In fact,

$$\begin{aligned} \lambda_1\boldsymbol{\sigma} &= \boldsymbol{\Sigma}^{(1)} + \boldsymbol{\Sigma}^{(2)}, \\ \lambda_1\boldsymbol{\alpha} &= \boldsymbol{\Sigma}^{(1)} - \boldsymbol{\Sigma}^{(2)}. \end{aligned} \tag{3.9}$$

The finite-dimensional irreducible representations $D(s_1, s_2)$ of $SO(3, 1)$ are characterized by the two invariants

$$\mathbf{\Sigma}^{(1)^2} = s_1(s_1 + 1)I, \quad \mathbf{\Sigma}^{(2)^2} = s_2(s_2 + 1)I. \quad (3.10)$$

s_1 and s_2 are both integers or half-integers for bosons, while for fermions one of them is an integer and the other is a half-integer; both are nonnegative. We denote the matrix elements of an operator A, connecting two irreducible representations $D(s_1, s_2)$ and $D(s_1', s_2')$, by $\langle s_1, s_2| A |s_1', s_2'\rangle$. Also the Poincaré component of ψ, transforming according to $D(s_1, s_2)$, is denoted by $\psi(s_1, s_2)$. Then, in the irreducible representation $R_5(\lambda_1, \lambda_2)$ of $SO(4, 1)$, the Poincaré component of $A\psi$ is given by

$$(A\psi)(s_1, s_2) = \sum_{s_1', s_2'} \langle s_1, s_2| A |s_1', s_2'\rangle \psi(s_1', s_2'), \quad (3.11)$$

where

$$\delta \leq |s_1' - s_2'| \leq \lambda_2 \leq s_1' + s_2' \leq \lambda_1, \quad (3.12)$$

and $\delta = 0$ for bosons and $\delta = \frac{1}{2}$ for fermions. In this representation

$$\langle s_1, s_2| \mathbf{\Sigma}^{(1)^2} |s_1', s_2'\rangle = \delta_{s_1 s_1'}\delta_{s_2 s_2'} s_1(s_1 + 1)I,$$
$$\langle s_1, s_2| \mathbf{\Sigma}^{(2)^2} |s_1', s_2'\rangle = \delta_{s_1 s_1'}\delta_{s_2 s_2'} s_2(s_2 + 1)I, \quad (3.13)$$

such that

$$\langle s_1, s_2| S_{\mu\nu}S_{\nu\mu} |s_1', s_2'\rangle = \delta_{s_1 s_1'}\delta_{s_2 s_2'}(4/\lambda_1^2)[s_1(s_1 + 1) + s_2(s_2 + 1)]I \quad (3.14)$$

and [cf. (2.22)]:

$$\lambda_1(\lambda_1 + 1)\langle s_1, s_2| \Gamma_5 |s_1', s_2'\rangle = \langle s_1, s_2| \mathbf{\Sigma}^{(1)^2} - \mathbf{\Sigma}^{(2)^2} |s_1', s_2'\rangle = \delta_{s_1 s_1'}\delta_{s_2 s_2'}(s_1 - s_2)(s_1 + s_2 + 1)I. \quad (3.15)$$

Here we have suppressed two additional indices, which characterize the representation completely. The two indices may characterize the connection between the eigenvalues of each of $\Sigma_3^{(1)}$ and $\Sigma_3^{(2)}$. This is Bhabha's[3] direct product representation $O(3, 1) = O(3) \times \bar{O}(3)$; the matrix elements are

$$\langle s_1, m_1; s_2, m_2| A |s_1', m_1'; s_2', m_2'\rangle.$$

In this representation

$$\langle s_1, m_1; s_2, m_2| \Sigma_3^{(1)} |s_1', m_1'; s_2', m_2'\rangle = m_1\delta_{s_1 s_1'}\delta_{s_2 s_2'}\delta_{m_1 m_1'}\delta_{m_2 m_2'},$$
$$\langle s_1, m_1; s_2, m_2| \Sigma_3^{(2)} |s_1', m_1'; s_2', m_2'\rangle = m_2\delta_{s_1 s_1'}\delta_{s_2 s_2'}\delta_{m_1 m_1'}\delta_{m_2 m_2'},$$
$$|m_1| \leq s_1, \quad |m_2| \leq s_2. \quad (3.16)$$

We may also consider the representation in which $\boldsymbol{\sigma}^2$ and σ_3 are diagonal. This representation was studied by Wild[37] and Le Couteur.[38,39] The unitary transformation connecting these two representations was derived by Wild. In the latter representation the matrix elements are denoted by

$$\langle s_1, s_2, s, m| A |s_1', s_2', s', m'\rangle$$

such that

$$\langle s_1, s_2, s, m| \boldsymbol{\sigma}^2 |s_1', s_2', s', m'\rangle = [s(s + 1)/\lambda_1^2]\delta_{s_1 s_1'}\delta_{s_2 s_2'}\delta_{ss'}\delta_{mm'},$$
$$\langle s_1, s_2, s, m| \sigma_3 |s_1', s_2', s', m'\rangle = (m/\lambda_1)\delta_{s_1 s_1'}\delta_{s_2 s_2'}\delta_{ss'}\delta_{mm'},$$
$$|s_1 - s_2| \leq s \leq s_1 + s_2, \quad |m| \leq s. \quad (3.17)$$

Further denote

$$\gamma^{a\dot{b}} = \lambda_1 s_\mu^{a\dot{b}}\gamma_\mu, \quad (3.18)$$

such that

$$\gamma_\mu = -(1/2\lambda_1)s_{\mu\dot{b}a}\gamma^{a\dot{b}}. \quad (3.19)$$

The spinors $\Gamma^{a\dot{b}}$, $g^{a\dot{b}}$, and $\Delta^{a\dot{b}}$ are defined similarly for Γ_μ, g_μ, and Δ_μ, respectively. Then, from (2.18), (2.24), and (2.30), we get

$$\Gamma^{a\dot{b}} = \lambda_1[\Gamma_5, \gamma^{a\dot{b}}], \quad (3.20)$$
$$g^{a\dot{b}} = \lambda_1[\Gamma_5, \Gamma^{a\dot{b}}], \quad (3.21)$$
$$\Delta^{a\dot{b}} = \lambda_1[\Gamma_5, g^{a\dot{b}}]. \quad (3.22)$$

Bhabha[3] has shown that the nonvanishing matrix elements of the γ's are those connecting two consecutive representations of $SO(3, 1)$, such that $|s_1 - s_1'| = |s_2 - s_2'| = \frac{1}{2}$. From (3.15) and (3.20)–(3.22), it follows that the nonvanishing matrix elements of $\Gamma^{a\dot{b}}$, $g^{a\dot{b}}$, and $\Delta^{a\dot{b}}$ are

$$\langle s_1, s_2| \Gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 - \frac{\sigma}{2} \right\rangle = \sigma x \langle s_1, s_2| \gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 - \frac{\sigma}{2} \right\rangle,$$
$$\langle s_1, s_2| \Gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 + \frac{\sigma}{2} \right\rangle = \sigma y \langle s_1, s_2| \gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 + \frac{\sigma}{2} \right\rangle, \quad (3.23)$$
$$\langle s_1, s_2| g^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 - \frac{\sigma}{2} \right\rangle = x^2 \langle s_1, s_2| \gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 - \frac{\sigma}{2} \right\rangle,$$
$$\langle s_1, s_2| g^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 + \frac{\sigma}{2} \right\rangle = y^2 \langle s_1, s_2| \gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 + \frac{\sigma}{2} \right\rangle, \quad (3.24)$$

[37] C. Wild, Proc. Roy. Soc. (London) **A191**, 253 (1947).
[38] K. J. Le Couteur, Proc. Roy. Soc. (London) **A202**, 284 (1950).
[39] K. J. Le Couteur, Proc. Roy. Soc. (London) **A202**, 394 (1950).

$$\langle s_1, s_2|\, \Delta^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 - \frac{\sigma}{2} \right\rangle = \sigma x^3 \langle s_1, s_2|\, \gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 - \frac{\sigma}{2} \right\rangle,$$

$$\langle s_1, s_2|\, \Delta^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 + \frac{\sigma}{2} \right\rangle = \sigma y^3 \langle s_1, s_2|\, \gamma^{a\dot{b}} \left| s_1 - \frac{\sigma}{2}, s_2 + \frac{\sigma}{2} \right\rangle, \quad (3.25)$$

where

$$\sigma = \pm 1, \quad x = \mu/(\lambda_1 + 1), \quad y = (\lambda + 1)/(\lambda_1 + 1),$$

and

$$\mu = s_1 - s_2, \quad \lambda = s_1 + s_2, \quad (3.26)$$

such that, [cf. (3.12)],

$$\delta \leq |\mu| \leq \lambda_2 \leq \lambda \leq \lambda_1. \quad (3.12')$$

B. The Proportionality Coefficients

Bhabha[3] has shown that the nonvanishing matrix elements of $\gamma^{a\dot{b}}$ are

$$\langle s_1 s_2|\, \gamma^{a\dot{b}} \,|s_1 - \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle = i d_1 (c_{--})^{\frac{1}{2}} u^a(s_1) u^{\dot{b}}(s_2),$$

$$\langle s_1 - \tfrac{1}{2}, s_2 - \tfrac{1}{2}|\, \gamma^{a\dot{b}} \,|s_1, s_2\rangle = -i d_1^{-1} (c_{--})^{\frac{1}{2}} v^a(s_1) v^{\dot{b}}(s_2),$$

$$\langle s_1, s_2|\, \gamma^{a\dot{b}} \,|s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle = d_2 (c_{+-})^{\frac{1}{2}} v^a(s_1 + \tfrac{1}{2}) u^{\dot{b}}(s_2),$$

$$\langle s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}|\, \gamma^{a\dot{b}} \,|s_1, s_2\rangle = -d_2^{-1} (c_{+-})^{\frac{1}{2}} u^a(s_1 + \tfrac{1}{2}) v^{\dot{b}}(s_2). \quad (3.27)$$

d_1 and d_2 can be chosen arbitrarily, but they do not assume zero values. c_{--} and c_{+-} depend on s_1 and s_2, but not on m_1 and m_2 (or s and m). They are also independent of the spinor indices a and $\dot{b}$. The Dirac matrices[6] $u^a(s)$ and $v^a(s)$ of $(2s + 1)$ rows and $2s$ columns, and $2s$ rows and $(2s + 1)$ columns, respectively. The multiplication $u^a(s_1)v^{\dot{b}}(s_2)$ implies the direct product of the two matrices. These matrices satisfy the following relations:

$$v_a(s)u^a(s) = -u_a(s + \tfrac{1}{2})v^a(s + \tfrac{1}{2}) = 2s + 1,$$
$$v_a(s)v^a(s + \tfrac{1}{2}) = u_a(s + \tfrac{1}{2})u^a(s) = 0, \quad (3.28)$$

$$\Sigma_b^{(1)a}(s) = -u^a(s)v_b(s) + s\delta_b^a = -v^a(s + \tfrac{1}{2})u_b(s + \tfrac{1}{2}) - (s + 1)\delta_b^a, \quad (3.29)$$

and similar relations for $\Sigma_{\dot{b}}^{(2)\dot{a}}$. These matrices differ from Dirac's original matrices by a trivial factor $(-1)^{s+\frac{1}{2}}$, which can be absorbed into d_1 and d_2. Equations (3.28) and (3.29) are identical with Bhabha's assignments.

Le Couteur[39] has shown that for Bhabha's matrices satisfying (2.9) and (2.10), c_{--} and c_{+-} satisfy the following relations:

$$(s_1 + 1)(c_{+-} - c_{++}) - s_1(c_{-+} - c_{--}) = 1,$$
$$(s_2 + 1)(c_{-+} - c_{++}) - s_2(c_{+-} - c_{--}) = 1, \quad (3.30)$$

and

$$s_1 \left[\frac{c_{++}(s_1 - \frac{1}{2}, s_2 + \frac{1}{2})}{c_{++}(s_1, s_2)}\right]^{\frac{1}{2}} = (s_1 + 1)\left[\frac{c_{-+}(s_1 + \frac{1}{2}, s_2 + \frac{1}{2})}{c_{-+}(s_1, s_2)}\right]^{\frac{1}{2}},$$

$$s_2 \left[\frac{c_{++}(s_1 + \frac{1}{2}, s_2 - \frac{1}{2})}{c_{++}(s_1, s_2)}\right]^{\frac{1}{2}} = (s_2 + 1)\left[\frac{c_{+-}(s_1 + \frac{1}{2}, s_2 + \frac{1}{2})}{c_{+-}(s_1, s_2)}\right]^{\frac{1}{2}}. \quad (3.31)$$

The latter relations imply the vanishing of the matrix elements

$$\langle s_1, s_2|\, [\gamma_\mu, \gamma_\nu] \,|s_1 + 1, s_2\rangle$$

and

$$\langle s_1, s_2|\, [\gamma_\mu, \gamma_\nu] \,|s_1, s_2 + 1\rangle.$$

Here we have denoted

$$c_{--} = c_{--}(s_1, s_2), \quad c_{++} = c_{--}(s_1 + \tfrac{1}{2}, s_2 + \tfrac{1}{2}),$$
$$c_{+-} = c_{+-}(s_1, s_2), \quad c_{-+} = c_{+-}(s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2}). \quad (3.32)$$

Equations (3.30) are two difference equations for c_{--} and c_{+-}. Instead of solving these equations, we evaluate the four constants c_{--}, c_{++}, c_{+-}, and c_{-+} by solving these two equations simultaneously with the two equations defining the two invariants D_1 and D_2 of the group. From (3.14) and (3.15) we get

$$\lambda_1^2(\lambda_1 + 1)^2 D_1 = \lambda_1^2(\lambda_1 + 1)^2 \langle s_1, s_2|\, \Gamma_A \Gamma_A \,|s_1, s_2\rangle = \mu^2(\lambda + 1)^2 - \tfrac{1}{2}(\lambda_1 + 1)^2 \epsilon_{ac}\epsilon_{\dot{b}\dot{d}} \langle s_1, s_2|\, \Gamma^{a\dot{b}}\Gamma^{c\dot{d}} \,|s_1, s_2\rangle \quad (3.33)$$

and

$$\lambda_1^2 D_2 = (\lambda_1^2/2) \langle s_1, s_2|\, S_{AB}S_{BA} \,|s_1, s_2\rangle = \lambda(\lambda + 2) + \mu^2 + \tfrac{1}{2}\epsilon_{ac}\epsilon_{\dot{b}\dot{d}} \langle s_1, s_2|\, \gamma^{a\dot{b}}\gamma^{c\dot{d}} \,|s_1, s_2\rangle, \quad (3.34)$$

where λ and μ are given by (3.26). Using (3.23), (3.27)–(3.29), (3.33), and (3.34), one gets after some manipulations:

$$s_1 s_2 c_{--} + (s_1 + 1)(s_2 + 1)c_{++} + s_2(s_1 + 1)c_{+-} + s_1(s_2 + 1)c_{-+} = \tfrac{1}{2}\lambda_1^2 D_2 - \tfrac{1}{2}[\lambda(\lambda + 2) + \mu^2], \quad (3.35)$$

and

$$\mu^2[s_1s_2c_{--} + (s_1+1)(s_2+1)c_{++}] + (\lambda+1)^2[s_2(s_1+1)c_{+-} + s_1(s_2+1)c_{-+}] = \tfrac{1}{2}\lambda_1^2(\lambda_1+1)^2D_1 - \tfrac{1}{2}\mu^2(\lambda+1)^2. \quad (3.36)$$

Solving (3.30), (3.35), and (3.36) for c_{--} and c_{++}, using[2]

$$\lambda_1^2 D_2 = \lambda_1(\lambda_1+3) + \lambda_2(\lambda_2+1), \quad (3.37)$$

we get

$$4(s_1+1)(s_2+1)(2s_1+1)(2s_2+1)c_{++} + \lambda_1^2(\lambda_1+1)^2D_1 = (\lambda+1)(\lambda+2)[\lambda_1(\lambda_1+3) - \lambda(\lambda+3) + \lambda_2(\lambda_2+1)]. \quad (3.38)$$

Now, since the maximum value of λ is λ_1 [cf. Eq. (3.12a)], the matrix element of $\gamma^{a\dot{b}}$, connecting $D(s_1, s_2)$ to $D(s_1+\frac{1}{2}, s_2+\frac{1}{2})$, should vanish for the finite-dimensional representations. Thus $c_{++} = 0$ for $\lambda = \lambda_1$. This gives the value

$$D_1 = \lambda_2(\lambda_2+1)(\lambda_1+2)/\lambda_1^2(\lambda_1+1). \quad (3.39)$$

Hence,

$$4s_1s_2(2s_1+1)(2s_2+1)c_{--} = (\lambda-\lambda_2)(\lambda+\lambda_2+1)(\lambda_1-\lambda+1)(\lambda_1+\lambda+2),$$
$$4(s_1+1)(s_2+1)(2s_1+1)(2s_2+1)c_{++} = (\lambda-\lambda_2+1)(\lambda+\lambda_2+2)(\lambda_1-\lambda)(\lambda_1+\lambda+3). \quad (3.40)$$

Solving (3.30), (3.35), and (3.36) for c_{+-} and c_{-+}, using (3.37) and (3.39), we get

$$4s_2(s_1+1)(2s_1+1)(2s_2+1)c_{+-} = (\lambda_2-\mu)(\lambda_2+\mu+1)(\lambda_1-\mu+1)(\lambda_1+\mu+2),$$
$$4s_1(s_2+1)(2s_1+1)(2s_2+1)c_{-+} = (\lambda_2+\mu)(\lambda_2-\mu+1)(\lambda_1-\mu+2)(\lambda_1+\mu+1). \quad (3.41)$$

One easily verifies that the values (3.40) and (3.41) are consistent with (3.31), as should be the case.

We consider two particular representations:

(1) $R_5(\lambda_1, \lambda_1)$. In this representation

$$\lambda_1 = \lambda_2 = \lambda = s_1 + s_2 \quad (3.42)$$

and

$$c_{+-} = c_{-+} = 1, \quad c_{++} = c_{--} = 0. \quad (3.43)$$

These coefficients were obtained by Bhabha.[3] In this case

$$R_5(\lambda_1, \lambda_1) = D(\lambda_1, 0) \oplus D(\lambda_1 - \tfrac{1}{2}, \tfrac{1}{2}) \oplus \cdots \oplus D(\tfrac{1}{2}, \lambda_1 - \tfrac{1}{2}) \oplus D(0, \lambda_1). \quad (3.44)$$

From (3.23), (3.24), and (3.25) we get

$$g_\mu = \gamma_\mu, \quad \Delta_\mu = \Gamma_\mu. \quad (3.45)$$

From (2.25) we get

$$\Theta_{AB} = S_{AB}. \quad (3.46)$$

Denoting $S_{A6} = -S_{6A} = \pm i\Gamma_A$, we see that the S_{AB} are the generators of $SO(6)$, where A and B run over the six values of the indices. In fact, in favor of (2.16), (2.23), and (3.45), one verifies that $\Sigma_{AB} = i\lambda_1 S_{AB}$ satisfies the integrability conditions (2.5) of $SO(6)$. Hence, $R_5(\lambda_1, \lambda_1)$ is a particular representation of $SO(6)$. Dirac's electron theory ($\lambda_1 = \lambda_2 = \frac{1}{2}$) and Kemmer's vector-meson theory ($\lambda_1 = \lambda_2 = 1$) belong to this class of representations.

(2) $R_5(\lambda_1, 0)$. In this representation $\lambda_2 = \mu = 0$:

$$s_1 = s_2 = \lambda/2, \quad 0 \le \lambda \le \lambda_1, \quad (3.47)$$

$$c_{+-} = c_{-+} = 0, \quad (3.48)$$

$$c_{--}(\tfrac{1}{2}\lambda, \tfrac{1}{2}\lambda) = (\lambda_1 - \lambda + 1)(\lambda_1 + \lambda + 2)/\lambda(\lambda+1),$$
$$c_{++}(\tfrac{1}{2}\lambda, \tfrac{1}{2}\lambda) = (\lambda_1 - \lambda)(\lambda_1 + \lambda + 3)/(\lambda+1)(\lambda+2). \quad (3.49)$$

This representation decomposes under $SO(3, 1)$ into

$$R_5(\lambda_1, 0) = D[(\lambda_1/2), (\lambda_1/2)] \oplus D\{[(\lambda_1 - 1)/2], [(\lambda_1 - 1)/2]\} \oplus \cdots \oplus D(0, 0). \quad (3.50)$$

From (3.23) and (3.48) we get

$$\Gamma_A = \Theta_{AB} = \Delta_\mu = D_1 = 0. \quad (3.51)$$

The dual equations (4.6) become (see Sec. 4):

$$\boldsymbol{\sigma} \cdot \mathbf{p}\psi = 0, \quad (\sigma p_0 + i\boldsymbol{\alpha} \wedge \mathbf{p})\psi = 0. \quad (3.52)$$

The particle is longitudinally polarized; its spin is $\lambda_2 = 0$. Kemmer's scalar-meson theory ($\lambda_1 = 1$, $\lambda_2 = 0$) belongs to this representation.

The representations $R_5(\lambda_1, \lambda_1)$ and $R_5(\lambda_1, 0)$ are the only ones with simple chains. All other representations $\lambda_1 \neq \lambda_2 \neq 0$ are mixtures of two chains $(s_1, s_2) \to (s_1 \pm \frac{1}{2}, s_2 \pm \frac{1}{2})$ and $(s_1, s_2) \to (s_1 \pm \frac{1}{2}, s_2 \mp \frac{1}{2})$.

The representations discussed here are the finite-dimensional representations of $SO(4, 1)$ and are all nonunitary. The other unitary representations were discussed by several authors.[40] The c coefficients for these representations were derived by Dixmier. His coefficients are identical with ours for the infinite tail of the Bhabha series. However, our derivation is much simpler, avoiding the solution of difference equations.

[40] L. H. Thomas, Ann. Math. **42**, 113 (1941); T. D. Newton, Ann. Math. **51**, 730 (1950); J. Dixmier, Bull. Soc. Math. France **89**, 9 (1961); R. Takahashi, Bull. Soc. Math. France **91**, 289 (1963); A. Kihlberg and S. Ström, Arkiv Fysik **31**, 491 (1966).

C. Reflections

The matrix elements of the spatial inversion operator β [cf. Eq. (2.34)] were evaluated by Wild[37] and Le Couteur.[38,39] However, they did not give the explicit form of the operator β in terms of the generators of $SO(5)$. Since their matrices satisfy $\beta^2 = 1$, our β is equal to theirs, up to a phase factor $\xi = \pm 1$. We use Le Couteur's representation.[39] Our β is proportional to his $\eta_0^{(+)}$. Thus the nonvanishing matrix elements of β for $s_1 \neq s_2$ are

$$\langle s_1, s_2| \beta |s_2', s_1'\rangle = \xi \delta_{s_1 s_1'} \delta_{s_2 s_2'} I. \tag{3.53}$$

On the other hand, for $s_1 = s_2$, the matrix elements depend on s, defined by (3.17); thus

$$\begin{aligned}\langle s_1, s_1, s, m| \beta |s_1', s_2', s', m'\rangle &= \langle s_1', s_2', s', m'| \beta |s_1, s_1, s, m\rangle \\ &= (-1)^s \xi \delta_{s_1 s_1'} \delta_{s_1 s_2'} \delta_{ss'} \delta_{mm'}. \end{aligned} \tag{3.54}$$

This latter case belongs to integral spin. ξ is evaluated below and found to be $\xi = (-1)^{[\lambda_2]}$, where $[\lambda_2]$ is the integral part of λ_2 [cf. Eq. (5.41) below].

In order to evaluate the space–time reflection operator ρ [cf. Eqs. (2.39) and (2.40)], we notice that $[\rho, S_{\mu\nu}] = 0$. Then,

$$\langle s_1, s_2| \rho |s_1', s_2'\rangle = (-1)^{r(s_1, s_2)} \delta_{s_1 s_1'} \delta_{s_2 s_2'} I. \tag{3.55}$$

$r(s_1, s_2)$ is an integer, because $\rho^2 = I$. Further, taking the matrix elements of the equation $\beta\rho = b_s \rho\beta$, between (s_1, s_2, s) and (s_2, s_1, s), then using (3.53) and (3.54) we get

$$\langle s_2, s_1| \rho |s_2, s_1\rangle = b_s \langle s_1, s_2| \rho |s_1, s_2\rangle. \tag{3.56}$$

Hence,

$$(-1)^{r(s_1, s_2)} = b_s (-1)^{r(s_2, s_1)}. \tag{3.57}$$

Further, from $\rho\gamma^{a\dot{b}} + \gamma^{a\dot{b}}\rho = 0$ we get

$$\begin{aligned}(-1)^{r(s_1+\frac{1}{2}, s_2+\frac{1}{2})} &= (-1)^{r(s_1-\frac{1}{2}, s_2-\frac{1}{2})} = (-1)^{r(s_1+\frac{1}{2}, s_2-\frac{1}{2})} \\ &= (-1)^{r(s_1-\frac{1}{2}, s_2+\frac{1}{2})} = (-1)^{1+r(s_1, s_2)}.\end{aligned} \tag{3.58}$$

One verifies easily that

$$\langle s_1, s_2| \rho |s_1, s_2\rangle = (-1)^{2s_1} I. \tag{3.59}$$

[We may also replace the factor $(-1)^{2s_1}$ by $(-1)^{2s_2}$.] We notice that

$$\langle s_2, s_1| \rho |s_2, s_1\rangle = (-1)^{2(s_2 - s_1)} \langle s_1, s_2| \rho |s_1, s_2\rangle.$$

For bosons (fermions), $s_1 - s_2$ is an integer (half-integer), such that $(-1)^{2(s_2 - s_1)} = b_s$, which proves (3.56).

Further, the nonvanishing matrix elements of τ [cf. Eq. (2.42)] are

$$\langle s_1, s_2| \tau |s_2, s_1\rangle = (-1)^{2s_1} a_s I, \quad \text{for} \quad s_1 \neq s_2, \tag{3.60}$$

and

$$\langle s_1, s_1, s, m| \tau |s_1, s_1, s, m\rangle = (-1)^{2s_1 + s} a_s. \tag{3.61}$$

D. Dirac's Equations

We show now that Dirac's spinor equations[6] belong to the particular representation $R_5(\lambda_1, \lambda_1)$. Contracting the field equations

$$(S_{\mu\nu} p_\nu + m\gamma_\mu)\psi = p_\mu \psi$$

and the dual equations [cf. Eq. (4.5) below]

$$(\tilde{S}_{\mu\nu} p_\nu + m\Gamma_\mu)\psi = p_\mu \Gamma_5 \psi$$

with $\lambda_1 s_\mu^{a\dot{b}}$, and using (3.3), (3.5), (3.6), and (3.18), we get

$$\begin{aligned}\{\Sigma_c^{(1)a} p^{c\dot{b}} - (\lambda_1/2)(1 + \Gamma_5) p^{a\dot{b}'}\}\psi \\ = -(m/2)(\gamma^{a\dot{b}} + \Gamma^{a\dot{b}})\psi,\end{aligned} \tag{3.62}$$

$$\begin{aligned}\{\Sigma_c^{(2)\dot{b}} p^{a\dot{c}} - (\lambda_1/2)(1 - \Gamma_5) p^{a\dot{b}}\}\psi \\ = -(m/2)(\gamma^{a\dot{b}} - \Gamma^{a\dot{b}})\psi,\end{aligned} \tag{3.63}$$

where

$$p^{a\dot{b}} = s_\mu^{a\dot{b}} p_\mu. \tag{3.64}$$

Now, in the representation $R_5(\lambda_1, \lambda_1)$ the admissible Poincaré components $\psi(s_1, s_2)$ are such that $s_1 + s_2 = \lambda_1$, where λ_1 is the spin of the particle. The nonvanishing matrix elements of $\gamma^{a\dot{b}}$ are [cf. Eqs. (3.27) and (3.43), and take $d_2 = 1$]:

$$\begin{aligned}\langle s_1, s_2| \gamma^{a\dot{b}} |s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2}\rangle &= -u^a(s_1) v^{\dot{b}}(s_2 + \tfrac{1}{2}), \\ \langle s_1, s_2| \gamma^{a\dot{b}} |s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}\rangle &= v^a(s_1 + \tfrac{1}{2}) u^{\dot{b}}(s_2).\end{aligned} \tag{3.65}$$

Also from (3.23) the nonvanishing matrix elements of $\Gamma^{a\dot{b}}$ are

$$\begin{aligned}\langle s_1, s_2| \Gamma^{a\dot{b}} |s_1 \mp \tfrac{1}{2}, s_2 \pm \tfrac{1}{2}\rangle \\ = \pm \langle s_1, s_2| \gamma^{a\dot{b}} |s_1 \mp \tfrac{1}{2}, s_2 \pm \tfrac{1}{2}\rangle.\end{aligned} \tag{3.66}$$

Further, from (3.15) we have

$$\langle s_1, s_2| \Gamma_5 |s_1', s_2'\rangle = [(s_1 - s_2)/\lambda_1] \delta_{s_1 s_1'} \delta_{s_2 s_2'} I. \tag{3.67}$$

From the preceding we get

$$\begin{aligned}u^a(s_1) v_c(s_1) p^{c\dot{b}} \psi(s_1, s_2) \\ = -m u^a(s_1) v^{\dot{b}}(s_2 + \tfrac{1}{2}) \psi(s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2})\end{aligned} \tag{3.68}$$

and

$$\begin{aligned}u^{\dot{b}}(s_2) v_{\dot{c}}(s_2) p^{a\dot{c}} \psi(s_1, s_2) \\ = m v^a(s_1 + \tfrac{1}{2}) u^{\dot{b}}(s_2) \psi(s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}).\end{aligned} \tag{3.69}$$

Contracting (3.68) with $v_a(s_1)$ and (3.69) with $v_{\dot{b}}(s_2)$, and using (3.28), we arrive at Dirac's equations

$$v_c(s_1)p^{c\dot{b}}\psi(s_1, s_2) = -mv^{\dot{b}}(s_2 + \tfrac{1}{2})\psi(s_1 - \tfrac{1}{2}, s_2 + \tfrac{1}{2}),$$
$$v_{\dot{c}}(s_2)p^{a\dot{c}}\psi(s_1, s_2) = mv^{a}(s_1 + \tfrac{1}{2})\psi(s_1 + \tfrac{1}{2}, s_2 - \tfrac{1}{2}). \tag{3.70}$$

We must remark, however, that Dirac's equations are not equivalent to the energy–momentum field equations, except for $\lambda_1 = \frac{1}{2}$, because they connect only two successive Poincaré components.

4. FORMAL STRUCTURE OF THE THEORY

A. The Dual Equations

From Eqs. (2.55) one obtains the following equations[1] by contracting with Γ_A and S_{ABC}, and using (2.29):

$$\Gamma_A p_A \psi \equiv (\Gamma_\mu p_\mu + m\Gamma_5)\psi = 0, \tag{4.1}$$

$$S_{ABC}p_C\psi = (\Gamma_A p_B - \Gamma_B p_A)\psi, \tag{4.2}$$

$$\Theta_{AB}p_B\psi = [\lambda_2(\lambda_2 + 1)/\lambda_1(\lambda_1 + 1)]p_A\psi. \tag{4.3}$$

In the last equation, use has been made of the value

$$\lambda_1^2(\lambda_1 + 1)^2 D_1 = \lambda_2(\lambda_2 + 1)(\lambda_1 + 1)(\lambda_1 + 2) \tag{4.4}$$

of the invariant D_1 in the irreducible representation $R_5(\lambda_1, \lambda_2)$, obtained in the preceding section. Taking $A = 5$ and $B = \mu$ in (4.2), we obtain the dual equations

$$(\tilde{S}_{\mu\nu}p_\nu + m\Gamma_\mu)\psi = \Gamma_5 p_\mu \psi. \tag{4.5}$$

In the three-dimensional notation, these equations read

$$(\boldsymbol{\sigma}\cdot\mathbf{p} + m\Gamma_0)\psi = \Gamma_5 p_0 \psi,$$
$$(\boldsymbol{\sigma}p_0 + i\boldsymbol{\alpha}\wedge\mathbf{p} + m\boldsymbol{\Gamma})\psi = \Gamma_5\mathbf{p}\psi, \tag{4.6}$$

while the energy–momentum equations are

$$P_0\psi \equiv (\boldsymbol{\alpha}\cdot\mathbf{p} + m\gamma_0)\psi = p_0\psi,$$
$$\mathbf{P}\psi \equiv (\boldsymbol{\alpha}p_0 + i\boldsymbol{\sigma}\wedge\mathbf{p} + m\boldsymbol{\gamma})\psi = \mathbf{p}\psi. \tag{4.7}$$

The rest of Eqs. (4.2) reads

$$(m\boldsymbol{\alpha} + \boldsymbol{\gamma}p_0 - \gamma_0\mathbf{p})\psi = i\boldsymbol{\Gamma}\wedge\mathbf{p}\psi,$$
$$(m\boldsymbol{\sigma} - i\boldsymbol{\gamma}\wedge\mathbf{p})\psi = (\Gamma_0\mathbf{p} - \boldsymbol{\Gamma}p_0)\psi. \tag{4.8}$$

These equations may be used to eliminate p_0 and obtain supplementary conditions containing only spatial derivatives.

The fifth of Eqs. (4.3) reads

$$\left(g_\mu p_\mu + m\frac{\lambda_2(\lambda_2 + 1)}{\lambda_1(\lambda_1 + 1)}\right)\psi = 0. \tag{4.9}$$

Further, contracting (4.5) with g_μ, then using (2.28) and (2.30), and noting that $g_\mu\Gamma_\mu = \Theta_\mu\Gamma_{\mu 5}$, we get

$$\{\Delta_\mu p_\mu + [m/(\lambda_1 + 1)]^2[\lambda_2(\lambda_2 + 1) + \lambda_1 + 1]\Gamma_5\}\psi = 0. \tag{4.10}$$

Equations (1.2), (4.1), (4.9), and (4.10) serve to obtain the relation between the successive Poincaré components of ψ by simple algebraic methods, without the troublesome solution of difference equations.

B. Observables

The definition of an observable, as that operator O which commutes with the Hamiltonian P_0 of the system, should be modified in such a way that $O\psi$ is a solution of the same field equations satisfied by ψ. O need not commute with $p_\mu - P_\mu$; it is sufficient that

$$[p_\mu - P_\mu, O]\psi = 0. \tag{4.11}$$

Trivially, the energy and momentum are observables. Also the angular momentum is an observable. Let

$$J_{\mu\nu} = L_{\mu\nu} + i\lambda_1 S_{\mu\nu},$$
$$L_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu. \tag{4.12}$$

Then, in favor of (2.4),

$$[J_{\mu\nu}, P_\alpha] = i(\delta_{\mu\alpha}P_\nu - \delta_{\nu\alpha}P_\mu). \tag{4.13}$$

The same relations hold for p_α. Hence

$$i[p_\alpha - P_\alpha, J_{\mu\nu}]\psi = \delta_{\mu\alpha}(p_\nu - P_\nu)\psi - \delta_{\nu\alpha}(p_\mu - P_\mu)\psi = 0.$$

Also the Pauli–Lubanski pseudovector (1.5) is observable. In fact, $[p_\mu, W_\nu] = 0$ and

$$\lambda_1[P_\mu, W_\nu] = p_\mu W_\nu - p_\nu W_\mu - m\lambda_1 S_{\mu\nu A}p_A.$$

Using (4.2), it follows from the last relation that $[P_\mu, W_\nu]\psi = 0$. Further, denoting

$$U_\mu = \lambda_1(\Gamma_5 p_\mu - \Gamma_\mu p_5), \tag{4.14}$$

the dual equations (4.5) read

$$W_\mu\psi = U_\mu\psi. \tag{4.15}$$

One expects that U_μ also is observable. Indeed,

$$[P_\mu, U_\nu]\psi = -m\delta_{\mu\nu}\Gamma_A p_A\psi = 0. \tag{4.16}$$

One verifies easily that $[W_\mu, U_\mu] = 0$. Thus, if we choose ψ as the eigenfunction of one W_μ, it is also an eigenfunction of the corresponding U_μ. W_μ acts irreducibly on each Poincaré component, while U_ν connects several Poincaré components. We may construct ψ as the eigenfunction of $W_\mu W_\mu$ and the helicity $W_0 = -iW_4$. A state of definite helicity satisfies

$$W_0\psi \equiv \lambda_1\boldsymbol{\sigma}\cdot\mathbf{p}\psi = w_0\psi, \tag{4.17}$$

and

$$U_0\psi \equiv \lambda_1(p_0\Gamma_5 - m\Gamma_0)\psi = w_0\psi. \tag{4.18}$$

We can also construct the simultaneous eigenfunctions of the angular momentum

$$\mathbf{J} = \mathbf{L} + \lambda_1\boldsymbol{\sigma}, \quad \mathbf{L} = \mathbf{r} \wedge \mathbf{p}, \tag{4.19}$$

such that

$$\mathbf{J}^2\psi = J(J+1)\psi,$$
$$J_3\psi = M\psi, \quad |M| \leq J. \tag{4.20}$$

In fact, J_k commutes with $W_\mu W_\mu$, W_0, U_0, p_0, P_0, and $\mathbf{p}^2$. Thus we may construct the wavefunction ψ as the eigenfunction of p_0, P_0, $\mathbf{p}^2$, $W_\mu W_\mu$, W_0, $\mathbf{J}^2$, and J_3, such that

$$p_0\psi \equiv i\partial_t\psi = \epsilon E_p\psi, \tag{4.21}$$

$$P_0\psi \equiv (-i\boldsymbol{\alpha}\cdot\nabla + m\gamma_0)\psi = \epsilon E_p\psi, \tag{4.22}$$

$$\mathbf{p}^2\psi \equiv -\nabla^2\psi = p^2\psi, \tag{4.23}$$

where $\epsilon = \pm 1$ is the energy sign, and

$$E_p = (p^2 + m^2)^{\frac{1}{2}}. \tag{4.24}$$

The helicity is $\sigma = w_0/p$. We can also define the helicity representation as the eigenfunction of p_0, P_0, p_3, P_3, $\mathbf{p}^2$, $W_\mu W_\mu$, W_0, and J_3. However, neither case is suitable for the connection with the wavefunction in the rest system, and they are not states of definite parity.

In order to define states of definite parity, we construct states of definite spin and orbital angular momentum. Since $\boldsymbol{\sigma}$ and $\mathbf{L}$ are not observable, we have to modify the spin in an observable way. This can be done in several ways, as in Dirac's theory.[30] We prefer to define the particle's spin in the Poincaré space in terms of the observable generators W_μ of PG:

$$\mathbf{S} = (\epsilon/m)\mathbf{W} - [m(E_p + m)]^{-1}W_0\mathbf{p}. \tag{4.25}$$

Shirokov[41] has shown that $\mathbf{S}$ is related to the spin in the rest system. It satisfies

$$[S_k, S_j] = i\epsilon_{kjn}S_n, \tag{4.26}$$

and

$$m^2\mathbf{S}^2 = W_\alpha W_\alpha, \quad \mathbf{S}\cdot\mathbf{p} = W_0. \tag{4.27}$$

The observable orbital angular momentum then is

$$\mathfrak{L} = \mathbf{J} - \mathbf{S}$$
$$= \mathbf{L} - i(\lambda_1\epsilon/m)\boldsymbol{\alpha} \wedge \mathbf{p}$$
$$+ \lambda_1[m(E_p + m)]^{-1}(\boldsymbol{\sigma} \wedge \mathbf{p}) \wedge \mathbf{p}. \tag{4.28}$$

It satisfies

$$[\mathfrak{L}_k, \mathfrak{L}_j] = i\epsilon_{kjn}\mathfrak{L}_n, \quad [\mathfrak{L}_k, S_j] = 0. \tag{4.29}$$

$\mathfrak{L}_k$ and S_k commute with the parity operator $P = \beta\Pi_\tau$.

[41] Iu. M. Shirokov, Zh. Eksp. Teor. Fiz. 33, 1196 (1957) [Sov. Phys.—JETP 6, 919 (1958)].

We can thus construct the canonical representations: (1) with definite momentum as eigenfunctions of p_μ, P_μ, S_3, and $\mathbf{S}^2$; (2) with definite angular momentum as eigenfunctions of p_0, P_0, $\mathbf{p}^2$, P, $\mathfrak{L}^2$, $\mathbf{S}^2$, J_3, and $\mathbf{J}^2$ or S_3, such that

$$\mathfrak{L}^2\psi = L(L+1)\psi,$$
$$\mathfrak{L}_3\psi = M_L\psi, \quad |M_L| \leq L, \tag{4.30}$$
$$\mathbf{S}^2\psi = \lambda_2(\lambda_2 + 1)\psi,$$
$$S_3\psi = M_s\psi, \quad |M_s| \leq \lambda_2. \tag{4.31}$$

L is a nonnegative integer.

C. Conserved Currents

The number of conserved currents increases with the dimension of the representation. They can be obtained from the complete reduction of the direct product $R_5(\lambda_1, \lambda_2) \otimes R_5(\lambda_1, \lambda_2)$ under hLG. However, these currents are not all independent, and are of two different types: (1) those formed by the five-dimensional metric $\psi^\dagger\tau O\psi$, like $\bar\psi\rho\Gamma_\mu\psi$ and $\bar\psi\rho\Delta_\mu\psi$, and related to the Lorentz invariant $\bar\psi\rho\Gamma_5\psi$; (2) those formed by the four-dimensional metric $\psi^\dagger\beta O\psi$, like $\bar\psi\gamma_\mu\psi$ and $\bar\psi g_\mu\psi$, and related to $\bar\psi\psi$.

Consider Eqs. (1.1) and (1.2) in configuration space, and their Hermitian conjugates

$$\gamma_\mu\partial_\mu\psi + im\psi = 0,$$
$$\partial_\mu\bar\psi\gamma_\mu - im\bar\psi = 0, \tag{4.32}$$

and

$$\partial_\mu\psi - S_{\mu\nu}\partial_\nu\psi - im\gamma_\mu\psi = 0,$$
$$\partial_\mu\bar\psi + \partial_\nu\bar\psi S_{\mu\nu} + im\bar\psi\gamma_\mu = 0. \tag{4.33}$$

Denoting

$$j_\mu = \bar\psi\gamma_\mu\psi, \tag{4.34}$$

$$t_{\mu\nu} = (i/2)[(\partial_\mu\bar\psi)\gamma_\nu\psi - \bar\psi\gamma_\nu\partial_\mu\psi], \tag{4.35}$$

and

$$\tau_{\mu\nu} = \tau_{\mu\nu} = \lambda_1\bar\psi(\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu)\psi + \delta_{\mu\nu}\bar\psi\psi, \tag{4.36}$$

we find that

$$\partial_\nu j_\nu = \partial_\nu t_{\mu\nu} = \partial_\nu\tau_{\nu\mu} = \partial_\nu\tau_{\mu\nu} = 0. \tag{4.37}$$

The last relation is obtained by eliminating $\bar\psi\gamma_\mu\psi$ from (4.33), substituting $S_{\mu\nu} = \lambda_1[\gamma_\mu, \gamma_\nu]$, and then using (4.32). These are the conservation laws of probability and energy–momentum. $t_{\mu\nu}$ is the canonical stress–energy–momentum tensor. We show now that it is proportional to $\tau_{\mu\nu}$. Consider Eqs. (1.1) in momentum space, and take their Hermitian conjugates. From these equations it follows directly that

$$\bar\psi\gamma_\mu\psi = (p_\mu/m)\bar\psi\psi. \tag{4.38}$$

Hence,

$$t_{\mu\nu} = p_\mu\bar\psi\gamma_\nu\psi = (p_\mu p_\nu/m)\bar\psi\psi. \tag{4.39}$$

Turning now to the rest system, $p_k = 0$ and $p_0 = \epsilon m$, Eqs. (1.1) become

$$\gamma_0\psi_0 = \epsilon\psi_0, \tag{4.40}$$

$$\gamma_k\psi_0 = -\epsilon\alpha_k\psi_0. \tag{4.41}$$

Hence β, defined by (2.34), satisfies

$$\beta\psi_0 = \beta_\epsilon\psi_0, \tag{4.42}$$

where

$$\beta_\epsilon = \begin{cases} 1, & \text{for bosons,} \\ \epsilon, & \text{for fermions.} \end{cases} \tag{4.43}$$

Since $\bar{\psi}\psi$ is Lorentz-invariant, then

$$\bar{\psi}\psi = \bar{\psi}_0\psi_0 = \beta_\epsilon\psi_0^\dagger\psi_0 \tag{4.44}$$

and

$$\bar{\psi}\gamma_0\psi = \epsilon\beta_\epsilon(E_p/m)\psi_0^\dagger\psi_0. \tag{4.45}$$

The probability density $\bar{\psi}\gamma_0\psi$ is positive-definite for fermions, but not for bosons. However, in the quantized theory $\bar{\psi}\gamma_0\psi$ becomes positive-definite for fermions and bosons.[37]

Further, taking the Hermitian conjugate of (4.41), one verifies easily that

$$\psi_0^\dagger\gamma_k\psi_0 = \psi_0^\dagger\alpha_k\psi_0 = 0, \tag{4.46}$$

$$\bar{\psi}_0\gamma_k = \epsilon\bar{\psi}_0\alpha_k. \tag{4.47}$$

From (4.40) and (4.46) we get

$$\begin{aligned}\tau_{4k}^o &= i\lambda_1\bar{\psi}_0(\gamma_k\gamma_0 + \gamma_0\gamma_k)\psi_0 \\ &= 2i\lambda_1\epsilon\beta_\epsilon\psi_0^\dagger\gamma_k\psi_0 = 0.\end{aligned} \tag{4.48}$$

Further, using $\lambda_1[\gamma_k, \alpha_j] = \gamma_0\delta_{kj}$, (4.40), (4.41), and (4.47) we get

$$\begin{aligned}\lambda_1\bar{\psi}_0(\gamma_k\gamma_j + \gamma_j\gamma_k)\psi_0 &= \lambda_1\epsilon\bar{\psi}_0(\alpha_k\gamma_j - \gamma_j\alpha_k)\psi_0 \\ &= -\epsilon\bar{\psi}_0\gamma_0\psi_0\delta_{kj} = -\bar{\psi}_0\psi_0\delta_{kj}.\end{aligned} \tag{4.49}$$

Hence,

$$\tau_{kj}^o = 0. \tag{4.50}$$

The only nonvanishing element of $\tau_{\mu\nu}^o$ is then

$$\tau_{44}^o = \bar{\psi}_0\psi_0 - 2\lambda_1\bar{\psi}_0\gamma_0^2\psi_0 = (1 - 2\lambda_1)\bar{\psi}\psi. \tag{4.51}$$

To obtain $\tau_{\mu\nu}$ in an arbitrary inertial system, we apply the Lorentz transformation

$$\tau_{\mu\nu} = a_{\mu\alpha}a_{\nu\beta}\tau_{\alpha\beta}^o = a_{\mu4}a_{\nu4}\tau_{44}^o, \tag{4.52}$$

where

$$a_{\mu4} = -ip_\mu/m. \tag{4.53}$$

Hence,

$$\tau_{\mu\nu} = (2\lambda_1 - 1)(p_\mu p_\nu/m^2)\bar{\psi}\psi. \tag{4.54}$$

Comparing this with (4.39) we get

$$m\tau_{\mu\nu} = (2\lambda_1 - 1)t_{\mu\nu}. \tag{4.55}$$

$\tau_{\mu\nu}$ vanishes only for Dirac's theory ($\lambda_1 = \frac{1}{2}$).

Consider further the dual equations (4.1) and (4.5) in configuration space and their Hermitian conjugates

$$\begin{aligned}\Gamma_\mu\partial_\mu\psi + im\Gamma_5\psi &= 0, \\ \partial_\mu\bar{\psi}\rho\Gamma_\mu - im\bar{\psi}\rho\Gamma_5 &= 0.\end{aligned} \tag{4.56}$$

It follows that

$$\partial_\mu(\bar{\psi}\rho\Gamma_\mu\psi) = 0. \tag{4.57}$$

However, this conserved current vanishes identically, except for Dirac's theory ($\lambda_1 = \frac{1}{2}$). To prove this we turn to the momentum space. From (4.5) it follows that

$$\bar{\psi}\rho\Gamma_\mu\psi = (p_\mu/m)\bar{\psi}\rho\Gamma_5\psi. \tag{4.58}$$

In the rest system we have

$$\Gamma_0\psi_0 = \epsilon\Gamma_5\psi_0. \tag{4.59}$$

Using $\Gamma_0 = \lambda_1[\Gamma_5, \gamma_0]$, we get

$$\begin{aligned}\bar{\psi}\rho\Gamma_5\psi &= \bar{\psi}_0\rho\Gamma_5\psi_0 = \epsilon\beta_\epsilon\psi_0^\dagger\rho\Gamma_0\psi_0 \\ &= \lambda_1\epsilon\beta_\epsilon(\psi_0^\dagger\rho\Gamma_5\gamma_0\psi_0 - \psi_0^\dagger\rho\gamma_0\Gamma_5\psi_0) \\ &= \lambda_1\epsilon\beta_\epsilon(\psi_0^\dagger\rho\Gamma_5\gamma_0\psi_0 + \psi_0^\dagger\gamma_0\rho\Gamma_5\psi_0) \\ &= 2\lambda_1\beta_\epsilon\psi_0^\dagger\rho\Gamma_5\psi_0 = 2\lambda_1\bar{\psi}\rho\Gamma_5\psi.\end{aligned} \tag{4.60}$$

This proves that $\bar{\psi}\rho\Gamma_5\psi = 0$, unless $\lambda_1 = \frac{1}{2}$. The case $\lambda_1 = \lambda_2 = \frac{1}{2}$ corresponds to $\rho = \Gamma_5$ and $\bar{\psi}\rho\Gamma_\mu\psi = \bar{\psi}\gamma_\mu\psi$. From (4.58) it follows that, for $\lambda_1 \neq \frac{1}{2}$,

$$\bar{\psi}\rho\Gamma_A\psi = 0. \tag{4.61}$$

Further, using (4.3), one verifies easily that

$$\bar{\psi}g_\mu\psi = [\lambda_2(\lambda_2 + 1)/\lambda_1(\lambda_1 + 1)]\bar{\psi}\gamma_\mu\psi. \tag{4.62}$$

By similar arguments one can prove that, for $\lambda_1 \neq \frac{1}{2}$,

$$\bar{\psi}\rho\Delta_\mu\psi = 0. \tag{4.63}$$

Hence, the only nontrivial current is $\bar{\psi}\gamma_\mu\psi$.

D. Lagrangian Formalism

Consider the Lagrangian density

$$\mathfrak{L} = -i\bar{\psi}\gamma_\mu\partial_\mu\psi + m\bar{\psi}\psi. \tag{4.64}$$

The canonical momenta are

$$\pi_\mu^\alpha = \partial\mathfrak{L}/\partial(\partial_\mu\psi^\alpha) = -i(\bar{\psi}\gamma_\mu)^\alpha,$$

where ψ^α are the components of ψ. For simplicity we suppress the index α and introduce the row matrix

$$\pi_\mu = -i\bar{\psi}\gamma_\mu. \tag{4.65}$$

The Lagrange-Euler equations $\partial_\mu\pi_\mu^\alpha = \partial\mathfrak{L}/\partial\psi^\alpha$ lead to Bhabha's equations

$$\partial_\mu\bar{\psi}\gamma_\mu = im\bar{\psi}, \quad \gamma_\mu\partial_\mu\psi = -im\psi. \tag{4.66}$$

Then the symmetrized stress–energy–momentum tensor may be written in the form

$$T_{\mu\nu} = -\mathfrak{L}\delta_{\mu\nu} + \tfrac{1}{2}(\pi_\mu\partial_\nu\psi + \pi_\nu\partial_\mu\psi) + \partial_\alpha f_{\alpha\mu\nu}, \quad (4.67)$$

where

$$f_{\alpha\mu\nu} = (\lambda_1/2)(\pi_\mu S_{\alpha\nu}\psi + \pi_\nu S_{\alpha\mu}\psi). \quad (4.68)$$

This form is slightly different, although essentially equivalent to the usual form.[42] For convenience we have symmetrized $f_{\alpha\mu\nu}$ with respect to μ and ν. Using only relations (2.9), we get

$$\partial_\alpha(\bar\psi\gamma_\mu S_{\alpha\nu}\psi) = -\bar\psi\gamma_\mu S_{\nu\alpha}\partial_\alpha\psi - (\partial_\alpha\bar\psi)S_{\nu\alpha}\gamma_\mu\psi + (1/\lambda_1)[(\partial_\alpha\bar\psi)\gamma_\alpha\delta_{\mu\nu} - (\partial_\mu\bar\psi)\gamma_\nu\psi].$$

Hence,

$$T_{\mu\nu} = -(\mathfrak{L} + \bar{\mathfrak{L}})\delta_{\mu\nu} + m\tau_{\mu\nu} - (\lambda_1 - 1)(t_{\mu\nu} + t_{\nu\mu}) + (\lambda_1/2)\bar\psi[\gamma_\mu(p_\nu - P_\nu)\psi + \gamma_\nu(p_\mu - P_\mu)]\psi + (\lambda_1/2)[\overline{(p_\nu - P_\nu)\psi}\gamma_\mu + \overline{(p_\mu - P_\mu)\psi}\gamma_\nu]\psi, \quad (4.69)$$

where

$$\bar{\mathfrak{L}} = i(\partial_\mu\bar\psi)\gamma_\mu\psi + m\bar\psi\psi$$

and

$$\overline{O\psi} = (O\psi)^\dagger\beta. \quad (4.70)$$

The field energy–momentum four-vector is

$$\langle P_\mu\rangle = \int T_{\mu\nu}\, d\sigma_\nu, \quad (4.71)$$

where

$$d\sigma_\mu = (dx_2\, dx_3\, dt,\ dx_3\, dx_1\, dt,\ dx_1\, dx_2\, dt,\ -i\, dV), \quad (4.72)$$

$dV = dx_1\, dx_2\, dx_3$, and the integration is carried out on a spacelike hypersurface.

One can also use the unsymmetrized stress–energy–momentum tensor

$$T^{(a)}_{\mu\nu} = \tfrac{1}{2}(T'_{\mu\nu} + T'^*_{\mu\nu}) \quad (4.73)$$

to define the field energy–momentum vector, where

$$T'_{\mu\nu} = -\mathfrak{L}\delta_{\mu\nu} + \pi_\mu\partial_\nu\psi. \quad (4.74)$$

From Eqs. (4.66) it follows that $\mathfrak{L} = 0$, such that

$$T^{(a)}_{\mu\nu} = t_{\mu\nu}, \quad (4.75)$$

where $t_{\mu\nu}$ is given by (4.35). Also

$$T_{\mu\nu} = T^{(a)}_{\mu\nu} + \partial_\alpha f_{\alpha\mu\nu}. \quad (4.76)$$

The term $\partial_\alpha f_{\alpha\mu\nu}$ cancels on integrating over the whole space, such that

$$\langle P_\mu\rangle = \int T_{\mu\nu}\, d\sigma_\nu = \int T^{(a)}_{\mu\nu}\, d\sigma_\nu = \int t_{\mu\nu}\, d\sigma_\nu. \quad (4.77)$$

[42] Confer J. Hamilton, *The Theory of Elementary Particles* (Clarendon Press, Oxford, England, 1959), p. 431.

However, in favor of the field equations

$$(p_\mu - P_\mu)\psi = 0,$$

we have the local equality

$$T_{\mu\nu} = T^{(a)}_{\mu\nu} = t_{\mu\nu}. \quad (4.78)$$

Thus the Lagrangian formalism should be supplemented by the local equality (4.78), in order to derive the field equations. It remains, however, to justify this condition. Also the uniqueness of the choice

$$(p_\mu - P_\mu)\psi = 0$$

to ensure (4.78) is an open question.

E. Expectation Values

We may define the expectation value of an operator O in a covariant way as

$$\langle O\rangle = \frac{1}{2}\int(\bar\psi\gamma_\mu O\psi + \overline{O\psi}\gamma_\mu\psi)\, d\sigma_\mu, \quad (4.79)$$

where the integration is carried out on a spacelike hypersurface. On a flat surface, $t = \text{const}$, we have

$$\langle O\rangle = \frac{1}{2}\int(\bar\psi\gamma_0 O\psi + \overline{O\psi}\,\gamma_0\psi)\, dV. \quad (4.80)$$

We remark that, in order to evaluate the expectation value of any timelike component, we have to turn first from x_4 to the physical component $x_0 = -ix_4$, and then evaluate the expectation value. The expectation value of the identity operator

$$\langle I\rangle = \int\bar\psi\gamma_0\psi\, dV \quad (4.81)$$

is a constant of motion and may be normalized. Further, if $O\psi_n = O_n\psi_n$, then

$$\langle O\rangle_n = (\tfrac{1}{2})(O_n + O_n^*)\langle I\rangle. \quad (4.82)$$

The expectation value is equal to the real part of the eigenvalue. Also the canonical stress–energy–momentum tensor gives the expectation value of energy and momentum as

$$\langle p_k\rangle = -i\int t_{k4}\, dV,$$
$$\langle p_0\rangle = -\int t_{44}\, dV. \quad (4.83)$$

Consider further the expectation value of the Pauli–Lubanski pseudovector. Introducing

$$W_{\mu\alpha} = \lambda_1[\bar\psi\gamma_\alpha\tilde S_{\mu\nu}\partial_\nu\psi - (\partial_\nu\bar\psi)\tilde S_{\mu\nu}\gamma_\alpha\psi], \quad (4.84)$$

one verifies easily, using (4.57), that

$$\partial_\alpha W_{\mu\alpha} = 0. \quad (4.85)$$

Hence,

$$\langle W_\mu \rangle = -i \int W_{\mu 4}\, dV \tag{4.86}$$

are constants of motion. One verifies directly that $\langle W_k \rangle$ and $\langle W_0 \rangle = \lambda_1 \langle \boldsymbol{\sigma} \cdot \mathbf{p} \rangle$ are expectation values in the sense of (4.80). Introducing further the mixed third-rank tensor

$$\begin{aligned} \Sigma_{\mu\nu\alpha} = -\Sigma_{\nu\mu\alpha} = {} & (i\lambda_1/2)\bar{\psi}(\gamma_\alpha S_{\mu\nu} + S_{\mu\nu}\gamma_\alpha)\psi \\ & + (i/2)[(\partial_\nu \bar{\psi})x_\mu - (\partial_\mu \bar{\psi})x_\nu]\gamma_\alpha \psi \\ & - (i/2)\bar{\psi}\gamma_\alpha(x_\mu \partial_\nu - x_\nu \partial_\mu)\psi, \end{aligned} \tag{4.87}$$

one verifies easily, using (4.32) and (4.33), that

$$\partial_\alpha \Sigma_{\mu\nu\alpha} = 0. \tag{4.88}$$

Hence,

$$M_{\mu\nu} = -i \int \Sigma_{\mu\nu 4}\, dV \tag{4.89}$$

are constants of motion. Obviously, $M_{\mu\nu}$ is the expectation value of the angular-momentum operator

$$J_{\mu\nu} = -i(x_\mu \partial_\nu - x_\nu \partial_\mu) + i\lambda_1 S_{\mu\nu}. \tag{4.90}$$

We have thus established the conservation of the expectation values of the observables defined before.

F. Scalar Product

We define the Lorentz-invariant scalar product of two "functions" ψ_1 and ψ_2 in the following way:

$$(\psi_1, \psi_2) = \int \bar{\psi}_1 \gamma_\mu \psi_2\, d\sigma_\mu, \tag{4.91}$$

where the integration is carried out on a spacelike hypersurface σ. The scalar product is thus the flux of $\bar{\psi}_1 \gamma_\mu \psi_2$ across σ, due to independent disturbances, because disturbances on σ are connected by signals moving faster than light. With this definition of the scalar product, the expectation value (4.79) of an operator A becomes

$$\langle A \rangle = \tfrac{1}{2}[(A\psi, \psi) + (\psi, A\psi)]. \tag{4.92}$$

The scalar product has the following usual properties:

$$(\psi_1, \psi_2)^* = (\psi_2, \psi_1), \tag{4.93a}$$

$$(\psi_1, a\psi_2) = a(\psi_1, \psi_2),$$

$$(a\psi_1, \psi_2) = a^*(\psi_1, \psi_2), \tag{4.93b}$$

$$(\psi_1 + \psi_2, \phi) = (\psi_1, \phi) + (\psi_2, \phi),$$

$$(\phi, \psi_1 + \psi_2) = (\phi, \psi_1) + (\phi, \psi_2). \tag{4.93c}$$

All properties are obvious except the first. To prove it, we notice that, for a Hermitian operator, $A^* = A^{\mathrm{T}}$, such that

$$(\psi_1^\dagger A \psi_2)^* = \psi_1^{\mathrm{T}} A^* \psi_2^* = \psi_1^{\mathrm{T}} A^{\mathrm{T}} \psi_2^* = \psi_2^\dagger A \psi_1. \tag{4.94}$$

From this relation, and the fact that $\beta\gamma_0$ and $\beta\gamma_k$ are Hermitian, relation (4.93a) follows directly. For practical purposes we define the scalar product on a flat surface ($t = \text{const}$):

$$(\psi_1, \psi_2) = \int \bar{\psi}_1 \gamma_0 \psi_2\, dV. \tag{4.95}$$

We prove now that two eigenfunctions belonging to different energies are orthogonal, where orthogonality implies the vanishing of the scalar product as usual. Consider the wave equations for the two states ψ_1 and ψ_2:

$$p_0 \psi_1 = -i\boldsymbol{\alpha} \cdot \nabla \psi_1 + m\gamma_0 \psi_1,$$

$$p_0' \psi_2 = -i\boldsymbol{\alpha} \cdot \nabla \psi_2 + m\gamma_0 \psi_2.$$

Taking the adjoint of the first equation, one verifies that

$$\bar{\psi}_1 \gamma_0 \psi_2 = [(p_0 + p_0')/2m]\bar{\psi}_1 \psi_2 + (i/2m)\,\mathrm{div}\,(\bar{\psi}_1 \boldsymbol{\alpha} \psi_2). \tag{4.96}$$

Integrating over the whole space (we may consider a finite normalization volume, on the boundary of which ψ vanishes), the divergence term cancels, and we get

$$(\psi_1, \psi_2) = \frac{p_0 + p_0'}{2m} \int \bar{\psi}_1 \psi_2\, dV. \tag{4.97}$$

Now, any eigenfunction of P_0 of eigenvalue $p_0 = \epsilon E_p = \epsilon(p^2 + m^2)^{\frac{1}{2}}$ is also an eigenfunction of $\mathbf{p}^2 = -\nabla^2$. If

$$(\nabla^2 + p^2)\psi_1 = [\nabla^2 + (p')^2]\psi_2 = 0,$$

then obviously $\int \bar{\psi}_1 \psi_2\, dV = 0$, if $p \neq p'$. On the other hand, if $p = p'$, then $p_0 = \epsilon E_p$ and $p_0' = \epsilon' E_p$. If $\epsilon \neq \epsilon'$, then $\epsilon' = -\epsilon$, $p_0' = -p_0$, and $(\psi_1, \psi_2) = 0$, from (4.97). This completes the proof that $(\psi_1, \psi_2) = 0$, if $p_0 \neq p_0'$.

Introducing the Fourier transform

$$\psi(\mathbf{r}, t) = (2\pi)^{-\frac{3}{2}} \int \phi(\mathbf{p}, t) e^{i\mathbf{r}\cdot\mathbf{p}}\, d^3p, \tag{4.98}$$

then

$$(\psi_1, \psi_2) = \int \bar{\phi}_1 \gamma_0 \phi_2\, d^3p, \tag{4.99}$$

where ϕ_1 and ϕ_2 are the Fourier transforms of ψ_1 and ψ_2 respectively, satisfying the field equations in momentum space. Consider the energy equation and its transpose

$$(\boldsymbol{\alpha} \cdot \mathbf{p} + m\gamma_0)\phi_2 = p_0 \phi_2,$$

$$-\bar{\phi}_1 \boldsymbol{\alpha} \cdot \mathbf{p} + m\bar{\phi}_1 \gamma_0 = p_0 \bar{\phi}_1,$$

Hence,

$$\bar{\phi}_1 \gamma_0 \phi_2 = (p_0/m)\bar{\phi}_1 \phi_2, \tag{4.100}$$

such that

$$(\psi_1, \psi_2) = \frac{\epsilon}{m} \int \bar{\phi}_1 \phi_2 E_p\, d^3p. \tag{4.101}$$

Relation (4.101) is equivalent to Bargmann–Wigner's scalar product[13] for positive energies. The only difference is that we are using an irreducible representation of the matrices γ_μ, while they use the reducible representation of the direct product of $2\lambda_1$ representations $R_5(\frac{1}{2}, \frac{1}{2})$.

5. CANONICAL REPRESENTATION

A. Chakrabarti Transformation

Consider the pure Lorentz transformation which brings the wavefunction to the rest system. This is obtained from (2.57) by choosing $\mathbf{v} = -\mathbf{p}/p_0$:

$$V = \exp(-\lambda_1\theta_1\boldsymbol{\alpha}\cdot\mathbf{p}/p), \tag{5.1}$$

where

$$\cosh\theta_1 = E_p/m, \quad \sinh\theta_1 = \epsilon p/m, \tag{5.2}$$
$$E_p = (p^2 + m^2)^{\frac{1}{2}},$$

and $\epsilon = \pm 1$ is the energy sign. This transformation was used by Chakrabarti[31] to connect the solutions of Dirac's equations for $\lambda_1 = \frac{1}{2}$ to Shirokov's[41] canonical representations of PG. It was used recently by Sesma[43] in connection with Kemmer's theory $\lambda_1 = 1$, $\lambda_2 = 1, 0$. We have made here a slight modification in the Chakrabarti transformation, by using (5.2) instead of $\cosh\theta_1 = p_0/m$ and $\sinh\theta_1 = p/m$. This modifies Chakrabarti's results by replacing m by the rest energy ϵm. For plane-wave solutions, one can study the wave equations directly in the rest system, $p_k = 0$, $p_0 = \epsilon m$. However, the transformation to the Chakrabarti representation is more useful, because if we construct states of definite angular momentum, the linear momentum is not observable. We prefer the Chakrabarti representation to the Foldy–Wouthuysen representation, because, in the former, each Poincaré component transforms separately. If ϕ is the wavefunction in this representation, then

$$\phi = V\psi, \tag{5.3}$$

and its Poincaré components are given by

$$\phi(s_1, s_2) = V(s_1, s_2)\psi(s_1, s_2) \tag{5.4}$$

where

$$V(s_1, s_2) = \langle s_1, s_2 | V | s_1, s_2\rangle. \tag{5.5}$$

For plane-wave solutions, ϕ is the wavefunction in the rest system.

The operators in this representation are given by

$$O' = VOV^{-1}. \tag{5.6}$$

[43] J. Sesma, J. Math. Phys. 7, 1300 (1966).

Using the expansion theorem (2.36) we get

$$\Gamma'_5 = \Gamma_5, \tag{5.7}$$

$$m\boldsymbol{\sigma}' = E_p\boldsymbol{\sigma} - \zeta_p(\boldsymbol{\sigma}\cdot\mathbf{p})\mathbf{p} - i\epsilon\boldsymbol{\alpha}\wedge\mathbf{p},$$
$$m\boldsymbol{\alpha}' = E_p\boldsymbol{\alpha} - \zeta_p(\boldsymbol{\alpha}\cdot\mathbf{p})\mathbf{p} - i\epsilon\boldsymbol{\sigma}\wedge\mathbf{p}, \tag{5.8}$$

and

$$m\gamma'_0 = E_p\gamma_0 + \epsilon\boldsymbol{\gamma}\cdot\mathbf{p},$$
$$m\boldsymbol{\gamma}' = m\boldsymbol{\gamma} + \zeta_p(\boldsymbol{\gamma}\cdot\mathbf{p})\mathbf{p} + \epsilon\gamma_0\mathbf{p}, \tag{5.9}$$

where $\zeta_p = (E_p + m)^{-1}$. Due to the invariance of Γ_5, we obtain for Γ'_μ, g'_μ, and Δ'_μ relations similar to (5.9). The generators of hLG become

$$\mathbf{J}' = \mathbf{J} = \mathbf{L} + \lambda_1\boldsymbol{\sigma},$$
$$\mathbf{G}' = \mathbf{N} + \lambda_1\epsilon\zeta_p\boldsymbol{\sigma}\wedge\mathbf{p}, \tag{5.10}$$

where

$$\mathbf{G} = \mathbf{N} + i\lambda_1\boldsymbol{\alpha},$$
$$\mathbf{N} = t\mathbf{p} - \mathbf{r}p_0. \tag{5.11}$$

The Pauli–Lubanski pseudovector

$$\mathbf{W} = \lambda_1(\boldsymbol{\sigma}p_0 + i\boldsymbol{\alpha}\wedge\mathbf{p}),$$
$$W_0 = \lambda_1\boldsymbol{\sigma}\cdot\mathbf{p} \tag{5.12}$$

becomes

$$\mathbf{W}' = \lambda_1\epsilon[m\boldsymbol{\sigma} + \zeta_p(\boldsymbol{\sigma}\cdot\mathbf{p})\mathbf{p}],$$
$$W'_0 = W_0. \tag{5.13}$$

$\mathbf{J}'$, $\mathbf{G}'$, and $\mathbf{W}'$ are identical with Shirokov's expressions[41] obtained by transforming W_μ to the laboratory system as a four-vector. From (5.12) one verifies directly that the observable spin and orbital angular momenta [cf. (4.25) and (4.28)] are related directly to the spin and orbital angular momenta in the canonical representation

$$\mathbf{S} = \lambda_1 V^{-1}\boldsymbol{\sigma}V, \quad \mathfrak{L} = V^{-1}\mathbf{L}V. \tag{5.14}$$

From (5.9) and the similar relations for Γ'_μ, g'_μ, and Δ'_μ, Eqs. (1.2), (4.1), (4.9), and (4.10) become

$$\begin{aligned}
&\gamma_0\phi = \epsilon\phi,\\
&\Gamma_0\phi = \epsilon\Gamma_5\phi,\\
&g_0\phi = \epsilon[\lambda_2(\lambda_2+1)/\lambda_1(\lambda_1+1)]\phi,\\
&\Delta_0\phi = \epsilon\{[\lambda_2(\lambda_2+1) + \lambda_1 + 1]/(\lambda_1+1)^2\}\Gamma_5\phi.
\end{aligned} \tag{5.15}$$

These are just the rest-system equations. The energy–momentum operators (2.76) become

$$P'_0 = E_p\gamma_0 + \epsilon\mathbf{A}^{(\epsilon)}\cdot\mathbf{p},$$
$$\mathbf{P}' = \epsilon\gamma_0\mathbf{p} + m\mathbf{A}^{(\epsilon)} + \zeta_p(\mathbf{A}^{(\epsilon)}\cdot\mathbf{p})\mathbf{p}, \tag{5.16}$$

where

$$\mathbf{A}^{(\epsilon)} = \boldsymbol{\gamma} + \epsilon\boldsymbol{\alpha}. \tag{5.17}$$

From the field equations $P'_\mu\phi = p_\mu\phi$ and $\gamma_0\phi = \epsilon\phi$, we get

$$\mathbf{A}^{(\epsilon)}\phi \equiv (\boldsymbol{\gamma} + \epsilon\boldsymbol{\alpha})\phi = 0. \tag{5.18}$$

$\boldsymbol{\sigma}$, $\mathbf{A}^{(\epsilon)}$, and γ_0 are the generators of the little group of $SO(4, 1)$. One verifies that

$$\begin{aligned} [A_k^{(\epsilon)}, A_l^{(\epsilon)}] &= 0, \\ \lambda_1[\sigma_k, A_l^{(\epsilon)}] &= i\epsilon_{kln}A_n^{(\epsilon)}, \\ \lambda_1[\gamma_0, A_k^{(\epsilon)}] &= \epsilon A_k^{(\epsilon)}, \\ [\sigma_k, \gamma_0] &= 0. \end{aligned} \tag{5.19}$$

Finally, the dual equations (4.6) lead to

$$(\Gamma + \epsilon\sigma)\phi = 0. \tag{5.20}$$

B. The Poincaré Components

We are now in a position to discuss the solutions of the field equations. We drop the time factor $\exp(-ip_0t)$. We consider the direct-sum representation (3.17). The plane-wave solutions with definite spin have the Poincaré components

$$\phi_{\mathbf{p},\epsilon,\lambda_2,M_s}(s_1, s_2, s) = a(\epsilon, s_1, s_2)\delta_{s\lambda_2}\eta_{\lambda_2 M_s}e^{i\mathbf{r}\cdot\mathbf{p}}, \tag{5.21}$$

where $\eta_{\lambda_2 M_s}$ is a realization of the representation $D(\lambda_2)$ of $O(3)$, such that

$$\begin{aligned} \mathbf{S}'^2\eta_{\lambda_2 M_s} &\equiv \lambda_1^2\boldsymbol{\sigma}^2\eta_{\lambda_2 M_s} = \lambda_2(\lambda_2 + 1)\eta_{\lambda_2 M_s}, \\ S_3'\eta_{\lambda_2 M_s} &\equiv \lambda_1\sigma_3\eta_{\lambda_2 M_s} = M_s\eta_{\lambda_2 M_s}, \quad |M_s| \leq \lambda_2. \end{aligned} \tag{5.22}$$

The terms $a(\epsilon, s_1, s_2)$ are constant coefficients which are determined from the solution of (5.15) and the normalization of the wavefunction. In the original representation

$$\begin{aligned} &\psi(s_1, s_2, s, m_s) \\ &\quad = a(\epsilon, s_1, s_2)e^{i\mathbf{r}\cdot\mathbf{p}}\langle s_1, s_2, s, m_s|\, V^{-1}\,|s_1, s_2, \lambda_2, M_s\rangle. \end{aligned} \tag{5.23}$$

These matrix elements of V^{-1} were studied by Pursey.[23] Here λ_2 and M_s are the eigenvalues of the observable spin (4.25).

Consider next the Poincaré component in the angular-momentum basis. Let it be an eigenfunction of $\boldsymbol{\sigma}^2$, σ_3, $\mathbf{L}^2$, and L_3. Here $\mathbf{p}^2$ is also an observable. Then

$$\begin{aligned} \mathfrak{L}'^2\phi(s_1, s_2) &\equiv \mathbf{L}^2\phi(s_1, s_2) = -(\mathbf{r}\wedge\nabla)^2\phi(s_1, s_2) \\ &= L(L + 1)\phi(s_1, s_2), \\ \mathfrak{L}_3'\phi(s_1, s_2) &\equiv L_3\phi(s_1, s_2) = M_L\phi(s_1, s_2), \\ &\qquad |M_L| \leq L, \end{aligned} \tag{5.24}$$

$$(\nabla^2 + p^2)\phi(s_1, s_2) = 0. \tag{5.25}$$

Thus in spherical polar coordinates (r, θ, φ),

$$\begin{aligned} &\phi_{\mathbf{p},\epsilon,L,M_L,\lambda_2,M_s}(s_1, s_2, s) \\ &\quad = a'(\epsilon, s_1, s_2)\delta_{s\lambda_2}c_L(pr)Y_{LM_L}(\theta, \varphi)\eta_{\lambda_2 M_s}, \end{aligned} \tag{5.26}$$

where $c_L(pr)$ is a spherical cylindrical function and Y_{LM_L} is a spherical harmonic. The parity $P = \beta\Pi_r$ of the state is $(-1)^L\beta_\epsilon$:

$$\begin{aligned} P\phi(\mathbf{r}, t) &\equiv \beta\phi(-\mathbf{r}, t) = (-1)^L\beta\phi(\mathbf{r}, t) \\ &= (-1)^L\beta_\epsilon\phi(\mathbf{r}, t). \end{aligned} \tag{5.27}$$

The intrinsic parity is β_ϵ; for bosons, $\beta_\epsilon = 1$, and $\beta_\epsilon = \epsilon$ for fermions. The coefficients $a'(\epsilon, s_1, s_2)$ are equal to $a(\epsilon, s_1, s_2)$ up to a normalization factor. The transformation to the original representation can be carried out in principle, thus

$$\begin{aligned} &\psi_{\mathbf{p},\epsilon,L,M_L,\lambda_2,M_s}(s_1, s_2, s, m_s) \\ &\quad = a'(\epsilon, s_1, s_2)\langle s_1, s_2, s, m_s|\, V^{-1}\,|s_1, s_2, \lambda_2, M_s\rangle \\ &\qquad\qquad \times c_L(pr)Y_{LM_L}(\theta, \varphi). \end{aligned} \tag{5.28}$$

The operator V^{-1} may be written in the form

$$V^{-1} = e^{\lambda_1 c_p\boldsymbol{\alpha}\cdot\mathbf{p}} = e^{c_p\boldsymbol{\Sigma}^{(1)}\cdot\mathbf{p}}\cdot e^{c_p\boldsymbol{\Sigma}^{(2)}\cdot\mathbf{p}}, \tag{5.29}$$

where $c_p = \theta_1/p$ is a constant. The operators $\exp(c_p\boldsymbol{\Sigma}^{(s)}\cdot\mathbf{p})$, $s = 1, 2$, may be expanded in terms of spin-projection operators, as shown by Shaw.[4] However, a direct derivation of $\psi(s_1, s_2)$ in the original representation is more practical. We remark that the coupling coefficients $a(\epsilon, s_1, s_2)$ of the various Poincaré components are the same in the original and the canonical representations.

C. The Coupling Coefficients

The equation $\gamma_0\phi = \epsilon\phi$ couples five components $\phi(s_1, s_2)$, $\phi(s_1 \pm \frac{1}{2}, s_2 \pm \frac{1}{2})$, and $\phi(s_1 \pm \frac{1}{2}, s_2 \mp \frac{1}{2})$. In principle the relation between any two Poincaré components can be obtained by using (5.18). However, it is more practical to use Eqs. (5.15). Denoting

$$\begin{aligned} \phi_{\mp\mp} &= \langle s_1, s_2|\,\gamma_0\,|s_1 \mp \tfrac{1}{2}, s_2 \mp \tfrac{1}{2}\rangle\,\phi(s_1 \mp \tfrac{1}{2}, s_2 \mp \tfrac{1}{2}), \\ \phi_{\mp\pm} &= \langle s_1, s_2|\,\gamma_0\,|s_1 \mp \tfrac{1}{2}, s_2 \pm \tfrac{1}{2}\rangle\,\phi(s_1 \mp \tfrac{1}{2}, s_2 \pm \tfrac{1}{2}), \end{aligned} \tag{5.30}$$

Eqs. (5.15) become, in favor of Eqs. (3.23)–(3.25),

$$\begin{aligned} &\phi_{--} + \phi_{++} + \phi_{-+} + \phi_{+-} = \epsilon\phi(s_1, s_2), \\ &x(\phi_{--} - \phi_{++}) + y(\phi_{-+} - \phi_{+-}) \\ &\qquad = \epsilon\left(\frac{\lambda_1 + 1}{\lambda_1}\right)xy\phi(s_1, s_2), \\ &x^2(\phi_{--} + \phi_{++}) + y^2(\phi_{-+} + \phi_{+-}) \\ &\qquad = \epsilon\left[\frac{\lambda_2(\lambda_2 + 1)}{\lambda_1(\lambda_1 + 1)}\right]\phi(s_1, s_2), \\ &x^3(\phi_{--} - \phi_{++}) + y^3(\phi_{-+} - \phi_{+-}) \\ &\qquad = \epsilon\left[\frac{\lambda_2(\lambda_2 + 1) + \lambda_1 + 1}{\lambda_1(\lambda_1 + 1)}\right]xy\phi(s_1, s_2). \end{aligned} \tag{5.31}$$

The solution is

$$\begin{aligned}\phi_{--}/(\lambda-\lambda_2)(\lambda+\lambda_1+2)(\lambda+\lambda_2+1)&=\phi_{++}/(\lambda_1-\lambda)(\lambda-\lambda_2+1)(\lambda+\lambda_2+2)\\&=\phi_{-+}/(\lambda_2+\mu)(\lambda_1+\mu+1)(\lambda_2-\mu+1)\\&=\phi_{+-}/(\lambda_2-\mu)(\lambda_1-\mu+1)(\lambda_2-\mu+1)\\&=\epsilon\phi(s_1,s_2)/2\lambda_1(2s_1+1)(2s_2+1).\end{aligned}\tag{5.32}$$

Now we take Le Couteur's representation[39] of γ_0:

$$\begin{aligned}2\lambda_1\langle s_1,s_2,\lambda_2|\gamma_0|s_1+\tfrac12,s_2+\tfrac12,\lambda_2\rangle&=[(\lambda+\lambda_2+2)(\lambda-\lambda_2+1)c_{++}]^{\frac12},\\2\lambda_1\langle s_1,s_2,\lambda_2|\gamma_0|s_1-\tfrac12,s_2-\tfrac12,\lambda_2\rangle&=[(\lambda+\lambda_2+1)(\lambda-\lambda_2)c_{--}]^{\frac12},\\2\lambda_1\langle s_1,s_2,\lambda_2|\gamma_0|s_1+\tfrac12,s_2-\tfrac12,\lambda_2\rangle&=c(\mu)[(\lambda_2+\mu+1)(\lambda_2-\mu)c_{+-}]^{\frac12},\\2\lambda_1\langle s_1,s_2,\lambda_2|\gamma_0|s_1-\tfrac12,s_2+\tfrac12,\lambda_2\rangle&=c(\mu+1)[(\lambda_2-\mu+1)(\lambda_2+\mu)c_{-+}]^{\frac12},\end{aligned}\tag{5.33}$$

where the c's are given by (3.40) and (3.41), and

$$c(\mu)=\begin{cases}(-1)^{[\lambda_2]}, & \text{if } \mu=-\tfrac12,-1,\\ 1, & \text{otherwise,}\end{cases}\tag{5.34}$$

and $[\lambda_2]$ is the integral part of λ_2. Denoting

$$\Phi(s_1,s_2)=[(2s_1+1)(2s_2+1)]^{-\frac12}\phi(s_1,s_2,\lambda_2),\tag{5.35}$$

from (5.32) and (5.33) we get

$$\Phi(s_1-\tfrac12,s_2-\tfrac12)=\epsilon[(\lambda_1+\lambda+2)/(\lambda_1-\lambda+1)]^{\frac12}\Phi(s_1,s_2),\tag{5.36}$$

$$\Phi(s_1+\tfrac12,s_2-\tfrac12)=\epsilon c(\mu)[(\lambda_1-\mu+1)/(\lambda_1+\mu+2)]^{\frac12}\Phi(s_1,s_2).\tag{5.37}$$

In particular for fermions, we get from the last relation

$$\phi(s_1,s_1+\tfrac12,\lambda_2)=(-1)^{[\lambda_2]}\epsilon\phi(s_1+\tfrac12,s_1,\lambda_2).\tag{5.38}$$

On the other hand, for bosons we have

$$\Phi(s_1,s_1+1)=(-1)^{\lambda_2}\epsilon[(\lambda_1+1)/(\lambda_1+2)]^{\frac12}\Phi(s_1+\tfrac12,s_1+\tfrac12)$$

and

$$\Phi(s_1+\tfrac12,s_1+\tfrac12)=\epsilon[(\lambda_1+2)/(\lambda_1+1)]^{\frac12}\Phi(s_1+1,s_1),$$

such that

$$\phi(s_1,s_1+1,\lambda_2)=(-1)^{\lambda_2}\phi(s_1+1,s_1,\lambda_2).\tag{5.39}$$

Equations (5.38) and (5.39) enable us to determine the factor in the matrix elements (3.53) and (3.54) of β. In fact, from these latter relations and the eigenvalue equation $\beta\phi=\beta_\epsilon\phi$, we get

$$\begin{aligned}\beta_\epsilon\phi(s_1,s_2,\lambda_2)&=\xi\phi(s_2,s_1,\lambda_2),\quad\text{for}\quad s_1\neq s_2,\\ \phi(s_1,s_1,\lambda_2)&=(-1)^{\lambda_2}\xi\phi(s_1,s_1,\lambda_2).\end{aligned}\tag{5.40}$$

Noticing that $\beta_\epsilon=1$ for bosons, $\beta_\epsilon=\epsilon$ for fermions, and comparing (5.38) and (5.39) with (5.40), we get

$$\xi=(-1)^{[\lambda_2]}.\tag{5.41}$$

Now, by successive application of (5.36), we get

$$\Phi\left(\frac{\lambda+\mu}{2},\frac{\lambda-\mu}{2}\right)=\epsilon^{\lambda-\lambda_2}\left[\frac{(\lambda_1-\lambda_2)!\,(\lambda_1+\lambda_2+2)!}{(\lambda_1-\lambda)!\,(\lambda_1+\lambda+2)!}\right]^{\frac12}\times\Phi\left(\frac{\lambda_2+\mu}{2},\frac{\lambda_2-\mu}{2}\right).\tag{5.42}$$

Further, by successive application of (5.37), for $s_1\geq s_2$ we get

$$\Phi\left(\frac{\lambda_2+\mu}{2},\frac{\lambda_2-\mu}{2}\right)=\epsilon^{\lambda_2-\mu}\left[\frac{(\lambda_1+\lambda_2+1)!\,(\lambda_1-\lambda_2+1)!}{(\lambda_1+\mu+1)!\,(\lambda_1-\mu+1)!}\right]^{\frac12}\Phi(\lambda_2,0).\tag{5.43}$$

Combining (5.42) and (5.43), for $s_1\geq s_2$ we get

$$\phi(s_1,s_2,\lambda_2)=\epsilon^{2s_2}C\left\{[(\lambda+1)^2-\mu^2]\binom{2\lambda_1+2}{\lambda_1-\lambda}\binom{2\lambda_1+2}{\lambda_1-\mu+1}\right\}^{\frac12}\times\phi(\lambda_2,0,\lambda_2),\tag{5.44}$$

where C is a normalization constant. For $s_1<s_2$ we use (5.40). $\phi(\lambda_2,0,\lambda_2)$ is just Wigner's canonical representation[5] of PG; we take it normalized. In order to normalize ϕ, we have to sum the series

$$Z^2=\sum_{\lambda=\lambda_2+1}^{N}\sum_{\mu=-\lambda_2}^{\lambda_2}f(\lambda,\mu),\tag{5.45}$$

where $N=\lambda_1+1$, and

$$f(\lambda,\mu)=(\lambda^2-\mu^2)\binom{2N}{N-\lambda}\binom{2N}{N-\mu},\tag{5.46}$$

such that

$$\phi^\dagger\phi=|C|^2Z^2.\tag{5.47}$$

In favor of the symmetry properties

$$\begin{aligned}f(\lambda,\mu)&=f(\lambda,-\mu)=f(-\lambda,\mu)\\&=-f(\mu,\lambda),\end{aligned}\tag{5.48}$$

we have

$$\sum_{\lambda=-\lambda_2}^{\lambda_2}\sum_{\mu=-\lambda_2}^{\lambda_2} f(\lambda,\mu)=0. \tag{5.49}$$

Hence,

$$\begin{aligned} Z^2 &= \tfrac{1}{2}\sum_{\lambda=-N}^{N}\sum_{\mu=-\lambda_2}^{\lambda_2} f(\lambda,\mu) \\ &= \tfrac{1}{2}\sum_{n=0}^{2N}\sum_{m=N-\lambda_2}^{N+\lambda_2}\Big[m(2N-m) \\ &\quad - n(2N-n)\binom{2N}{n}\binom{2N}{m}\Big] \\ &= 2^{2N-2}N(2N-1)[4J(N-1,\lambda_2)-J(N,\lambda_2)], \end{aligned} \tag{5.50}$$

where

$$J(N,M)=B(2N,N+M)-B(2N,N-M-1) \tag{5.51}$$

and

$$B(N,M)=\sum_{m=0}^{M}\binom{N}{m},\quad M\le N. \tag{5.52}$$

The truncated binomial series $B(N,M)$ is related to the hypergeometric series as follows[44]:

$$\begin{aligned} B(N,M) &= \binom{N}{M}F(-M,1;N-M+1;-1) \\ &= 2^N-\frac{1}{M+1}\binom{N}{M} \\ &\quad\times F(M-N+1,1;M+2;-1). \end{aligned} \tag{5.53}$$

Using the first form for $B(2N, N+M)$ and the second form for $B(2N, N-M)$ we get

$$\begin{aligned} &J(N,M)+2^{2N} \\ &\quad=\frac{N+M+2}{N+M+1}\binom{2N}{N-M} \\ &\quad\quad\times F(-N-M,1;N-M+1;-1). \end{aligned} \tag{5.54}$$

Hence

$$\begin{aligned} Z^2 &= [2^{2\lambda_1}(2\lambda_1+2)!/(\lambda_1-\lambda_2+1)!\,(\lambda_1+\lambda_2+2)!] \\ &\quad\times[2(\lambda_1-\lambda_2+1)(\lambda_1+\lambda_2+2)^2 \\ &\quad\times F(-\lambda_1-\lambda_2,1;\lambda_1-\lambda_2+1;-1) \\ &\quad-(\lambda_1+1)(2\lambda_1+1)(\lambda_1+\lambda_2+3) \\ &\quad\times F(-\lambda_1-\lambda_2-1,1;\lambda_1-\lambda_2+2;-1)]. \end{aligned} \tag{5.55}$$

D. Scalar Product

The Chakrabarti transformation (5.1) depends on p and ϵ. Thus states with different p_0 have different transformations. We write explicitly

$$\psi(\mathbf{r},t;p,\epsilon)=V^{-1}(p,\epsilon)\phi(\mathbf{r},t;p,\epsilon). \tag{5.56}$$

Although V is not unitary, it satisfies

$$\beta V^{\dagger}\beta=\beta V\beta=V^{-1}. \tag{5.57}$$

Consider two states $\psi_1(\mathbf{r},t;p,\epsilon)$ and $\psi_2(\mathbf{r},t;p',\epsilon')$, and let their Chakrabarti transforms be $\phi_1(\mathbf{r},t;p,\epsilon)$ and $\phi_2(\mathbf{r},t;p',\epsilon')$. Then their scalar product, defined by (4.95), becomes, in favor of (4.97),

$$\begin{aligned} (\psi_1,\psi_2) &= \frac{p_0+p_0'}{2m}\int\bar{\psi}_1(\mathbf{r},t;p,\epsilon)\psi_2(\mathbf{r},t;p',\epsilon')\,dV \\ &= \frac{p_0+p_0'}{2m}\int\bar{\phi}_1(\mathbf{r},t;p,\epsilon)V(p,\epsilon) \\ &\quad\times V^{-1}(p',\epsilon')\phi_2(\mathbf{r},t;p',\epsilon')\,dV. \end{aligned} \tag{5.58}$$

Now, since $(\psi_1,\psi_2)=0$ unless $p_0=p_0'$ (i.e., $p=p'$, $\epsilon=\epsilon'$), we may substitute $V(p,\epsilon)V^{-1}(p',\epsilon')=I\delta_{\epsilon\epsilon'}$ in the integrand, so that

$$\begin{aligned} (\psi_1,\psi_2) &= (p_0/m)\delta_{\epsilon\epsilon'} \\ &\quad\times\int\bar{\phi}_1(\mathbf{r},t;p,\epsilon)\phi_2(\mathbf{r},t;p',\epsilon')\,dV. \end{aligned} \tag{5.59}$$

For plane waves of definite spin

$$\begin{aligned} \phi_1(\mathbf{r},t;p,\epsilon) &= (2\pi)^{-\frac{3}{2}}e^{i(\mathbf{p}\cdot\mathbf{r}-\epsilon E_p t)}\phi_M, \\ \phi_2(\mathbf{r},t;p',\epsilon') &= (2\pi)^{-\frac{3}{2}}e^{i(\mathbf{p}'\cdot\mathbf{r}-\epsilon' E_{p'} t)}\phi_{M'}, \end{aligned} \tag{5.60}$$

where

$$\phi_M=Z^{-1}\sum_{s_1,s_2}\oplus\, a(\epsilon,s_1,s_2)\eta_{\lambda_2 M}(s_1,s_2), \tag{5.61}$$

and $\eta_{\lambda_2 M}(s_1,s_2)$ is the spin eigenfunction in the representation $D(s_1,s_2)$, assumed normalized,

$$\eta^{\dagger}_{\lambda_2 M}\eta_{\lambda_2 M}=I,$$

and Z is given by (3.55). If ϕ_M is normalized, such that

$$\phi^{\dagger}_M\phi_{M'}=\delta_{MM'}, \tag{5.62}$$

then

$$\begin{aligned} a(\epsilon,s_1,s_2) &= \zeta_{s_1 s_2}\epsilon^{2s_1}\Big[(2s_1+1)(2s_2+1) \\ &\quad\times\binom{2\lambda_1+2}{\lambda_1-s_1-s_2}\binom{2\lambda_1+2}{\lambda_1-s_1+s_2+1}\Big]^{\frac{1}{2}}, \end{aligned} \tag{5.63}$$

where

$$\zeta_{s_1 s_2}=\begin{cases}1, & \text{for } s_1\ge s_2,\\ (-1)^{[\lambda_2]}, & \text{for } s_1<s_2.\end{cases} \tag{5.64}$$

From (5.59) we get

$$(\psi_1,\psi_2)=\epsilon\beta_\epsilon\delta_{\epsilon\epsilon'}\delta_{MM'}(E_p/m)\delta^3(\mathbf{p}-\mathbf{p}'). \tag{5.65}$$

The scalar product is nonnegative for fermions ($\beta_\epsilon=\epsilon$), but not for bosons ($\beta_\epsilon=1$). Any arbitrary

[44] *Bateman Manuscript Project, Higher Transcendental Functions*, E. Erdélyi, Ed. (McGraw-Hill Book Company, New York, 1953), Vol. I, p. 87.

solution of the field equations may be written in the form

$$\Psi(\mathbf{r}, t) = \sum_{\epsilon=\pm 1} \sum_{M=-\lambda_2}^{\lambda_2} \int C_M(\mathbf{p}, \epsilon)\psi_M(\mathbf{r}, t; \mathbf{p}, \epsilon)\, d^3p$$
$$= (2\pi)^{-\frac{3}{2}} \sum_{\epsilon=\pm 1} \sum_{M=-\lambda_2}^{\lambda_2} \int C_M(\mathbf{p}, \epsilon)V^{-1}(p, \epsilon)\phi_M$$
$$\times\, e^{i(\mathbf{r}\cdot\mathbf{p}-\epsilon E_p t)}\, d^3p, \tag{5.66}$$

where $\psi_M(\mathbf{r}, t; \mathbf{p}, \epsilon)$ is normalized according to (5.65). Then the scalar product of two functions Ψ_1 and Ψ_2, of corresponding coefficients $C_M^{(1)}(\mathbf{p}, \epsilon)$ and $C_M^{(2)}(\mathbf{p}, \epsilon)$, is given by

$$(\Psi_1, \Psi_2) = \sum_{\epsilon=\pm 1} \sum_{M=-\lambda_2}^{\lambda_2} \epsilon\beta_\epsilon \int C_M^{(1)*}(\mathbf{p}, \epsilon)C_M^{(2)}(\mathbf{p}, \epsilon)(E_p/m)\, d^3p, \tag{5.67}$$

which corresponds to Bargmann–Wigner's[13] scalar product.

4. EXTREME RELATIVISTIC REPRESENTATION

A. Cini–Touschek Transformation

In order to discuss the extreme relativistic limit, we consider the generalization of the Cini–Touschek transformation[32] for $s = \frac{1}{2}$ to arbitrary spin

$$U = \exp(-\lambda_1\theta_2\boldsymbol{\gamma}\cdot\mathbf{p}/p), \tag{6.1}$$

where

$$\cos\theta_2 = p/E_p, \quad \sin\theta_2 = m/E_p. \tag{6.2}$$

This transformation was used by Mathews and Sankranarayanan[45] in connection with Kemmer's theory ($s = 0, 1$). The wavefunction in this representation is given by

$$\chi = U\psi. \tag{6.3}$$

The operators are transformed as follows:

$$\begin{aligned} E_p\gamma_0' &\equiv E_pU\gamma_0U^{-1} = p\gamma_0 + (m/p)\boldsymbol{\alpha}\cdot\mathbf{p}, \\ E_p\boldsymbol{\gamma}' &= p\boldsymbol{\gamma} + \xi_p(\boldsymbol{\gamma}\cdot\mathbf{p})\mathbf{p} + (im/p)\boldsymbol{\sigma}\wedge\mathbf{p}, \\ E_p\boldsymbol{\sigma}' &= p\boldsymbol{\sigma} + \xi_p(\boldsymbol{\sigma}\cdot\mathbf{p})\mathbf{p} - (im/p)\boldsymbol{\gamma}\wedge\mathbf{p}, \\ E_p\boldsymbol{\alpha}' &= E_p\boldsymbol{\alpha} - \mathbf{p}[\xi_p(\boldsymbol{\alpha}\cdot\mathbf{p}) + (m/p)\gamma_0], \end{aligned} \tag{6.4}$$

where $\xi_p = (E_p - p)/p^2$. Hence,

$$\gamma_\mu' p_\mu = \boldsymbol{\gamma}\cdot\mathbf{p} - \epsilon p\gamma_0 - (\epsilon m/p)(\boldsymbol{\alpha}\cdot\mathbf{p}), \tag{6.5}$$

$$P_0' = E_p(\boldsymbol{\alpha}\cdot\mathbf{p})/p, \tag{6.6}$$

$$\mathbf{P}' = (E_p/p)(\epsilon p\boldsymbol{\alpha} + i\boldsymbol{\sigma}\wedge\mathbf{p}) - \epsilon\xi_p(\boldsymbol{\alpha}\cdot\mathbf{p})\mathbf{p} + (m/p^2)\mathbf{p}(\boldsymbol{\gamma}\cdot\mathbf{p} - \epsilon p\gamma_0). \tag{6.7}$$

[45] P. M. Mathews and A. Sankranarayanan, Nuovo Cimento **34**, 101 (1964).

The field equations become

$$(\boldsymbol{\gamma}\cdot\mathbf{p} - \epsilon p\gamma_0)\chi = 0, \tag{6.8}$$

$$\boldsymbol{\alpha}\cdot\mathbf{p}\chi = \epsilon p\chi, \tag{6.9}$$

$$(\epsilon p\boldsymbol{\alpha} + i\boldsymbol{\sigma}\wedge\mathbf{p})\chi = \mathbf{p}\chi. \tag{6.10}$$

These are just the field equations for massless particles of energy $p_0' = \epsilon p$. From (6.9) and (6.10) one obtains

$$\boldsymbol{\sigma}\cdot\mathbf{p}\chi = \epsilon p\Gamma_5\chi \tag{6.11}$$

and

$$(\epsilon p\boldsymbol{\sigma} + i\boldsymbol{\alpha}\wedge\mathbf{p})\chi = \mathbf{p}\Gamma_5\chi, \tag{6.12}$$

which are the transformed dual equations (4.7).

B. Plane-Wave Solutions

This representation is particularly useful in obtaining the helicity solutions. We consider the plane-wave solutions and choose $p_1 = p_2 = 0$, $p_3 = p$. Equations (6.9) and (6.10) are satisfied by each Poincaré component $\chi(s_1, s_2)$ separately, and read now

$$\alpha_3\chi(s_1, s_2) = \epsilon\chi(s_1, s_2), \tag{6.13}$$

$$\sigma_3\chi(s_1, s_2) = \epsilon\Gamma_5\chi(s_1, s_2) = \epsilon S\chi(s_1, s_2), \tag{6.14}$$

where

$$S = \langle s_1, s_2|\,\Gamma_5\,|s_1, s_2\rangle = (s_1 - s_2)(s_1 + s_2 + 1)/\lambda_1(\lambda_1 + 1). \tag{6.15}$$

Also,

$$\begin{aligned} (\alpha_1 + i\epsilon\sigma_2)\chi(s_1, s_2) &= 0, \\ (\alpha_2 - i\epsilon\sigma_1)\chi(s_1, s_2) &= 0. \end{aligned} \tag{6.16}$$

From (6.13) and (6.14) we see that $\chi(s_1, s_2)$ is an eigenfunction of both of $\Sigma_3^{(1)} = (\lambda_1/2)(\sigma_3 + \alpha_3)$ and $\Sigma_3^{(2)} = (\lambda_1/2)(\sigma_3 - \alpha_3)$. It is suitable, therefore, to consider Bhabha's direct-product representation (3.13) and (3.16), such that

$$\begin{aligned} \Sigma_3^{(1)}\chi(s_1, m_1; s_2, m_2) &= m_1\chi(s_1, m_1; s_2, m_2), \\ \Sigma_3^{(2)}\chi(s_1, m_1; s_2, m_2) &= m_2\chi(s_1, m_1; s_2, m_2), \\ \boldsymbol{\Sigma}^{(1)^2}\chi(s_1, m_1; s_2, m_2) &= s_1(s_1 + 1)\chi(s_1, m_1; s_2, m_2), \\ \boldsymbol{\Sigma}^{(2)^2}\chi(s_1, m_1; s_2, m_2) &= s_2(s_2 + 1)\chi(s_1, m_1; s_2, m_2), \end{aligned} \tag{6.17}$$

where

$$2m_1 = \epsilon\lambda_1(S + 1), \quad 2m_2 = \epsilon\lambda_1(S - 1). \tag{6.18}$$

Further, denoting

$$\Sigma_\epsilon^{(s)} = \Sigma_1^{(s)} + i\epsilon\Sigma_2^{(s)}, \quad s = 1, 2, \tag{6.19}$$

Eqs. (6.16) read as follows:

$$\begin{aligned} \Sigma_\epsilon^{(1)}\chi(s_1, m_1; s_2, m_2) &= 0, \\ \Sigma_{-\epsilon}^{(2)}\chi(s_1, m_1; s_2, m_2) &= 0. \end{aligned} \tag{6.20}$$

Now, in the direct-product representation,

$$\Sigma_k^{(1)} = S_k(s_1) \otimes I(s_2), \quad \Sigma_k^{(2)} = I(s_1) \otimes S_k(s_2) \quad (6.21)$$

and

$$\chi(s_1, m_1; s_2, m_2) = \eta_{s_1 m_1} \otimes \eta_{s_2 m_2}, \quad (6.22)$$

where $I(s)$ is the $(2s + 1) \times (2s + 1)$ unit matrix, and $\mathbf{S}(s)$ is the spin in the $(2s + 1)$-dimensional irreducible representation $D(s)$ of $O(3)$, such that

$$\mathbf{S}^2(s)\eta_{sm} = s(s + 1)\eta_{sm},$$
$$S_3\eta_{sm} = m\eta_{sm}, \quad (6.23)$$

and S_3 is diagonal. Now, $\Sigma_\epsilon^{(1)}$ and $\Sigma_\epsilon^{(2)}$ are the ladder operators increasing or decreasing m by one unit, according as $\epsilon = 1$ or $\epsilon = -1$, respectively. Thus Eq. (6.20) implies that

$$m_1 = \epsilon s_1, \quad m_2 = -\epsilon s_2. \quad (6.24)$$

From (6.18) and (6.24) we get

$$\lambda = s_1 + s_2 = \lambda_1, \quad \mu = s_1 - s_2 = \lambda_1 S. \quad (6.25)$$

Thus each solution with a definite energy sign ϵ and a definite helicity

$$\Sigma = \langle \lambda_1(\boldsymbol{\sigma} \cdot \mathbf{p})/p \rangle = \langle \lambda_1 \sigma_3 \rangle = \epsilon\lambda_1 S = \epsilon(s_1 - s_2) \quad (6.26)$$

corresponds to a single Poincaré component

$$\chi_\epsilon(s_1, \epsilon s_1; s_2, -\epsilon s_2),$$

where

$$2s_1 = \lambda_1 + \epsilon\Sigma, \quad 2s_2 = \lambda_1 - \epsilon\Sigma. \quad (6.27)$$

Since $|s_1 - s_2| \leq \lambda_2$ [cf. Eq. (3.12)], then

$$\delta \leq |\Sigma| \leq \lambda_2, \quad (6.28)$$

where $\delta = 0$ for bosons and $\delta = \frac{1}{2}$ for fermions.

In the direct-sum representation, we get from the Clebsch–Gordan theorem that

$$\chi_{\epsilon,\Sigma}(s_1, \epsilon s_1; s_2, -\epsilon s_2) = \sum_{j=|\Sigma|}^{\lambda_1} c\left(\frac{\lambda_1 + \epsilon\Sigma}{2}, \frac{\lambda_1 - \epsilon\Sigma}{2}, j; \frac{\epsilon}{2}(\lambda_1 + \epsilon\Sigma), \frac{\epsilon}{2}(\lambda_1 - \epsilon\Sigma), \Sigma\right)\eta_{j,\Sigma}(s_1, s_2). \quad (6.29)$$

The inverse is only one term,

$$\chi_{j,\Sigma}(s_1, s_2) = c(s_1, s_2, j; \epsilon s_1, -\epsilon s_2, \Sigma) \times \chi_\epsilon(s_1, \epsilon s_1; s_2, -\epsilon s_2), \quad (6.30)$$

where $|\Sigma| \leq j \leq \lambda_1$, $c(s_1, s_2, j; m_1, m_2, m)$ is a Clebsch–Gordan coefficient.

We see that, in the extreme relativistic limit $\psi \to \chi$, the wavefunction is the direct product of two wavefunctions of two massless particles of spins s_1 and s_2 with opposite maximum helicities ϵs_1 and $-\epsilon s_2$, respectively. In this limit, several spin states (6.30) occur, while in the nonrelativistic limit only $j = \lambda_2$ occurs.

C. Self-Consistency of the Solution

We have considered so far only Eqs. (6.9) and (6.10), and Eq. (6.11), which is derived from them. These equations are satisfied by each Poincaré component separately. Since we have seen that we have only one Poincaré component, we should make sure that this solution is consistent with the additional condition (6.8). This condition follows from the energy–momentum equations, and connects the different admissible Poincaré components. Now, for $p_1 = p_2 = 0$, $p_3 = p$, Eq. (6.8) reads

$$(\gamma_0 - \epsilon\gamma_3)\chi_\epsilon = 0.$$

Hence, in the spinor notation [cf. (3.18)],

$$\gamma^{11}\chi_- = 0, \quad \gamma^{22}\chi_+ = 0, \quad (6.31)$$

where $\chi_\pm$ correspond to $\epsilon = \pm 1$ respectively. From Eqs. (6.8)–(6.11) it is obvious that

$$\chi_{-\epsilon,\Sigma} = \beta\chi_{\epsilon,\Sigma}. \quad (6.32)$$

Using (3.53) and (3.54), and dropping an arbitrary phase factor ξ, we get

$$\chi_{-\epsilon,\Sigma}(s_1, s_2) = \chi_{\epsilon,\Sigma}(s_2, s_1) = \chi_\epsilon(s_2, \epsilon s_2; s_1, -\epsilon s_1). \quad (6.33)$$

It suffices, therefore, to consider the self-consistency of one of the two equations (6.31). We take $\gamma^{11}\chi_- = 0$, and notice that

$$\chi_- = \chi_-(s_1, -s_1; s_2, s_2). \quad (6.34)$$

Since we have only one Poincaré component, $\gamma^{11}\chi_- = 0$ yields

$$\langle s_1', m_1'; s_2', m_2' | \gamma^{11} | s_1, -s_1; s_2, s_2 \rangle \times \chi(s_1, -s_1; s_2, s_2) = 0, \quad (6.35)$$

for all admissible values of s_1', s_2', m_1', and m_2', subject to the condition $s_1 + s_2 = \lambda_1$. Bhabha[3] has shown that the nonvanishing matrix elements of γ^{11} are

$$\langle s_1', m_1'; s_2', m_2' | \gamma^{11} | s_1, m_1; s_2, m_2 \rangle = [c(s_1', s_2'; s_1, s_2)]^{\frac{1}{2}} \langle s_1', m_1' | A | s_1, m_1 \rangle \times \langle s_2', m_2' | A | s_2, m_2 \rangle, \quad (6.36)$$

where

$$|s_1' - s_1| = |s_2' - s_2| = m_1 - m_1' = m_2 - m_2' = \tfrac{1}{2}. \tag{6.37}$$

Here $c(s_1', s_2'; s_1, s_2)$ is the corresponding c given by (3.40) and (3.41), and $\langle s_1', m_1' | A | s_1, m_1 \rangle$ satisfies the following recurrence relation:

$$\begin{aligned}\left\langle s - \frac{\sigma}{2}, m - \tfrac{1}{2} \middle| A \middle| s, m \right\rangle & [(s + m + 1)(s - m)]^{\frac{1}{2}} \\ &= \left\langle s - \frac{\sigma}{2}, m + \tfrac{1}{2} \middle| A \middle| s, m + 1 \right\rangle \\ &\quad \times \left[\left(s + m + \frac{1 - \sigma}{2}\right)\left(s - m + \frac{1 - \sigma}{2}\right)\right]^{\frac{1}{2}},\end{aligned} \tag{6.38}$$

$\sigma = \pm 1$. From (6.37) and (6.38) it follows directly that

$$\langle s_1 - \tfrac{1}{2}, m_1' | A | s_1, -s_1 \rangle = \langle s_2 - \tfrac{1}{2}, m_2' | A | s_2, s_2 \rangle = 0 \tag{6.39}$$

for all m_1' and m_2'. This proves that the matrix elements of γ^{11} vanish identically for all values concerned, except $s_1' - s_1 = s_2' - s_2 = \frac{1}{2}$. We are thus left with the matrix element

$$\begin{aligned}\langle s_1 + \tfrac{1}{2}, -(s_1 + \tfrac{1}{2}); s_2 + \tfrac{1}{2}, s_2 - \tfrac{1}{2} | \gamma^{11} | s_1, -s_1; s_2, s_2 \rangle \\ = (c_{++})^{\frac{1}{2}} \langle s_1 + \tfrac{1}{2}, -(s_1 + \tfrac{1}{2}) | A | s_1, -s_1 \rangle \\ \times \langle s_2 + \tfrac{1}{2}, s_2 - \tfrac{1}{2} | A | s_2, s_2 \rangle.\end{aligned} \tag{6.40}$$

This matrix element vanishes also, because $c_{++} = 0$ for $\lambda = s_1 + s_2 = \lambda_1$. This completes the proof that Eq. (6.8) is satisfied automatically by the solution of the preceding section.

In considering the Cini–Touschek representation, we have discussed implicitly the generalization of Stepanovskii's equations[7] for massless particles. Stepanovskii's equations correspond to the representation $R_5(\lambda_1, \lambda_1)$, where

$$2s_1 = (1 + \epsilon)\lambda_1, \quad 2s_2 = (1 - \epsilon)\lambda_1. \tag{6.41}$$

This is also Hammer–Good's theory[16] but with the required supplementary conditions. We have seen that $\gamma_\mu p_\mu \psi = 0$ is satisfied for $p_0 = \epsilon p$, but in general it is not sufficient. We should consider the field equations $(S_{\mu\nu} p_\nu - p_\mu)\psi = 0$. This is true for any arbitrary representation $R_5(\lambda_1, \lambda_2)$. λ_1 is the maximal spin and λ_2 is the maximal helicity. A state of helicity Σ belongs to the representation $D(s_1, s_2)$ of hLG, with s_1 and s_2 as given by (6.27). However, this is of rather formal importance. The two known massless particles, the neutrino ($s = \frac{1}{2}$) and the photon ($s = 1$), belong to the special representation (6.41).

D. Relation to the Canonical Representation

A direct normalization of χ to yield (5.62) with the help of (5.58) is impractical. In fact, although U is unitary, its unitarity is destroyed by β; thus

$$\beta U^\dagger \beta = \beta U^{-1} \beta = U,$$

such that $\bar{\psi}\psi = \bar{\chi} U^{-2} \chi$. A direct evaluation of this expression is tedious. Instead, we consider the relation

$$\chi = U\psi = UV^{-1}\phi, \tag{6.42}$$

and simplify UV^{-1} such that $\chi^\dagger \chi$ may be related directly to $\phi^\dagger \phi$. For this purpose let

$$A = UV^{-1} = e^{-\lambda_1 \theta_2 \gamma \cdot \mathbf{n}} e^{\lambda_1 \theta_1 \alpha \cdot \mathbf{n}}, \tag{6.43}$$

where $\mathbf{n} = \mathbf{p}/p$. Let further

$$B = e^{\epsilon \lambda_1 c \alpha \cdot \mathbf{n}} e^{-\lambda_1 b \gamma \cdot \mathbf{n}} e^{\lambda_1 a(\gamma + \epsilon\alpha)\cdot \mathbf{n}}. \tag{6.44}$$

We choose a, b, and c such that $A = B$. Then from (5.18) and (6.9) we get

$$\begin{aligned}e^{-\lambda_1 c}\chi = e^{-\epsilon\lambda_1 c\alpha\cdot\mathbf{n}}\chi &= e^{-\epsilon\lambda_1 c\alpha\cdot\mathbf{n}} B\phi \\ &= e^{-\lambda_1 b\gamma\cdot\mathbf{n}} e^{\lambda_1 a(\gamma+\epsilon\alpha)\cdot\mathbf{n}}\phi \\ &= e^{-\lambda_1 b\gamma\cdot\mathbf{n}}\phi.\end{aligned}$$

Hence,

$$\chi = e^{\lambda_1 c} e^{-\lambda_1 b \gamma \cdot \mathbf{n}} \phi. \tag{6.45}$$

b and c turn out to be real, such that

$$\chi^\dagger \chi = e^{2\lambda_1 c} \phi^\dagger \phi. \tag{6.46}$$

In order to choose $A = B$, it is sufficient to consider the equality

$$A\gamma_\mu A^{-1} = B\gamma_\mu B^{-1} \tag{6.47}$$

for all μ, such that

$$[AB^{-1}, \gamma_\mu] = [AB^{-1}, S_{\mu\nu}] = 0. \tag{6.48}$$

By Schur's lemma, AB^{-1} is a multiple of the identity in each irreducible representation. It remains then to prove the equality of A and B after determining a, b, and c. After some lengthy manipulations, using (5.8), (5.9), and (6.4), it is found that

$$a = p/E_p, \quad b = \pi/2, \quad \mathbf{e}^c = E_p/m. \tag{6.49}$$

In order to show that $A = B$ for these values, we expand A and B in ascending powers of p/m, and consider the first approximation. Let $\theta_2 = \pi/2 - \theta_0$. Then

$$\cos\theta_0 = m/E_p \approx 1 - p^2/2m^2 \approx 1,$$

$$\sin\theta_0 = p/E_p \approx p/m.$$

Further,

$$\cosh\theta_1 = E_p/m \approx 1, \quad \sinh\theta_1 = \epsilon p/m.$$

Hence, in a first approximation,

$$a = \theta_0 = \epsilon\theta_1 = p/m,$$

while

$$c = \log (E_p/m) \approx \log (1 + p^2/2m^2) \approx p^2/2m^2$$

may be neglected, such that $\exp (\lambda_1 \epsilon c\boldsymbol{\alpha} \cdot \mathbf{n}) \approx 1$, and

$$B = e^{-(\lambda_1\pi/2)\boldsymbol{\gamma}\cdot\mathbf{n}} e^{\lambda_1\theta_0(\boldsymbol{\gamma}+\epsilon\boldsymbol{\alpha})\cdot\mathbf{n}}$$
$$\approx e^{-(\lambda_1\pi/2)\boldsymbol{\gamma}\cdot\mathbf{n}}[1 + \lambda_1\theta_0(\boldsymbol{\gamma} + \epsilon\boldsymbol{\alpha}) \cdot \mathbf{n}].$$

Also

$$A = e^{-(\pi/2-\theta_0)\lambda_1\boldsymbol{\gamma}\cdot\mathbf{n}} e^{\lambda_1\epsilon\theta_0\boldsymbol{\alpha}\cdot\mathbf{n}}$$
$$\approx e^{-(\lambda_1\pi/2)\boldsymbol{\gamma}\cdot\mathbf{n}}[1 + \lambda_1\theta_0(\boldsymbol{\gamma} + \epsilon\boldsymbol{\alpha}) \cdot \mathbf{n}],$$

which proves that $A = B$.

From (6.45), (6.46), and (6.49), it follows that

$$\chi = (E_p/m)^{\lambda_1} e^{-(\lambda_1\pi\boldsymbol{\gamma}\cdot\mathbf{p}/2p)}\phi \tag{6.50}$$

and

$$\chi^\dagger\chi = (E_p/m)^{2\lambda_1}, \tag{6.51}$$

provided that ϕ is normalized: $\phi^\dagger\phi = 1$. The normalized helicity solution (6.22) is then

$$\psi_{s_1m_1,s_2m_2}(\mathbf{r}, t; p, \epsilon) = (2\pi)^{-\frac{3}{2}}(E_p/m)^{\lambda_1} e^{i(pz-\epsilon E_p t)} e^{\lambda_1\theta_2\gamma_3}(\eta_{s_1m_1} \otimes \eta_{s_2m_2}). \tag{6.52}$$

On the other hand, if we express the solutions of the field equations in terms of normalized χ, such that $\chi^\dagger\chi = 1$,

$$\Psi(\mathbf{r}, t) = \sum_{\epsilon=\pm1} \sum_{M=-\lambda_2}^{\lambda_2} \int b_M(\mathbf{p}, \epsilon) \times U^{-1}(p, \epsilon)\chi_M e^{i(\mathbf{p}\cdot\mathbf{r}-\epsilon E_p t)}\, d^3p, \tag{6.53}$$

then the scalar product of two functions Ψ_1 and Ψ_2, with corresponding coefficients $b_M^{(1)}(\mathbf{p}, \epsilon)$ and $b_M^{(2)}(\mathbf{p}, \epsilon)$, is given by

$$(\Psi_1, \Psi_2) = \sum_{\epsilon=\pm1} \sum_{M=-\lambda_2}^{\lambda_2} \epsilon\beta_\epsilon \int b_M^{(1)*}(\mathbf{p}, \epsilon) \times b_M^{(2)}(\mathbf{p}, \epsilon)\left(\frac{m}{E_p}\right)^{2\lambda_1-1} d^3p. \tag{6.54}$$

This is analogous to Bargmann–Wigner's scalar product.[13]

E. Relation to the Foldy–Wouthuysen Representation

The Foldy–Wouthuysen transformation[33,46] and its relation to the Chakrabarti transformation[31,47,48] were discussed in the fusion representation

$$\gamma_\mu = \frac{1}{2s}\sum_{n=1}^{2s}\gamma_\mu^{(n)}$$

where

$$[\gamma_\mu^{(n)}, \gamma_\nu^{(n')}] = 0, \quad \text{for} \quad n \neq n',$$

and

$$\gamma_\mu^{(n)}\gamma_\nu^{(n)} + \gamma_\nu^{(n)}\gamma_\mu^{(n)} = -2\delta_{\mu\nu}I.$$

We establish here the connection between this transformation and the Chakrabarti and Cini–Touschek transformations without referring to a particular representation. The Foldy–Wouthuysen transformation is

$$W = \exp (\lambda_1\theta_3\boldsymbol{\gamma} \cdot \mathbf{p}/p), \tag{6.55}$$

where

$$\cos\theta_3 = m/E_p, \quad \sin\theta_3 = p/E_p. \tag{6.56}$$

Under this transformation the energy–momentum operators become

$$P_0' = WP_0W^{-1} = E_p\gamma_0,$$
$$\mathbf{P}' = \epsilon\gamma_0\mathbf{p} + E_p\{\mathbf{A}^{(\epsilon)} - (\mathbf{A}^{(\epsilon)} \cdot \mathbf{p})\mathbf{p}/E_p(E_p + m)\}, \tag{6.57}$$

where $\mathbf{A}^{(\epsilon)}$ is given by (5.17). The wavefunction becomes

$$\zeta = W\psi. \tag{6.58}$$

One verifies easily that ζ satisfies the same equations as ϕ. Now, from (6.2) and (6.56), one finds that

$$\theta_3 = (\pi/2) - \theta_2. \tag{6.59}$$

Hence

$$W = U \exp (\lambda_1\pi\boldsymbol{\gamma} \cdot \mathbf{p}/2p). \tag{6.60}$$

From (6.50) we get finally

$$\zeta = WU^{-1}\chi = e^{(\lambda_1\pi\boldsymbol{\gamma}\cdot\mathbf{p}/2p)}\chi$$
$$= (E_p/m)^{\lambda_1}\phi. \tag{6.61}$$

This is a generalization of the known relation for Dirac's theory[48] ($s = \frac{1}{2}$).

ACKNOWLEDGMENTS

The author wishes to express his thanks to M. I. Hessein and A. F. Ghaleb for reading a preliminary draft of the manuscript and for checking the calculations.

[46] D. L. Pursey, Nucl. Phys. **53**, 174 (1964).
[47] C. G. Bollini and J. J. Giambiagi, Nuovo Cimento **21**, 107 (1961).
[48] R. H. Good, Jr., and M. E. Rose, Nuovo Cimento **24**, 864 (1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Inversion Problem with Separable Potentials*†

ROBERT L. MILLS
The Ohio State University, Columbus, Ohio
AND
JOHN F. READING‡
Laboratory for Nuclear Science and Physics Department, Massachusetts Institute of Technology Cambridge, Massachusetts

(Received 24 March 1967)

The problem of finding a potential, expressible as the sum of a finite number of separable terms, to fit a given phase shift at all energies is solved quite generally for nonrelativistic scattering within a single channel. The most general solution is found, and the necessary restrictions on the phase shift and the role of bound states are studied. The minimum number of separable terms needed to fit a given phase shift in a given channel is found to be determined, normally, by

$$1 + \max_{x_2 > x_1} \{[\delta(x_2)/\pi] - [\delta(x_1)/\pi]\},$$

where x_1 and x_2 range through all positive energies, and $[y]$ is the largest integer less than or equal to y. Our method differs from that of Kh. Chadan [Nuovo Cimento **10**, 892 (1958); **47**, 510 (1967)] who has solved what is, in essence, the same problem, in that it involves spherical Bessel transforms throughout, rather than the use at each stage of a representation determined by the solution of a previous stage.

I. INTRODUCTION

Interest in separable potentials[1] has been revived recently by Lovelace,[2] who pointed out that a two-body T matrix dominated by poles or resonances is of a separable form, or is the sum of separable forms. In low-energy nucleon–nucleon scattering, for example, we have such a situation,[3] in that the deuteron pole dominates the ${}^3S_1$ phase shift, so that we have some justification for the excellent work done by Mitra[4] and others on the three-body problem using a separable-potential approach. It is not immediately obvious, however, that writing the interaction as a sum of separable potentials is really justifiable,[5] but because such good results for the three-body binding energy have been obtained by this approach, we would like to understand why this is so. Any Hermitian potential, local or nonlocal, can of course be written as a sum of separable terms, although the number of terms is, in general, infinite.

The inverse problem is to find the energy-independent potential, given the phase shifts at all energies, a problem which has been studied very thoroughly for local potentials,[6] and investigated in some detail for separable potentials.[7–9] The phase shifts given experimentally are probably too inaccurate and given in too limited an energy range for such a program to be carried out with precision, but if one assumes some experimental error and a reasonable high-energy behavior, one can test the sensitivity of the potential and physically observed phenomena to these, which, in itself, is of interest. In most cases it is desirable to have analytic expressions for potentials, but these can always be obtained by fitting the potential data obtained from such a program. However, the main results of this paper are that they lead to a deeper insight into the properties of such potentials, their analytic structure,[3] and their nonuniqueness.

We restrict our attention to the case in which scattering is within uncoupled channels, i.e., in which the conservation laws reduce the Schrödinger equation to uncoupled radial equations. This means that the S matrix is diagonal and is given entirely by the phase shifts for the different channels. Thus spin–orbit interactions could be included (if the target is spinless, say), but not tensor forces. Each term in a separable potential can act only within a single channel, so that every channel is an independent

* Part of this work was done at the University of Birmingham, Birmingham, England, and was presented as part of a Ph.D. thesis by one of us (J. F. R.). A part was also done at the State University of New York, Stony Brook, N.Y.

† Work supported in part by the Atomic Energy Commission, in part by the Department of Scientific & Industrial Research of the British Government, and in part by the National Science Foundation.

‡ Present address: Theoretical Physics Division, Aere, Harwell, Didcot, Berks., England.

[1] Y. Yamaguchi, Phys. Rev. **95**, 1628 (1954); Y. Yamaguchi and Y. Yamaguchi, *ibid.* **95**, 1635 (1954).

[2] C. Lovelace, Phys. Rev. **135**, B1225 (1964).

[3] J. F. Reading, "Perey Effect in the Deuteron" (submitted to Physics Letters).

[4] A. N. Mitra, Nucl. Phys. **32**, 529 (1962).

[5] F. Tabakin, Phys. Rev. **137**, B75 (1965); Ph.D. thesis, Dept. of Physics, Massachusetts Institute of Technology.

[6] R. Jost and W. Kohn, Phys. Rev. **87**, 977 (1952); **88**, 382 (1952); I. M. Gel'fand and B. M. Levitan, Dokl. Akad. Nauk SSSR **77**, 557 (1951); Izv. Akad. Nauk SSSR **15**, 309 (1951). See particularly the thorough review paper of R. G. Newton, J. Math. Phys. **1**, 319 (1960).

[7] M. Gourdin and A. Martin, Nuovo Cimento **8**, 699 (1958).

[8] Kh. Chadan, Nuovo Cimento **10**, 892 (1958); **47**, 510 (1967).

[9] M. Bolsterli and J. MacKenzie, Physics **2**, 141 (1965).

problem. Different channels may, of course, be related, as by rotation symmetry or isotopic spin symmetry.

Let $\mathbf{P}$ denote the total linear momentum of the two-particle system, and let the set of other conserved quantities, which must all commute with $\mathbf{P}$, be denoted by τ. We must be able to separate the Schrödinger equation down to a radial equation in the magnitude p of the relative momentum $\mathbf{p}$, and so we must require that τ also commute with the operator $\mathbf{p}^2$. This will suffice for most of our discussion, but in fact one would usually expect a radial equation in r to be valid also, in which case τ must commute with $\mathbf{r}^2$ as well; if it commutes with both $\mathbf{r}^2$ and $\mathbf{p}^2$, it also commutes with $(\mathbf{r} \times \mathbf{p})^2$, and l is a good quantum number. We assume for convenience that this is the case, although our requirement of uncoupled scattering channels is, in fact, equivalent to saying merely that $\mathbf{P}$, τ, and $\mathbf{p}^2$ are a complete set of commuting observables for the problem, and that $\mathbf{P}$ and τ commute with the Hamiltonian.

II. THE ONE-TERM SEPARABLE POTENTIAL

For the sake of clarity, we start by reviewing the simple one-term case for spinless particles,[7,9] and then go on to the more general N-term case.[8]

We take as our Hamiltonian

$$H = \frac{\mathbf{P}^2}{2M} + \frac{\mathbf{p}^2}{2m} + V, \tag{1}$$

where M is the total mass and m the reduced mass, $\mathbf{P}$ the total momentum and $\mathbf{p}$ the relative momentum. The interaction V is assumed for now to consist of a single separable term in each angular-momentum channel:

$$\langle \mathbf{P}, \mathbf{r}| V |\mathbf{P}', \mathbf{r}'\rangle = \delta(\mathbf{P} - \mathbf{P}') \sum_{l,m} \frac{1}{\mu_l} u_l(r) Y_l^m(\theta, \varphi) u_l^*(r') Y_l^{m*}(\theta', \varphi'). \tag{2}$$

Galilean invariance demands that μ_l and $u_l(r)$ be independent of $\mathbf{P}$. Time-reversal invariance would require $u_l(r)$ to be real. The coefficient is here taken as μ_l^{-1} (μ_l is real, of course) because the case in which the coefficient is infinite is nontrivially interesting and useful; while it corresponds in a sense to a potential of infinite strength, its effect is simply to impose an orthogonality condition on the wavefunction. The zero-coefficient case, on the other hand, is of no interest at all. There would, of course, be no loss of generality in restricting μ_l to the values $0, \pm 1$.

It is convenient to use the observables $\mathbf{P}$, $\mathbf{L}^2$, $\mathbf{L}_z$, and $\mathbf{p}^2$ to define our representation. The state is represented by

$$\langle \mathbf{P}, l, m, p \mid \psi\rangle = \delta(\mathbf{P} - \mathbf{P}_1)\delta_{l,l_1}\delta_{m,m_1}\psi(p), \tag{3}$$

and the radial function $\psi(p)$, the spherical Bessel transform of the usual radial function $\psi(r)$, in a system for which $\hbar = 2m = 1$, satisfies

$$p^2\psi(p) + \alpha u(p) = k^2\psi(p), \tag{4}$$

where

$$\mu\alpha = \int_0^\infty u^*(p)\psi(p)p^2\,dp. \tag{5}$$

Here, $u(p)$ is the spherical Bessel transform of $u_l(r)$, k^2 is the center-of-mass energy, and the index l has been dropped, since the treatment in no way depends on the value of l. The only effect of l is to govern the low-energy behavior of $u(p)$ and of the phase shift for short-range potentials.

We speak of the "normal" case, both here and for the N-term problem, as that in which $u(r)$ is sufficiently smooth that $p^\nu |u(p)| \to 0$ as $p \to \infty$, with $\nu > \frac{3}{2}$, and $u_l(r)$ is of sufficiently short range that $u(p)$ is continuous and differentiable everywhere. This includes the case $\mu = 0$. A variety of conclusions can be drawn for different restrictions of varying degrees of stringency on the behavior of $u(r)$, and we make no effort to be exhaustive, although from time to time, we refer to the "normal" behavior. Actually the scattering problem can be solved and phase shifts obtained, even though $u(r)$ may be a generalized function not expressible as an ordinary function, provided $u(p)$ is any of a wide variety of ordinary functions, including discontinuous and unbounded ones, for which the integral in Eq. (10) is convergent for complex z.

The phase shift is, in the normal case, a continuous function going to zero at infinite energy, but in fact the inversion problem can also be solved for any of a wide range of functions for the phase shift, with restrictions, however, to be discussed in more detail later.

The solution of Eq. (4) may be taken as

$$\psi(p) = Z\delta(p - k) - \frac{\alpha u(p)}{p^2 - k^2 - i\epsilon}, \tag{6}$$

(here ϵ is a positive infinitesimal), where α can be determined by substituting (6) into Eq. (5):

$$\mu\alpha = Zk^2u^*(k) - \alpha\int_0^\infty \frac{|u(p)|^2\,p^2\,dp}{p^2 - k^2 - i\epsilon}. \tag{7}$$

The convergence of $\int^\infty |u|^2\,dp$ and of $\int_0 p^2\,|u|^2\,dp$ would thus ensure the existence of solutions of Eq. (4) at all positive energies. It would suffice, for instance, if $u(r)$ were bounded for finite r, behaved like

$r^{-\alpha}$, with $\alpha < \frac{5}{2}$ near the origin, and like $r^{-\beta}$, with $\beta > \frac{3}{2}$, at infinity.

The phase shift is determined by the asymptotic behavior of $\psi(r)$, and hence by the behavior of $\psi(p)$ near the singularity at $p = k$. For the usual definition of the phase shift δ, this turns out to be

$$\psi(p) \approx Z\left[\delta(p-k) + \frac{1}{\pi}\frac{e^{i\delta}\sin\delta}{p-k-i\epsilon}\right], \tag{8}$$

for $p \approx k$, so that, combining (6), (7), and (8), we may write

$$e^{i\delta}\sin\delta = -\tfrac{1}{2}\pi k\,|u(k)|^2 \Big/ \left[\mu + \int_0^\infty \frac{|u(p)|^2\,p^2\,dp}{p^2-k^2-i\epsilon}\right]. \tag{9}$$

This can at best be solved for $|u(p)|^2$, the phase of $u(p)$ being arbitrary if time-reversal invariance is not required; it is clear, on the other hand, that time-reversal invariance can be imposed, i.e., by choosing $u(p)$ to be real, if the problem is soluble at all. To solve for $|u|^2$, we note that the phase shift is simply minus the argument of the complex number in the denominator on the right which, in turn, is the boundary value of a function of the energy which is analytic in the upper half-plane, that is, $e^{i\delta}A(k^2+i\epsilon)$ is real, where

$$A(z) \equiv \mu + \int_0^\infty \frac{|u(p)|^2\,p^2\,dp}{p^2-z}. \tag{10}$$

The imaginary part of $\ln A(k^2+i\epsilon)$ is thus determined by the phase shift, which allows $A(z)$ to be determined,[9] while $\frac{1}{2}\pi k\,|u(k)|^2$ is then the imaginary part of $A(k^2+i\epsilon)$. [Provided $\mu \neq 0$, the function $(1/\mu)A(k^2-i\epsilon)$ bears the same relation to the phase shift as the Jost function $f_l(k)$ defined for local potentials.[6]]

It is convenient at this point to be using the energy p^2 as our variable, and to absorb a factor $(\frac{1}{2}\pi k)^{\frac{1}{2}}$ into the definition of the potential function; we let

$$\begin{aligned} v(x) &\equiv (\tfrac{1}{2}\pi k)^{\frac{1}{2}}u^*(k), \quad x = k^2 \geq 0, \\ &\equiv 0, \qquad\qquad\quad x < 0. \end{aligned} \tag{11}$$

[We use the complex conjugate $u^*(k)$ here for notational convenience because, when we generalize to the case of potentials of more than one term, the elements $u^*(k)$ constitute a column matrix.] Thus

$$A(z) = \mu + \frac{1}{\pi}\int_0^\infty \frac{|v(x')|^2\,dx'}{x'-z}, \tag{12}$$

where z denotes a complex energy, with real and imaginary parts x and y, respectively.

There is an obvious symmetry between the upper and lower half-planes, and we restrict our attention to the upper. For real z we take the limiting value:

$$\begin{aligned} A(x) &\equiv \lim_{\epsilon\to 0+} A(x+i\epsilon) \\ &= \mu + \frac{1}{\pi}\mathcal{P}\int_0^\infty \frac{|v(x')|^2\,dx'}{x'-x} + i\,|v(x)|^2, \end{aligned} \tag{13}$$

so that, from Eq. (9),

$$e^{i\delta(x)}\sin\delta(x) = -\frac{|v(x)|^2}{A(x)}. \tag{14}$$

Various facts are now easy to see:

(i) $$\operatorname{Im} A(x) \geq 0. \tag{15}$$

[This follows also from the reality of the phase shift $\delta(x)$ on the left-hand side of (14).]

(ii) For $y > b$,

$$\operatorname{Im} A(x+iy) = \frac{y}{\pi}\int_0^\infty \frac{|v(x')|^2\,dx'}{|x'-z|^2} > 0. \tag{16}$$

(iii) For $x < 0$, $A(x)$ is real and greater than μ; in fact,

$$\frac{d^r}{dx^r}[A(x)-\mu] = \frac{r!}{\pi}\int_0^\infty \frac{|v(x')|^2\,dx'}{(x'-x)^{r+1}} > 0, \quad r = 0, 1, 2, \cdots. \tag{17}$$

(iv) If, for some $x > 0$, $A(x) = 0$, then $v(x) = 0$ (Eq. 13) and

$$\frac{dA(x)}{dx} = \frac{1}{\pi}\int_0^\infty \frac{|v(x')|^2\,dx'}{(x'-x-i\epsilon)^2} > 0; \tag{18}$$

the derivative is real and positive, provided $v(x')$ is smoothly varying at the point x; it is sufficient, for example, that $|v(x')|^2/|x'-x|^{1+\nu}$ be bounded, for some $\nu > 0$.

Equation (14) determines the phase shift only to within an arbitrary multiple of π, and the fact that the value of $A(x)$ is restricted to the upper-half of the complex plane [Eq. (15)] in no way restricts the possible values of the phase shift. Indeed, it is a bit difficult to define the phase shift unambiguously in a satisfactory way. If one defines an "absolute phase shift," say, by counting the nodes of the wavefunction $\psi(r)$, one finds that as a function of energy it can have discontinuities of π, 2π, etc., since for a nonlocal potential the slope (or lowest-order nonvanishing derivative) at the origin can pass through zero, or the wavefunction can pass through a condition of zero

value and slope at points away from the origin. These discontinuities are not observable in scattering measurements, and are not obvious in the behavior of $A(x)$, which behaves quite innocently at such points. The problem of fitting such absolute phase shifts is thus very difficult, and we do not attempt it. In the normal case, however, the S matrix $e^{2i\delta}$ is a continuous function of the energy, so that, if we are not concerned with the absolute phase shift, the phase shift can be defined by requiring that it be a continuous function of the energy. In the normal case, moreover, $e^{2i\delta}$ tends to unity for very large energy, and the phase shift can thus be defined to have the limiting value zero for large energies.

Let us also define a modified phase shift $\eta(x)$ by the condition that it lie between $-\pi$ and zero. More precisely, for all real x, let

$$\eta(x) \equiv -\arg A(x), \tag{19}$$

with

$$-\pi \leq \eta(x) \leq 0. \tag{20}$$

Thus $\eta(x)$ is equal to 0 or $-\pi$ for negative x, and for positive x differs from the phase shift $\delta(x)$ by some multiple of π. It exhibits a discontinuity only when $A(x)$ passes through the value zero. At such a point the derivative of $A(x)$ is real and positive [Eqs. (17) and (18)], so that the discontinuity is $+\pi$. If this occurs for negative x, it corresponds to a bound solution of Eq. (4); at positive x it corresponds to a normalized solution,[10] or "bound state of positive energy" in addition to the normal scattering state.[11] Since $A(x)$ is real and monotonically increasing for negative x there can be, at most, one bound state, and this can occur only if μ is negative and sufficiently small (potential strength $1/\mu$ sufficiently large). If $\eta(x)$ has a discontinuity of this sort for positive x, then the phase shift $\delta(x)$ passes through a multiple of π with negative slope with respect to the energy. A discontinuity in $\eta(x)$ of other than $+\pi$ corresponds to a singularity in $A(x)$ and hence in $v(x)$. This may be tolerable, though inelegant, unless the discontinuity is $-\pi$, in which case $A(x)$ has a simple pole, and $|v(x)|^2$ would have to have a delta-function singularity as a function of the energy x. This is not possible, so that we must exclude a phase shift which increases continuously through a value $n\pi$, with n an integer.

Usually $\eta(+\infty)$ has the same value as $\eta(-\infty)$, namely $-\arg\mu$ (either 0 or $-\pi$), but if $\mu = 0$, then η will have a limiting value determined by the limiting behavior of $A(x)$. In this case $\eta(-\infty) = 0$ and, as $x \to +\infty$, η can approach any value from 0 to $-\pi$ or oscillate without any limit, although in the normal case it takes on the value $-\pi$. In any case, therefore, we must have

$$\eta(+\infty) \leq \eta(-\infty). \tag{21}$$

Similarly $\eta(0)$ usually has the value $-\arg A(0)$, (either 0 or $-\pi$), although if $A(0) = 0$ this is not defined, and $\eta(0)$ may take on other values. In this case (which may be thought of as the case of a zero-energy bound state) the limiting value $\eta(0-)$ for negative x is $-\pi$; the limiting value $\eta(0+)$ for positive x may be anything, though for short-range potentials it is typically, but not necessarily, $-\frac{1}{2}\pi$ for $l = 0$, and zero for $l > 0$. It is also possible for $\eta(x)$ to display a negative discontinuity at $x = 0$, if $A(x)$, and with it $v(x)$, is singular there, corresponding to an appropriate long-range tail in $u(r)$, but (just as for $x \neq 0$) the discontinuity cannot be $-\pi$, since this would require a delta-function singularity in $|v(x)|^2$.

In summary, the necessary restrictions on $\eta(x)$, apart from the restriction to the values 0 and $-\pi$ for negative x, are simply these, that there be no discontinuities of $-\pi$, and that $\eta(+\infty) \leq \eta(-\infty)$. The corresponding restrictions on the phase shift $\delta(x)$, defined for positive x only and without regard to whether or not there is a bound state, are that $\delta(x)$ may not increase (with increasing x) through a value equal to a multiple of π and that, if $\delta(x)$ is a multiple of π at $x = 0$, and is increasing for small x [this corresponds to $\eta(0+) = -\pi$], then $\delta(x)$ must approach a multiple of π from above as $x \to \infty$ [that is, $\eta(+\infty) = -\pi$]. There is no restriction on the number of times $\delta(x)$ may decrease through multiples of π. The above restrictions on the phase shift can be summarized in the simple requirement that $\delta(0+) - \delta(\infty) > (N-1)\pi$, where N is the number of zeros (not counting $x = 0, \infty$) in $\sin\delta(x)$. If we restrict our attention to the normal case with finite strength ($\mu \neq 0$), and no zero-energy bound state [$A(0) \neq 0$], then we find a modified version of Levinson's theorem,[10,12] namely, that the number of bound states (zero or one) is equal to $[\eta(0+) - \eta(+\infty)]/\pi$. The difference $(1/\pi)[\delta(0) - \delta(\infty)]$ is thus at least as great

[10] A. Martin, Nuovo Cimento 7, 607 (1958).

[11] For a negative-energy solution of (4) ($k^2 = x_0 < 0$), we see that $p^2 - x_0$ cannot vanish, so we must put $Z = 0$ in (6) and (7), and hence must have $A(x_0) = 0$. Conversely, if $A(x_0) = 0$ for $x_0 < 0$, then Eq. (6), with Z set equal to zero, yields a solution of (4) and, furthermore, $\psi(p)$ is normalizable, since the norm involves a more convergent integral than that in Eq. (7), which must be assumed convergent. If $A(x_0) = 0$ for some positive x_0 [a point of discontinuity for $\eta(x)$], then, by (13), $v(x_0) = 0$ and, hence, (7) is satisfied for arbitrary α and Z. The case $Z = 0$ gives the normalizable solution [Eq. (6) with $u(k_0) = 0$], or bound state of positive energy; while the case $\alpha = 0$ is the same as the solution of the unperturbed problem. The ratio α/Z has a definite finite limit, though, as $x \to x_0$, so that the "bound-state" solution never appears as x varies continuously through x_0; the finiteness of this limit can be seen to be a consequence of Eq. (18), which says that dA/dx does not vanish at x_0.

[12] N. Levinson, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. **25**, No. 9 (1949).

as the number of bound states in this case; that is, if $\delta(0) = \delta(\infty)$, there is no bound state.

The procedure for constructing the potential to fit a given phase shift satisfying the above restrictions is now straightforward.[9] For each positive x we construct $\eta(x)$ by changing $\delta(x)$ by a multiple of π in such a way that (20) is satisfied [$\eta(x)$ may have discontinuities of $+\pi$]. For x negative, $\eta(x)$ must take on only the values 0 and $-\pi$, it can have at most one discontinuity of $+\pi$ corresponding to a bound state, and it must satisfy $\eta(-\infty) \geq \eta(+\infty)$. Thus, if

$$\eta(+\infty) \neq -\pi,$$

we must take $\eta(x) = 0$ for all $x < 0$. If $\eta(+\infty) = -\pi$, then $\eta(x)$ can be 0 for all $x < 0$, or $-\pi$ for all $x < 0$, or may be $-\pi$ for $x < x_0$ and 0 for $x > x_0$, where x_0 is an arbitrary negative energy. If $\eta(0+) = -\pi$, we must have

$$\eta(x) = -\pi \tag{22}$$

for all $x < 0$.

We now must find a function $A(x)$ which is the boundary value of a function analytic in the upper half-plane, and which satisfies Eq. (19), or equivalently [since by Eq. (16), $A(z)$ can have no zeros in the upper half-plane], the equation

$$\eta(x) = -\text{Im ln } A(x). \tag{23}$$

If $\eta(\infty) = \eta(-\infty)$, then normally we have $\mu \neq 0$ and this limiting value of η must be zero or π, corresponding to positive or negative μ. The magnitude of μ is arbitrary, and the solution of (23) can be written

$$\ln \frac{A(x)}{\mu} = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\eta(x') - \eta(-\infty)}{x' - x - i\epsilon}\, dx'. \tag{24}$$

If the integral in Eq. (24) diverges at $+\infty$ either because $\eta(\infty) \neq \eta(-\infty)$ or because $\eta(x)$ does not approach its limiting value fast enough, then $\mu = 0$, and a single subtraction yields the convergent expression

$$\ln \frac{A(x)}{A(x_1)} = -\frac{x - x_1}{\pi}\int_{-\infty}^{\infty} \frac{\eta(x')\, dx'}{(x' - x - i\epsilon)(x' - x_1 - i\epsilon)}, \tag{25}$$

where x_1 may be any point for which $\eta(x_1)$ is continuous, and where $A(x_1)$ has arbitrary modulus, but its phase must be $-\eta(x_1)$ to be consistent with (23). It is of course convenient to choose $x_1 < 0$, so that $A(x_1)$ is real and positive (since $\mu = 0$). It is also clear that Eq. (25) is a quite satisfactory form for the case $\mu \neq 0$ also, provided x_1 does not coincide with a bound-state energy.

Once $A(x)$ is known, the modulus of the potential is determined from Eq. (13):

$$|v(x)|^2 = \text{Im } A(x), \tag{26}$$

which is guaranteed to be nonnegative by the restriction that $\eta(x)$ lies between $-\pi$ and 0. The phase of $v(x)$ is quite arbitrary, but $v(x)$ may be taken as real, as discussed above, in order to satisfy time-reversal invariance. Equation (11) now gives $u(k)$ which, in turn, determines $u(r)$ by spherical Bessel transform (the only place the value of l enters).

III. THE MANY-TERM POTENTIAL

A. Properties of the Phase Shift

We are now in a position to ask for the most general phase shift in a given channel which can be fitted by an N-term separable potential, or, equivalently, for the minimum number of terms needed to fit a given phase shift. We shall do this, in fact, by first studying the properties of the phase shift for the N-term case, and then explicitly constructing the most general N-term potential which fits a given phase shift.

Extending Eq. (2), we take as our basic nonlocal potential

$$\langle \mathbf{P}, \mathbf{r}|\, V\, |\mathbf{P}', \mathbf{r}'\rangle = \delta(\mathbf{P} - \mathbf{P}') \sum_{l,m} \sum_{i=1}^{N_l} \frac{1}{\mu_{li}} u_{li}(r) Y_l^m(\theta, \varphi) u_{li}^*(r') Y_l^{m*}(\theta', \varphi'). \tag{27}$$

We again restrict our attention to a single channel, described by the momentum-representation wavefunction $\psi(p)$ defined by Eq. (3), which satisfies equations, analogous to (4) and (5), of the form (again dropping the index l)

$$p^2\psi(p) + \sum_{i=1}^{N} \alpha_i u_i(p) = k^2\psi(p), \tag{28}$$

with

$$\mu_i\alpha_i = \int_0^{\infty} u_i^*(p)\psi(p)p^2\, dp. \tag{29}$$

The solution may be taken in a form similar to Eq. (6):

$$\psi(p) = Z\delta(p - k) - \frac{\sum \alpha_i u_i(p)}{p^2 - k^2 - i\epsilon}, \tag{30}$$

where here the coefficients α_i are determined by the coupled equations

$$\mu_i\alpha_i = Zk^2u_i^*(k) - \sum_j \alpha_j \int_0^{\infty} \frac{u_i^*(p)u_j(p)p^2\, dp}{p^2 - k^2 - i\epsilon}. \tag{31}$$

Introducing the energy variable x, which is equal to k^2, and following the pattern of Eqs. (11) and (12), we define a column vector $\mathbf{v}(x)$ with components

$$v_i(x) \equiv (\tfrac{1}{2}\pi k)^{\frac{1}{2}} u_i^*(k), \quad x \geq 0,$$
$$\equiv 0, \quad x < 0, \tag{32}$$

and an analytic matrix function $\mathbf{A}(z)$ with components given by

$$A_{ij}(z) \equiv \mu_i \delta_{ij} + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v_i(x')v_j^*(x')}{x' - z}\, dx'. \tag{33}$$

In a matrix notation, we write

$$\mathbf{A}(z) = \boldsymbol{\mu} + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathbf{v}(x')\bar{\mathbf{v}}(x')}{x' - z}\, dx', \tag{34}$$

where the row matrix $\bar{\mathbf{v}}(x')$ is the Hermitian conjugate of $\mathbf{v}(x')$, and $\boldsymbol{\mu}$ is the diagonal matrix whose elements are μ_i. The boundary value of $\mathbf{A}(z)$ for real z is given by

$$\mathbf{A}(x) \equiv \lim_{\epsilon \to 0+} \mathbf{A}(x + i\epsilon)$$
$$= \mathbf{T}(x) + i\mathbf{v}(x)\bar{\mathbf{v}}(x), \tag{35}$$

where $\mathbf{T}(x)$ is the Hermitian part of the matrix $\mathbf{A}(x)$:

$$\mathbf{T}(x) \equiv \boldsymbol{\mu} + \frac{1}{\pi}\mathcal{P}\int_{-\infty}^{\infty} \frac{\mathbf{v}(x')\bar{\mathbf{v}}(x')}{x' - x}\, dx'. \tag{36}$$

Solving Eqs. (30) and (31) for $\psi(p)$, and comparing with Eq. (8), we find that the phase shift for a given value of x satisfies an equation analogous to (14), namely

$$e^{i\delta} \sin \delta = -\bar{\mathbf{v}}\mathbf{A}^{-1}\mathbf{v}. \tag{37}$$

According to Eq. (35), the anti-Hermitian part of $\mathbf{A}(x)$ is the rank-one matrix $\mathbf{v}\bar{\mathbf{v}}$. Because of this special property of $\mathbf{A}(x)$, the right-hand side of Eq. (37) can be re-expressed in terms of the determinant

$$\bar{\mathbf{v}}\mathbf{A}^{-1}\mathbf{v} = \frac{\bar{\mathbf{v}}\mathbf{T}^{\mathrm{adj}}\mathbf{v}}{\det \mathbf{A}}, \tag{38}$$

so that from (37), the phase shift is just minus the phase of $\det \mathbf{A}(x)$, up to a multiple of π.

We therefore must study the properties of $\det \mathbf{A}(x)$, and the associated analytic function $\det \mathbf{A}(z)$, which we denote by $D(z)$, of which $\det \mathbf{A}(x)$ is the boundary value:

$$D(z) \equiv \det \mathbf{A}(z), \tag{39}$$
$$D(x) \equiv \lim_{\epsilon \to 0+} D(x + i\epsilon)$$
$$= \det \mathbf{A}(x). \tag{40}$$

Note that the asymptotic value of $D(z)$ is simply $\det \boldsymbol{\mu} = \prod_i \mu_i$.

We find first that the matrix $\mathbf{A}(z)$ has several properties, including generalizations of those [Eqs. (15)–(18)] found in the case of the one-term potential. Let us use the expression $\mathcal{K}\mathbf{X}$ to denote the Hermitian part of any matrix $\mathbf{X}$, and $\mathcal{A}\mathbf{X}$, the anti-Hermitian part. Furthermore, let $\mathbf{X} > 0$ signify that $\mathbf{X}$ is a positive-definite Hermitian form, and $\mathbf{X} \geq 0$ that $\mathbf{X}$ is positive-semidefinite. Then,

(i)
$$\mathcal{A}\mathbf{A}(x) = \mathbf{v}(x)\bar{\mathbf{v}}(x)$$
$$\geq 0. \tag{41}$$

(ii) For $y > 0$,

$$\mathcal{A}\mathbf{A}(x + iy) = \frac{y}{\pi}\int \frac{\mathbf{v}(x')\bar{\mathbf{v}}(x')}{|x' - z|^2}\, dx'$$
$$> 0. \tag{42}$$

[Note that $\bar{\mathbf{u}}\mathbf{v}(x) = 0$ (all x) would imply that the functions $v_i(x)$ are linearly dependent, and reduce the problem to that of an $(N-1)$-term potential; thus $\mathcal{A}\mathbf{A}$ is positive-definite and not merely semidefinite.]

(iii) For $x < 0$, $\mathbf{A}(x)$ is Hermitian and

$$\frac{d^r}{dx^r}[\mathbf{A}(x) - \boldsymbol{\mu}] = \frac{r!}{\pi}\int_{-\infty}^{\infty} \frac{\mathbf{v}(x')\bar{\mathbf{v}}(x')}{(x' - x)^{r+1}}\, dx'$$
$$> 0, \quad r = 0, 1, 2, \cdots. \tag{43}$$

(iv) For $y > 0$,

$$D(x + iy) \neq 0, \tag{44}$$

since $D = 0$ would imply the existence of a vector $\mathbf{u}$ such that $\mathbf{Au} = 0$, which would in turn imply

$$\mathrm{Im}\ \bar{\mathbf{u}}\mathbf{A}\mathbf{u} = \bar{\mathbf{u}}(\mathcal{A}\mathbf{A})\mathbf{u}$$
$$= 0, \tag{45}$$

contradicting Eq. (42).

(v) If, for some positive or negative x, $D(x) = 0$, then there exists a vector $\boldsymbol{\alpha}$ such that $\mathbf{A}(x)\boldsymbol{\alpha} = 0$ [see Eq. (31)], and hence a normalizable solution (30) with $Z = 0$. This is a normal bound state if $x < 0$, and a quasibound state if $x > 0$. Furthermore, any bound state of negative energy must correspond to a zero in $D(x)$. A higher-order zero in $D(x)$ can be seen to correspond to a degenerate bound state.

(vi) If, for some x_κ (positive or negative), $D(x_\kappa) = 0$, then there exist one or more orthonormal vectors $\mathbf{u}_{\kappa\alpha}$ such that

$$\mathbf{T}(x_\kappa)\mathbf{u}_{\kappa\alpha} = 0 \tag{46}$$

and (for $x_\kappa > 0$ in particular):

$$\bar{\mathbf{u}}_{\kappa\alpha}\mathbf{v}(x_\kappa) = 0; \tag{47}$$

and furthermore, for each such vector $\mathbf{u}_{\kappa\alpha}$,

$$e_{\kappa\alpha} \equiv \bar{\mathbf{u}}_{\kappa\alpha}\left(\frac{d\mathbf{T}}{dx}\right)_{x_\kappa}\mathbf{u}_{\kappa\alpha} = \frac{1}{\pi}\int \frac{|\bar{\mathbf{u}}_{\kappa\alpha}\mathbf{v}(x')|^2}{(x' - x_\kappa)^2}\, dx'$$
$$> 0. \tag{48}$$

(vii) It follows from this that for any nonsingular vector function $\mathbf{w}(x)$, the poles in the function $\bar{\mathbf{w}}\mathbf{T}^{-1}\mathbf{w}$ are of first order only, with real positive residues, given by $\sum_\alpha |\bar{\mathbf{w}}\mathbf{u}_{\kappa\alpha}|^2/e_{\kappa\alpha}$.

We proceed by studying the effect on the phase shift of adding an additional term to the potential which we shall call the potential increment. We find that the resultant increment to the phase shift, which we call a "partial phase shift," is a function of similar form to the phase shift for a one-term potential alone, and furthermore, that the addition of an arbitrary partial phase shift, subject to restrictions similar to those for the one-term phase shift, can always be effected by the addition of a single term to the potential.

Let the original potential be the N-term potential discussed above, and let the potential increment (in terms of the energy variable x) be $(1/\mu_0)v_0^*(x)v_0(x')$. Let the increment to the phase shift be $\delta_0(x)$, and the resultant phase shift, $\delta'(x)$:

$$\delta'(x) = \delta(x) + \delta_0(x). \tag{49}$$

The new $(N+1) \times (N+1)$ matrices corresponding to $\mathbf{A}(z)$ and $\mathbf{T}(x)$ shall be called $\mathbf{A}'(z)$ and $\mathbf{T}'(x)$, with $\mathbf{A}'(x)$ again denoting the boundary value approaching the real axis from above. The components of $\mathbf{T}'(x)$ include a new N-component column vector which we call $\mathbf{t}(x)$:

$$\mathbf{T}' \equiv \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{t}' & T_{00} \end{pmatrix}, \tag{50}$$

while similarly, $\mathbf{A}'(z)$ can be written

$$\mathbf{A}' \equiv \begin{pmatrix} \mathbf{A} & \mathbf{a} \\ \bar{\mathbf{b}} & A_{00} \end{pmatrix}, \tag{51}$$

where $\bar{\mathbf{b}}(z)$ is a row matrix, the Hermitian conjugate of a column matrix $\mathbf{b}(z)$ related to $\mathbf{a}(z)$ by

$$\mathbf{b}(z) = \mathbf{a}(z^*). \tag{52}$$

For real x, we have, extending (and in addition to) Eq. (35),

$$\mathbf{a}(x) = \mathbf{t}(x) + i\mathbf{v}(x)v_0^*(x), \tag{53}$$

$$\bar{\mathbf{b}}(x) = \bar{\mathbf{t}}(x) + iv_0(x)\bar{\mathbf{v}}(x), \tag{54}$$

$$A_{00}(x) = T_{00} + i\,|v_0(x)|^2. \tag{55}$$

Let the determinant of $\mathbf{A}'(z)$ be denoted by $D'(z)$, with $D_0(z)$ defined by

$$D'(z) \equiv D(z)D_0(z). \tag{56}$$

The asymptotic value of $D'(z)$ is $\mu_0 \prod_1^N \mu_i$, and that for $D_0(z)$ is simply μ_0. The partial phase shift $\delta_0(x)$ is given in the standard way by minus the phase of $D_0(x)$. To proceed, we find from the form of Eq. (51) that

$$\begin{aligned} D'(x) &= \det \mathbf{A}'(x) \\ &= A_{00} \det \mathbf{A} - \bar{\mathbf{b}}\mathbf{A}^{\mathrm{adj}}\mathbf{a} \\ &= D(x)[A_{00} - \bar{\mathbf{b}}\mathbf{A}^{-1}\mathbf{a}], \end{aligned} \tag{57}$$

so that

$$D_0(x) = A_{00}(x) - \bar{\mathbf{b}}(x)\mathbf{A}^{-1}(x)\mathbf{a}(x). \tag{58}$$

The inverse of $\mathbf{A}(x)$ can be expressed, because of the form of Eq. (35), as

$$\mathbf{A}^{-1} = \mathbf{T}^{-1} - [1/(\beta - i)]\mathbf{T}^{-1}\mathbf{v}\bar{\mathbf{v}}\mathbf{T}^{-1}, \tag{59}$$

where $\beta(x)$ is defined by

$$\beta = \bar{\mathbf{v}}\mathbf{T}^{-1}\mathbf{v}. \tag{60}$$

Substituting the expressions (53)–(55), (59) for $\mathbf{a}$, $\mathbf{b}$, A_{00}, and $\mathbf{A}^{-1}$ into Eq. (58), we obtain

$$\begin{aligned} D_0(x) = T_{00} &- \bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t} \\ &+ [(\beta + i)/(\beta^2 + 1)]\,|v_0 - \bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{v}|^2. \end{aligned} \tag{61}$$

At those points x_κ (both positive and negative in general) at which $D(x)$ vanishes, $D_0(x)$ may have poles, appearing in the term $\bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t}$, but not in the other terms, since $\mathbf{T}^{-1}\mathbf{v}$ is, in fact, finite at such points. By property (vii) of the matrix function $\mathbf{T}(x)$, there may be a simple pole, with negative residue d_κ, say, in $D_0(x)$ at each such point, and hence an additional imaginary term $i\pi d_\kappa \delta(x - x_\kappa)$ ($d_\kappa \geq 0$) not appearing explicitly in Eq. (61). A higher-order zero may appear in $D(x)$, but there can be only a first-order pole in $D_0(x)$, and indeed $D_0(x)$ may be finite or zero there, so that $D'(x)$ may exhibit a zero of the same order or of an order higher or lower by one. If $\mathbf{T}$ becomes singular at a point for which $D(x) \neq 0$, the singularities in the separate terms of Eq. (61) cancel, leaving $D_0(x)$ nonsingular.

We can now write

$$D_0(x) = \tau(x) + i\rho(x), \tag{62}$$

where

$$\tau(x) \equiv T_{00} - \bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t} + \frac{\beta}{\beta^2 + 1}\,|\zeta|^2, \tag{63}$$

$$\rho(x) = \frac{1}{\beta^2 + 1}\,|\zeta|^2 + \pi \sum_\kappa d_\kappa \delta(x - x_\kappa), \tag{64}$$

and

$$\zeta(x) = v_0 - \bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{v}. \tag{65}$$

As discussed above, x_κ are the zeros of $D(x)$, and d_κ are the residues of the poles in $\bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t}$ at these points, which may be positive or zero. The imaginary part $\rho(x)$ of $D_0(x)$ is thus nonnegative, and $D_0(x)$ may be

written

$$D_0(x) = \mu_0 + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\rho(x')}{x' - x - i\epsilon}\, dx'. \quad (66)$$

This is made possible by the fact that $D(z)$ and $D'(z)$ are analytic and without zeros in the upper half-plane, so that $D_0(z)$ has the same properties; it has a cut on the positive-real axis, and by (64) the discontinuity across the real axis is positive-imaginary, just as for the amplitude $A(z)$ in the one-term potential case [Eqs. (12), (13)].

The properties of the partial phase shift $\delta_0(x)$, which is simply minus the phase of $D_0(x)$ (up to a multiple of π), can now be found by straightforward extension of the reasoning in the one-term case.

If we define a modified partial phase shift $\eta_0(x)$, in exact analogy with $\eta(x)$ in the one-term case [Eqs. (19), (20)], by

$$\eta_0(x) = -\arg D_0(x); \quad (67)$$

$$-\pi \leq \eta_0(x) \leq 0, \quad (68)$$

we see that for positive x, $\eta_0(x)$ differs from $\delta_0(x)$ by an integral multiple of π, and for negative x it is restricted to the values 0 and $-\pi$. This modified partial phase shift may have discontinuities of $+\pi$ at any value of x, corresponding to zeros in $D_0(x)$ [where $\delta_0(x)$ passes through $n\pi$ with negative slope] but may have discontinuities of $-\pi$ [$\delta_0(x)$ passing through $n\pi$ with positive slope], corresponding to poles in $D_0(x)$, only at points where $D(x)$ has a zero. The restriction $\eta_0(+\infty) \leq \eta_0(-\infty)$ follows from Eq. (66) precisely as in the case of $\eta(x)$ with the one-term potential.

For negative x, it is clear from (66) that $D_0(x)$ is real and of positive slope everywhere except at poles. This implies that $D_0(x)$ has either the same number of zeros as poles, or else one more or less. Since a pole of D_0 removes a first-order zero of D, or reduces the order of a higher-order zero by one, we see that, if we count an nth-order zero as n coincident simple zeros (corresponding to n degenerate bound states) then $D'(x)$ has either the same number of zeros as $D(x)$ for negative x, or else one more or one less. It also follows that $D(x)$ can have at most N zeros for negative x, by successive applications of the above result, starting with the $N = 0$ case.

B. Construction of the Potential Increment

If we are given a partial phase shift corresponding to an $\eta_0(x)$ which satisfies the restrictions that it have discontinuities of $-\pi$ only at points for which $D(x)$ has a zero, and that $\eta_0(+\infty) \leq \eta_0(-\infty)$, we can in fact construct a potential function $v_0(x)$ which will yield this partial phase shift, although the procedure is not so straightforward as in the one-term case. We first construct $D_0(x)$ by means of an equation of exactly the same form as (25), with an additive constant which will give rise to an arbitrary positive scale factor. The imaginary part of $D_0(x)$ is $\rho(x)$, which has the form given in Eq. (64). Since β is known from the N-term potential we started with, we can obtain $|\zeta(x)|^2$ and the coefficients d_κ. We may now choose the function $\zeta(x)$ to have arbitrary phase, or to be real if time-reversal invariance is to be imposed. Neither $v_0(x)$ nor $\mathbf{t}(x)$ is known yet, however, so another step is needed. Consider the analytic vector function $\bar{\mathbf{b}}(z)\mathbf{A}^{-1}(z)$. For real z, using the expressions (54) and (59) for $\bar{\mathbf{b}}$ and $\mathbf{A}^{-1}$, we obtain

$$\bar{\mathbf{b}}(x)\mathbf{A}^{-1}(x) = \bar{\mathbf{t}}\mathbf{T}^{-1} + [(\beta + i)/(\beta^2 + 1)]\zeta\bar{\mathbf{v}}\mathbf{T}^{-1}. \quad (69)$$

Again we must include properly the contribution at points where $\mathbf{A}(x)$ is singular, that is, where $D(x)$ is zero (where $\mathbf{T}$ is singular but $\mathbf{A}$ is not, the singularities on the right cancel). According to property (vi) of the matrix $\mathbf{T}$ [Eqs. (46)–(48)], near each singular point x_κ, $\mathbf{T}^{-1}(x)$ has the dominant behavior:

$$\mathbf{T}^{-1}(x) \approx \sum_\alpha \frac{e_{\kappa\alpha}^{-1}\mathbf{u}_{\kappa\alpha}\bar{\mathbf{u}}_{\kappa\alpha}}{x - x_\kappa}. \quad (70)$$

Then the singularity in Eq. (69) has the form

$$\bar{\mathbf{t}}\mathbf{T}^{-1} \approx \frac{\bar{\mathbf{w}}_\kappa}{x - x_\kappa + i\epsilon}, \quad (71)$$

where

$$\bar{\mathbf{w}}_\kappa \equiv \sum_\alpha e_{\kappa\alpha}^{-1} f_{\kappa\alpha}\bar{\mathbf{u}}_{\kappa\alpha} \quad (72)$$

and

$$f_{\kappa\alpha} \equiv \bar{\mathbf{t}}(x_\kappa)\mathbf{u}_{\kappa\alpha}. \quad (73)$$

Here, the $f_{\kappa\alpha}$ are related to the residue d_κ in $\bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t}$: since

$$\bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t} \approx \sum_\alpha \frac{e_{\kappa\alpha}^{-1}|\bar{\mathbf{t}}\mathbf{u}_{\kappa\alpha}|^2}{x - x_\kappa}, \quad (74)$$

we have

$$d_\kappa = \sum_\alpha e_{\kappa\alpha}^{-1}|f_{\kappa\alpha}|^2; \quad (75)$$

beyond this the $f_{\kappa\alpha}$ are undetermined if $\mathbf{t}$ is not known. We can now write the discontinuity across the real axis of the analytic function $\bar{\mathbf{b}}(z)\mathbf{A}^{-1}(z)$. Calling it $\Delta[\bar{\mathbf{b}}\mathbf{A}^{-1}]$, we have

$$\frac{1}{2i}\Delta[\bar{\mathbf{b}}\mathbf{A}^{-1}] = \frac{1}{\beta^2 + 1}\zeta\bar{\mathbf{v}}\mathbf{T}^{-1} - \pi\sum_\kappa \bar{\mathbf{w}}_\kappa\delta(x - x_\kappa) \quad (76a)$$

$$\equiv \bar{\boldsymbol{\chi}}(x) - \pi\sum_\kappa \bar{\mathbf{w}}_\kappa\delta(x - x_\kappa), \quad (76b)$$

where $\bar{\mathbf{w}}_\kappa$ is given by Eq. (72), and $\bar{\boldsymbol{\chi}}(x)$ is defined to be the first term on the right-hand side of (76a). This

expression involves only known quantities, at this stage of our construction, except for the coefficients $f_{\kappa\alpha}$, which we here choose arbitrarily except for the restriction (75). This just means that the phase of $f_{\kappa\alpha}$ is arbitrary in the usual case of no degeneracy. We have to show later that the potential we construct reproduces the same values of $f_{\kappa\alpha}$. From (76) we can construct $\bar{\mathbf{b}}(x)\mathbf{A}^{-1}(x)$ itself; assuming no subtractions are necessary, we have

$$\bar{\mathbf{b}}(x)\mathbf{A}^{-1}(x) = \bar{\bar{\xi}} + \frac{1}{\pi}\int \frac{\bar{\chi}(x')\,dx'}{x' - x - i\epsilon} - \sum \frac{\bar{w}_\kappa}{x_\kappa - x - i\epsilon}, \tag{77}$$

where $\bar{\bar{\xi}}$ is an arbitrary constant vector. Letting

$$\bar{\varphi}(x) \equiv \bar{\bar{\xi}} + \frac{1}{\pi}\mathcal{P}\int \frac{\bar{\chi}(x')\,dx'}{x' - x} + \sum_\kappa \frac{\bar{w}_\kappa}{x - x_\kappa}, \tag{78}$$

we rewrite this as

$$\bar{\mathbf{b}}(x)\mathbf{A}^{-1}(x) = \bar{\varphi}(x) + i\bar{\chi}(x) - i\pi\sum_\kappa \bar{w}_\kappa\delta(x - x_\kappa). \tag{79}$$

Multiplying on the right by $\mathbf{A}(x)$ $(= \mathbf{T} + i\mathbf{v}\bar{\mathbf{v}})$ we obtain an expression for $\bar{\mathbf{b}}(x)$ from which, if it is of the proper form (54), we can read off $v_0(x)$. We get

$$\begin{aligned}\bar{\mathbf{b}}(x) = \bar{\varphi}\mathbf{T} - [\bar{\chi} - \pi\textstyle\sum \bar{w}_\kappa\delta(x - x_\kappa)]\mathbf{v}\bar{\mathbf{v}} \\ + i\{\bar{\chi}\mathbf{T} - \pi\textstyle\sum \bar{w}_\kappa\mathbf{T}\delta(x - x_\kappa) + \bar{\varphi}\mathbf{v}\bar{\mathbf{v}}\}.\end{aligned} \tag{80}$$

Since, by Eqs. (46), (47), and (72), $\bar{w}_\kappa\mathbf{T}(x_\kappa) = 0$ and $\bar{w}_\kappa\mathbf{v}(x_\kappa) = 0$, and since

$$\bar{\chi}\mathbf{T} = [1/(\beta^2 + 1)]\zeta\bar{\mathbf{v}}, \tag{81}$$

by the definition of $\bar{\chi}$ [Eqs. (76a) and (76b)], we see that

$$\bar{\mathbf{b}}(x) = \bar{\varphi}\mathbf{T} - \bar{\chi}\mathbf{v}\mathbf{v} + i[1/(\beta^2 + 1)\zeta + \bar{\varphi}\mathbf{v}]\bar{\mathbf{v}}, \tag{82}$$

which is indeed of the same form as (54), so that

$$v_0(x) = \zeta/(\beta^2 + 1) + \bar{\varphi}\mathbf{v}. \tag{83}$$

Recall that $\beta(x)$ is known from the N-term case [Eq. (60)], that $\zeta(x)$ is determined from $D_0(x)$ [Eqs. (62), (64)], and that $\bar{\varphi}(x)$ is given by Eq. (78).

This completes the construction; it remains to show consistency. The form of Eq. (77) assures that $\bar{\mathbf{b}}\mathbf{A}^{-1}$ is the boundary value of a function analytic in the upper half-plane, and hence that $\bar{\mathbf{b}}(x)$ itself [Eq. (82)] is also. This implies that the vector function $\bar{\mathbf{t}}(x)$ constructed from our new-found potential $v_0(x)$ coincides with that obtained by comparing (82) with (54). Thus we know that this function $\bar{\mathbf{t}}(x)$ is given by

$$\bar{\mathbf{t}}(x) = \bar{\varphi}\mathbf{T} - \bar{\chi}\mathbf{v}\bar{\mathbf{v}}. \tag{84}$$

When this is substituted into Eq. (73), only the singular terms in $\bar{\varphi}$ contribute anything, on account of (46) and (47), and we end up with the desired result that the coefficients $f_{\kappa\alpha}$ so obtained coincide with those chosen initially, regardless of how they were chosen. One next shows, quite readily, that the expression $\zeta(x)$ obtained from Eq. (65) coincides with the $\zeta(x)$ we used in the construction, and that the coefficients d_κ obtained from the residues in $\bar{\mathbf{t}}\mathbf{T}^{-1}\mathbf{t}$ [Eqs. (63), (64)] coincide with those given by (75). Then, since $\zeta(x)$ and the d_κ's determine the imaginary part of $D_0(x)$, and hence $D_0(x)$ itself, it is clear that the partial phase shift calculated from our constructed $v_0(x)$ coincides with the one we started with.

C. General Construction Procedure

To fit a given phase shift with an N-term potential now, we must be able to express it as a sum of N partial phase shifts, to each of which in sequence we can apply the construction of the preceding section. At the rth stage, the sum of the first r partial phase shifts is then correctly fitted by an r-term potential. The restrictions on the rth partial phase were found above, and may be summarized best in terms of the modified partial phase shift $\eta_r(x)$, which may have discontinuities of $-\pi$ only where one of the preceding η's had a discontinuity of $+\pi$, and must satisfy $\eta_r(\infty) \leq \eta_r(-\infty)$. [Further restrictions on the behavior of the potential would impose more stringent conditions on $\eta_r(x)$, which would normally be continuous except for the discontinuities of $\pm\pi$, and would normally satisfy $\eta(\infty) = \eta(-\infty)$.] The selection of N partial phase shifts subject to these restrictions allows a great deal of arbitrariness, in addition to the arbitrariness at each stage of the construction procedure, in the choice of the phase of $\zeta(x)$ and the constants $f_{\kappa\alpha}$ and $\bar{\bar{\xi}}$. Making these choices in all possible ways we generate, with minor duplications, all possible potentials which can yield the given phase shift.

The question now arises as to how many terms are needed to fit a given phase shift. The basic answer, ignoring exceptions and making no requirements on the number of bound states, is fairly simple: If the maximum number of multiples of π through which the phase shift increases between any two energies x_1 and x_2 is $R - 1$, then a minimum of R terms is needed to fit it. To make a more precise description, we make the following definitions:

$$\nu(x) \equiv \lim_{x' \to x}\left[\text{largest integer} \leq \frac{\delta(x')}{\pi}\right]. \tag{85}$$

[The lim is introduced merely to deal with the case $\delta(x) = n\pi$; but with $\delta(x') < n\pi$ for x' near x, if δ

crosses $n\pi$ at x, we may take $\nu(x)$ as undefined there.]

$$R \equiv 1 + \max_{(x_2>x_1)} [\nu(x_2) - \nu(x_1)], \tag{86}$$

$$R_0 \equiv 1 + \max \left[\nu(x) - \frac{\delta(0)}{\pi}\right], \tag{87}$$

$$R_\infty \equiv \max \left[\frac{\delta(\infty)}{\pi} - \nu(x)\right], \tag{88}$$

$$N_B \equiv \text{number of bound states}, \tag{89}$$

$$N \equiv \text{number of terms in potential}. \tag{90}$$

Since

$$\nu(x) \leq \frac{\delta(x)}{\pi} \leq \nu(x) + 1, \tag{91}$$

we can conclude that

$$0 \leq R_0 \leq R \tag{92}$$

and

$$0 \leq R_\infty \leq R. \tag{93}$$

From a discussion in the previous section, we found that

$$N_B \leq N. \tag{94}$$

For finding restrictions on N, we can ignore the possibilities of discontinuities of $-\pi$ in the modified partial phase shifts η_r, since each such is coupled to a discontinuity of $+\pi$ in another $\eta_{r'}$, and both can be removed together, whether they occur at positive or negative x. That is, two η's with the same sum can be found which have no discontinuity there. So for this purpose we can restrict ourselves to partial phase shifts of the single-term type, which do not increase through $n\pi$, and for which $\eta_r(\infty) \leq \eta_r(-\infty)$. The only restriction that this imposes for $x > 0$ is that if $\eta_r(0+) = -\pi$, then [since $\eta_r(-\infty)$ must equal $-\pi$ and $\eta_r(\infty) \leq \eta_r(-\infty)$], $\eta_r(\infty) = -\pi$ also. If there is a bound state corresponding to this partial phase shift, that is, a discontinuity of $+\pi$ for $x < 0$, then $\eta(0+) \neq -\pi$ and $\eta(\infty) = -\pi$.

Since a single partial phase shift cannot increase through a multiple of π, we conclude that the sum of N terms cannot increase through N multiples of π or, more precisely,

$$R \leq N. \tag{95}$$

For each bound state there must be one modified partial phase shift η_r which approaches $-\pi$ as $x \to \infty$, which reduces by π the possible increase that the total phase shift can undergo as $x \to \infty$. Thus,

$$R_\infty \leq N - N_B. \tag{96}$$

[R_∞ is expressed in terms of $\delta(\infty)$ rather than $\nu(\infty)$ because if $\delta(\infty)$ is a multiple of π, it may still approach its limiting value from above or below.]

If $R_0 = N$, then we must have $\eta_r(0+) = -\pi$ for every term, in which case there can be no bound states, and $\eta_r(\infty)$ is also $-\infty$ for every term. This implies that R_∞ must be an integer, and cannot be positive:

$$R_0 = N \Rightarrow N_B = 0 \quad \text{and} \quad R_\infty = 0, -1, -2, \cdots. \tag{97}$$

Unless $R_0 = N$, $\eta_r(0+)$ need not be $-\pi$ for any r, and so no similar restrictions need be considered except in this case.

Finally, if there are N bound states, then again $\eta_r(\infty) = -\pi$, for all r, so that

$$N_B = N \Rightarrow R_\infty = 0, -1, -2, \cdots. \tag{98}$$

If there are any fewer bound states, then some of the $\eta_r(\infty)$ are unrestricted, and only the condition (96) need be satisfied.

These five equations (94)–(98) represent the only absolute restrictions on the phase shift for it to be fitted by an N-term potential; if they are satisfied, the construction of partial phase shifts satisfying the necessary restrictions can be carried out straightforwardly. We can therefore find the minimum value of N for a given phase shift: in general, if the number of bound states is not specified or is zero, then

$$N_{\min} = R, \tag{99}$$

unless $R_0 = R$ and R_∞ is positive or nonintegral, in which case

$$N_{\min} = R + 1. \tag{100}$$

If the number of bound states is given and is not zero, then

$$N_{\min} = \max [N_B, R, R_\infty + N_B, R_0 + 1], \tag{101}$$

unless R_∞ is nonintegral and negative, in which case

$$N_{\min} = \max [N_B + 1, R, R_0 + 1]. \tag{102}$$

The location of these bound states is arbitrary.

In all of this discussion the most general possible potential has been considered. It is clear that as additional restrictions are imposed on the potential, one obtains corresponding restrictions on the phase shift, as well as on the number and location of bound states. If one avoids singular potentials for which $\mu_i = 0$ (infinite strength) or for which $\mathbf{A}(z) \to \infty$ as $|z| \to \infty$, and if one avoids values of μ_i which correspond to bound states of zero energy, then the behavior is much simpler. In this case, for each i, $\eta_i(\infty) = \eta_i(-\infty)$, and $\eta_i(x)$ is continuous at $x = 0$,

so that $(1/\pi)[\delta(0) - \delta(\infty)]$ is any integer at least as great as the number of bound states, and in fact equal to the number of bound states plus the number of quasibound states of positive energy, discussed previously, which correspond to discontinuities $+\pi$ in the modified partial phase shifts $\eta_l(x)$. This is the generalization of Levinson's theorem[8,10] for the N-term separable potential.

If one restricts one's attention to potentials of finite range, vanishing for sufficiently large r, then further properties of the phase shift are implied, in particular that the S matrix $e^{2i\delta(x)}$ is in fact analytic in the cut z plane, having a simple second-order branch point at the origin, and taking values on the second sheet reciprocal to those on the first sheet. The zeros of the S matrix are thus determined, and hence the location and number of bound states. If one performs the inversion construction using the same phase shift but different bound states, one simply obtains a potential that does not vanish for large r.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Classification and Complete Reducibility of Representations of Compact Inhomogeneous Groups and Algebras

R. MIRMAN
Physics Department, Long Island University, The Brooklyn Center, Brooklyn, New York

(Received 27 November 1967)

The representations of compact inhomogeneous groups are classified according to which type of operators are taken as diagonal. The question of complete reducibility of some of the different classes is discussed.

1. INTRODUCTION

A basic problem in the theory of group representations is that of determining whether the representations are completely reducible. We consider this question here for compact inhomogeneous groups and their algebras, which we have discussed previously.[1] Certain facts about the classification of such groups which we use here have been proven in that paper. The notation, definitions, and terminology are the same.

The question of the complete reducibility of the representations of inhomogeneous groups has been considered by Mackey,[2] but in a general, abstract manner, and in a way which pertains especially to only one class of representations (that considered in Sec. 4). A much more explicit, detailed, and particular discussion is needed before the results can be fully useful for physical applications, especially for the other classes of representations. For applications it is often necessary to go beyond the general theorems to get specific facts about specific representations. We attempt to start such an investigation here.

The bases of the representations, and so the representations, are classified according to the type of operators which are diagonal. The three classes which we define in Sec. 2 are called discrete, continuous, and mixed. By reducing a representation we mean finding an equivalent (reduced) representation which is in the same class. This gives a new basis, but one which is in the same class of bases. Therefore, we do not consider as reducible a representation which is not equivalent to a reduced representation in the same class, but is equivalent to a reduced representation of a different class (if that is possible).

Because of the requirement imposed by the classification we are very restricted in finding equivalent representations. To a large extent, the basis for the state vectors is unique once the class for the basis has been chosen, and we cannot take linear combinations of state vectors if we are to stay in the given class. That is, once the class (and, thus, the operators which are diagonal) is chosen and a representation is found, there is no other representation equivalent to it for many representations.

For the discrete representations, which are represented by matrices with a countable set of rows and

[1] R. Mirman, J. Math. Phys. **9**, 39 (1968).

[2] G. W. Mackey, Am. J. Math. **73**, 576 (1951); Ann. Math. **55**, 101 (1952); **58**, 193 (1953); in *Proceedings of the International Symposium on Linear Spaces at the Hebrew University of Jerusalem, July 5–12, 1960* (Pergamon Press Ltd., London 1961), pp. 319–326; "The Theory of Group Representations," University of Chicago Lecture Notes, 1955 (unpublished); I. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, R.I., 1963).

so that $(1/\pi)[\delta(0) - \delta(\infty)]$ is any integer at least as great as the number of bound states, and in fact equal to the number of bound states plus the number of quasibound states of positive energy, discussed previously, which correspond to discontinuities $+\pi$ in the modified partial phase shifts $\eta_l(x)$. This is the generalization of Levinson's theorem[8,10] for the N-term separable potential.

If one restricts one's attention to potentials of finite range, vanishing for sufficiently large r, then further properties of the phase shift are implied, in particular that the S matrix $e^{2i\delta(x)}$ is in fact analytic in the cut z plane, having a simple second-order branch point at the origin, and taking values on the second sheet reciprocal to those on the first sheet. The zeros of the S matrix are thus determined, and hence the location and number of bound states. If one performs the inversion construction using the same phase shift but different bound states, one simply obtains a potential that does not vanish for large r.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Classification and Complete Reducibility of Representations of Compact Inhomogeneous Groups and Algebras

R. MIRMAN
Physics Department, Long Island University, The Brooklyn Center, Brooklyn, New York

(Received 27 November 1967)

The representations of compact inhomogeneous groups are classified according to which type of operators are taken as diagonal. The question of complete reducibility of some of the different classes is discussed.

1. INTRODUCTION

A basic problem in the theory of group representations is that of determining whether the representations are completely reducible. We consider this question here for compact inhomogeneous groups and their algebras, which we have discussed previously.[1] Certain facts about the classification of such groups which we use here have been proven in that paper. The notation, definitions, and terminology are the same.

The question of the complete reducibility of the representations of inhomogeneous groups has been considered by Mackey,[2] but in a general, abstract manner, and in a way which pertains especially to only one class of representations (that considered in Sec. 4). A much more explicit, detailed, and particular discussion is needed before the results can be fully useful for physical applications, especially for the other classes of representations. For applications it is often necessary to go beyond the general theorems to get specific facts about specific representations. We attempt to start such an investigation here.

The bases of the representations, and so the representations, are classified according to the type of operators which are diagonal. The three classes which we define in Sec. 2 are called discrete, continuous, and mixed. By reducing a representation we mean finding an equivalent (reduced) representation which is in the same class. This gives a new basis, but one which is in the same class of bases. Therefore, we do not consider as reducible a representation which is not equivalent to a reduced representation in the same class, but is equivalent to a reduced representation of a different class (if that is possible).

Because of the requirement imposed by the classification we are very restricted in finding equivalent representations. To a large extent, the basis for the state vectors is unique once the class for the basis has been chosen, and we cannot take linear combinations of state vectors if we are to stay in the given class. That is, once the class (and, thus, the operators which are diagonal) is chosen and a representation is found, there is no other representation equivalent to it for many representations.

For the discrete representations, which are represented by matrices with a countable set of rows and

[1] R. Mirman, J. Math. Phys. **9**, 39 (1968).

[2] G. W. Mackey, Am. J. Math. **73**, 576 (1951); Ann. Math. **55**, 101 (1952); **58**, 193 (1953); in *Proceedings of the International Symposium on Linear Spaces at the Hebrew University of Jerusalem, July 5–12, 1960* (Pergamon Press Ltd., London 1961), pp. 319–326; "The Theory of Group Representations," University of Chicago Lecture Notes, 1955 (unpublished); I. Segal, *Mathematical Problems of Relativistic Physics* (American Mathematical Society, Providence, R.I., 1963).

columns, the complete reducibility of the representations of either the group or the algebra implies it for the other.[3] This is true even when the matrices are of infinite dimension, for, by a representation of the group corresponding to the representation of the algebra, we mean the set of operators obtained by exponentiating the operator of the algebra; if the matrices representing the p's leave a subspace invariant, continued application of them also leaves the subspace invariant. Here we do not assume that for every representation of the algebra there is a corresponding representation of the group, but, since we use the algebra to study the complete reducibility, we do need the converse. For the other classifications the situation is more complicated and is discussed below for the continuous case.

A representation, which is the product of representations belonging to different spaces, need not fall into our classification because the different terms in the product may be eigenfunctions of different operators. However, since each term in the product can be considered separately this does not affect the discussion.

Except where stated, we limit ourselves to homogeneous parts which are simple and the p's which form an irreducible representation of that simple algebra. The generalization can be obtained from our results.

The different classes of representations are considered separately: in Sec. 3, the discrete representations; in Sec. 4, the continuous; and in Sec. 5, the mixed. The Appendix discusses a property of some types of invariants of inhomogeneous groups.

2. PROOF OF CLASSIFICATION

The representations are divided into three classes. In the first class all of the p's are diagonal, none of the E's, and none of the homogeneous Casimir operators (polynomials in the E's only, commuting with all E's). In the second class, all of the E's in the Cartan subalgebra and all of the homogeneous Casimir operators are diagonal, and none of the p's. In the third class, the diagonal operators are some p's, and some E's and their Casimir operators. These are referred to as continuous, discrete, and mixed representations, respectively.

For the first two classes we must prove that, when we have diagonalized the operators in any one of the above sets, then no others are diagonal. We exclude from our considerations as trivial, p's which commute with all the E's.

[3] H. Boerner, *Representations of Groups* (North-Holland Publishing Company, Amsterdam, 1963), p. 90.

From this and the Jacobi identity, it immediately follows that no E commutes with all the members of each irreducible set of p's. Let E_1 commute with all p's, and let E_3 not commute with at least one p, say p_1. We can always choose such an E_1 and E_3, and an E_2 such that $[E_1, E_2] = E_3$. Then in the Jacobi identity

$$[E_1, [E_2, p_1]] + [E_2, [p_1, E_1]] + [p_1, [E_1, E_2]] = 0, \tag{2.1}$$

the first two terms are zero, and the last gives the commutator of p_1 and E_3 which was assumed not zero, giving a contradiction, with the result that no such E_1 exists.

We prove next that no p can commute with a homogeneous Casimir operator. So, from this fact and the statement that no E commutes with all p's, we see that, if all p's are diagonal, no E's are; and that if even one homogeneous Casimir operator is diagonal, no p's are. This means that we can choose the labeling operators as stated above.

First we assume that at least one p from each irreducible set of p's can always be found that does not commute with some homogeneous Casimir operator (that is, that such an operator is not an invariant), prove that no p from that set commutes with it, and then show that such a p can always be found.

The Casimir operator is denoted by C. We can always find an E_i, a p_α, and a p_β such that $[E_i, p_\alpha] = p_\beta$, where the sum on the right-hand side is limited to one term by defining p_β as equal to whatever sum appears there. We take p_α as commuting with C and p_β as not commuting with it.

Then, the Jacobi identity

$$[C, [E_i, p_\alpha]] + [E_i, [p_\alpha, C]] + [p_\alpha, [C, E_i]] = 0 \tag{2.2}$$

gives

$$[C, p_\beta] = 0, \tag{2.3}$$

which is a contradiction to the above assumption, with the result that, if p_β does not commute with C then there is no p_α which commutes with it.

We now turn to the proof that no polynomial in the E's is an invariant.

First consider a scalar consisting of only one term, so that the relevant commutation relation is

$$[p_\alpha, E_1 E_2 \cdots E_n] = -\sum_{a,\gamma} p_\gamma D^\gamma_{a\alpha} E_1 E_2 \cdots E_{a-1} E_{a+1} \cdots. \tag{2.4}$$

The p can be written in front because lower-order terms in the E's do not alter the following argument. We now take the coefficient of p_β in this sum and show that it is not zero, which proves that the sum,

and thus the commutator, is not zero. If it were zero, then its commutator with anything would be zero. Taking its commutator with p_α and taking the coefficient of p_β, we get a similar term with two D's. Note that not all the D's are zero, for we assume that no E commutes with all the p's. We continue this process and get finally

$$\sum_{a,b,\cdots,m} D^\beta_{a\alpha} D^\beta_{b\alpha} \cdots D^\beta_{m\alpha} = 0 \tag{2.5}$$

as the requirement.

But this cannot be, as all terms in the sum are identical except for the irrelevant ordering of the factors. So this sum, and therefore the original commutator, is nonzero, proving that a single term is not an invariant.

For a sum of terms we have

$$\left[p_\alpha, \sum_i R_i E_{i1} E_{i2} \cdots\right] = -\sum_{i,i_f,\gamma} R_i D_{i_f\alpha} E_{i_1} \cdots E_{i_{f-1}} E_{i_{f+1}} \cdots p_\gamma. \tag{2.6}$$

In order that this sum to zero, there must be a set of terms in the sum, all coefficients of p_β, with the same E's whose coefficients sum to zero. Now in the sum inside the commutator we must be able to take any two terms and find at least one E which differs between them. Different orderings merely introduce lower-order terms which do not affect the argument. If terms on the right-hand side sum to zero, then they must originate from terms inside the commutator which differ in only one E. Thus inside the commutator there is the term

$$R_1 E_a \cdots E_1 + R_2 E_a \cdots E_2 + \cdots, \tag{2.7}$$

and this term when commuted with any p gives a series of terms, including ones with the $E_1, E_2, \cdots$ removed and replaced by p's. This latter series sums to zero for each p by hypothesis. We now define an E_z, which is a linear combination of $E_1, E_2, \cdots$, such that

$$E_z = R_1 E_1 + R_2 E_2 + \cdots. \tag{2.8}$$

So the sum inside the commutator reduces to one term and the sum on the right-hand side also reduces to one term. But this term cannot be zero for all p's, because there is no E_z which commutes with all p's. Therefore, the commutator is not zero, and no polynomial in the E's only is an invariant. The proof does not hold if the polynomial is linear in the E's; but then it could be defined as an E, and we have shown above that no E commutes with all the p's.

There is one trivial exception to the above result, which is not, however, in contradiction to our main result about the choice of the diagonal operators. This is an E which commutes with all operators and so must be an invariant. For $IU(n)$, when the p's are written with an equal number of upper and lower indices, the generator E^i_i (sum over all i's) commutes with all other generators.

Even though the operators do not commute in general their commutators may be zero in some representations. For example, in representations in which the p's are zero, every commutator with them will be zero. Thus it is possible that in these representations (provided they exist, beyond the trivial case in which a set of operators is zero) we will need more labeling operators than are implied by our general considerations. That is, there may be vectors which have the same eigenvalues of the p's but are not identical, having different eigenvalues of some other operator.

As can be seen from the above considerations, the only way such a difficulty could occur is if all the p's are diagonal and there is a homogeneous Casimir operator which "accidentally" commutes with all of them.

We do not consider such representations here but regard them as a special case of mixed representations.

Besides the operators listed, there may be polynomials in the p's and E's which commute with all the p's but not all the E's. An example is the vector w_i of the Poincaré group, whose square forms the second invariant of the group. When all the p's are diagonal, can these furnish labeling operators? Can there be two vectors which have the same eigenvalues of the p's, but different eigenvalues of the w's?

For representations in which the p's are diagonal we use the realization given in the Appendix, and in this case the w's are identically zero; they cannot be constants because they do not commute with the E's. By the argument of the Appendix, we find that if the w's and the p's commute, then the p's (that is, the derivatives with respect to x) acting on the w's are zero. Hence the w's are zero.

Note that, for a product of a representation with the p's diagonal times a representation with the E's diagonal, the w's are not necessarily zero. An example is a representation of the Poincaré group of the form

$$z \exp(ip \cdot x), \tag{2.9}$$

where z is a spinor.

We come now to the possibility of polynomials in the p's and E's, which are not invariants, but do commute with the homogeneous Casimir operators. If they do not commute with all the E's, then they would provide additional labeling operators for the basis vectors of semisimple representations. Since this

is impossible, we need not be concerned with this possibility.

If they commute with all the E's, then they would label different, equivalent semisimple representations in the series of such representations which make up the single representation of the inhomogeneous group. We give no example of this, but see no reason why it should not occur.

We also note that if p and E commute with any operator C, then any p obtained by commutation of p with E any number of times, which we shall denote by p', also commutes with C. From the Jacobi identity

$$[p, [C, E]] + [C, [E, p]] + [E, [p, C]] = 0, \quad (2.10)$$

we get $[C, p'] = 0$, and repeating the procedure with p', p'', etc., we get the result stated.

It should also be noted that unlike the situation for semisimple algebras, the number of diagonal generators (the "rank") is not constant but varies with the representation. For example, consider the adjoint $ISU(2)$, which is the inhomogeneous rotation algebra. We can take as diagonal, p_x, p_y, and p_z; or p_z and J_z; or J_z and J.[2]

3. THE DISCRETE CLASS

All inhomogeneous algebras have noncompletely reducible discrete representations. In addition some, but not all, have infinite-dimensional, completely reducible discrete representations. We prove the existence of noncompletely reducible discrete representations and give examples of algebras having completely reducible discrete representations, as well as an algebra which has no completely reducible discrete representation. We do not, however, give criteria to determine whether or not an algebra has completely reducible discrete representations; nor do we discuss whether or not it is possible for discrete representations with finite-dimensional p's to be completely reducible.

For infinite-dimensional p's we exhibit a representation of each type, taken explicitly from the adjoint $ISU(2)$ group. Completely reduced discrete representations have been worked out by Pauli,[4] and we simply refer to that paper for the details. Another example [from adjoint $ISU(3)$] is given by Bose.[5]

We also use Pauli's work as the basis of the derivation of the noncompletely reducible representation. We take $(j|\,p\,|j) = 0$ and $(j+1|\,p\,|j) = 0$, and show that we still have a representation. This representation is, of course, noncompletely reducible.

It is, of course, also non-Hermitian. A representation in which the p's are Hermitian, so that $(j+1|\,p\,|j) = (j-1|\,p\,|j)^*$, is clearly completely reduced.

To construct the representation all that is necessary is to show that the commutator of two p's taken between the states $|jm)$ and $|j-2, m')$ vanish, for the commutators between any other states vanish automatically by the assumption that p only has a step-down matrix element. And the vanishing of the commutator between these states comes about because of the Clebsch–Gordan coefficients, not because of the reduced matrix elements of the p's. Thus all the required commutation relations are satisfied. Note that if we took p with step-up elements (on the superdiagonal) instead of step-down elements (on the subdiagonal), the result would be the same.

For all inhomogeneous algebras there are always noncompletely reducible discrete representations. To show this we first consider what restrictions must be satisfied by a representation.

By the conditions of this section we are considering states which are grouped into representations of the semisimple part. Now the p's form a representation of the semisimple part and are considered to be written, by the Wigner–Eckart theorem, as the product of a Clebsch–Gordan coefficient of the semisimple part times a reduced matrix element. The latter depends on the semisimple representations between which the p is taken, but not on the state of the representation.

For an irreducible representation every state can be reached from every other by the application of some set of operators. Clearly this is always true of the states within any one semisimple representation. In studying the reducibility of the representation, it is only necessary to consider the reduced matrix elements of the p's, and how they connect different semisimple representations. The commutation relations between two E's are satisfied automatically because of the use of the representations of the E's, those of the p's and E's by the explicit form of the matrix elements of the p's (through the use of the Wigner–Eckart theorem). Thus there is left only the requirement that the p's commute.

This last is satisfied by requiring one commutator for irreducible p's to be zero, since we now show that all other commutators are zero if this one is. Let the p with the highest weight be denoted by p_h (if the homogeneous part is semisimple, but not simple, p_h has the highest weight in the representation of each simple part) and the one with the lowest by p_l $(= p_{-h})$.

Then we require that

$$[p_h, p_l] = 0. \quad (3.1)$$

[4] W. Pauli, CERN Preprint 56-31, 1956 (unpublished), p. 12; B. Kursunoglu, *Modern Quantum Theory* (W. H. Freeman and Co., San Francisco, 1962), p. 120.

[5] S. K. Bose, Phys. Rev. **150**, 1231 (1966).

Taking the commutator of this commutator and $E_{+n\alpha}$, where n is any positive integer and α any positive root, we find that p_h commutes with $p_{l+n\alpha}$. Commuting $[p_h, p_{l+n\alpha}]$ with $E_{+m\beta}$, m and β positive, etc., we find p_h commutes with

$$p_{l+n\alpha+m\beta+\cdots}$$

However, any p can be written in this form for some $n, m \cdots$ and some α, $\beta \cdots$, if the representation formed by the p's is a simple completely reduced one. If it is not, then we must require Eq. (3.1) for each such representation; and in addition, for, say, two representations p and q, we require that

$$[p_h, q_l] = 0, \tag{3.2}$$

and then from this proof all p's and q's commute. Therefore, p_h commutes with all other p's. Consider the commutator

$$[p_h, p_a] = 0, \tag{3.3}$$

where p_a is arbitrary. Commuting it with E_{-np}, any arbitrary generator, we get

$$[p_{h-np}, p_a] = 0, \tag{3.4}$$

and commuting this last commutator with $E_{-m\sigma}$, etc., we find that p_a commutes with $p_{h-np-m\sigma\cdots}$. Since any p can be written in this form, we find that any p_a commutes with any other, and so all p's commute which proves the result.

Hence we must now require the reduced matrix elements of p to satisfy Eq. (3.1) when it is taken between any two states whatever.

The states are labeled j and m, where the set of numbers j labels the representation of the E's, while the set m labels the state in the representation. Suppressing the m's, we let j' be a representation such that $(j' | p | j) \neq 0$, and so that j' is one of the representations in the reduction of the direct product of j and p, and j'' be such that $(j'' | p | j) = 0$; but $(j'' | p | j') \neq 0$, for some j', so that $(j'' | pp | j) \neq 0$. If we now take the commutator between $| j)$ and some other state, then the only states that have to be considered belong to the set $\{| j), | j'), | j'')\}$, for all j, j', and j'', with j' and j'' depending on j. For all other states the commutator is automatically zero.

Considering first the set $\{| j'')\}$ we get

$$\sum_{j'} (j'' | p | j')(j' | p | j) \{C_1C_2 - C_3C_4\} = 0, \tag{3.5}$$

where the C's denote Clebsch–Gordan coefficients with the arguments suppressed, and where the sum is over all possible intermediate j'.

Let us denote the highest-weight vector of the representation $| j)$ by $\mathbf{w}_j$ and that of p by $\mathbf{w}_p$. Then, in the reduction of the product $p | j)$, the representation of highest weight $\mathbf{w}_j + \mathbf{w}_p$ occurs once. Now in Eq. (3.5), take j'' to be the representation with highest weight $\mathbf{w}_j + 2\mathbf{w}_p$. (Of course, the state on the left-hand side does not have this weight.) Then the sum has one term. Denoting the reduced matrix element of p between $| j)$ and the representation of highest weight $\mathbf{w}_j + \mathbf{w}_p$ by A, with the appropriate subscripts, which are usually suppressed, we get, with j' and j'' going with the representations with the highest weights just given, that Eq. (3.5) becomes

$$A_{jj'}A_{j'j''}(C_1C_2 - C_3C_4) = 0. \tag{3.6}$$

We show below that A is never zero. Hence this is an identity in the Clebsch–Gordan coefficients and does not place any restrictions on the reduced matrix elements.

That the term involving the Clebsch–Gordan coefficients must give zero and not the A's can be seen by considering the reduction of the direct product of any representation with the direct product of any representation with itself. The latter need not have anything to do with inhomogeneous groups. We take from this reduction the unique representation with $\mathbf{w}_j + 2\mathbf{w}_p$ as highest weight. Since the abstractly considered direct product is commutative, the order in which the terms are arranged is irrelevant. But since the representation with this highest weight occurs in the direct-product reduction, the commutivity must come about because of the Clebsch–Gordan coefficients.

Next we consider the commutation relation between $| j)$ and $| j')$, for any $| j')$. Suppressing symbols referring to states within a semisimple representation, we obtain

$$\sum_q [(j' | p_h | j')_q (j' | p_l | j) + (j' | p_h | j)(j | p_l | j)_q] + \sum_{\bar{j}} (j' | p_h | \bar{j})(\bar{j} | p_l | j) - (h \leftrightarrow l) = 0. \tag{3.7}$$

Finally, between $| j)$ and $| j)$ the commutation relation becomes

$$\sum_{qr} (j | p | j)_q (j | p | j)_r + \sum_{\bar{j}} |(j | p | \bar{j})|^2 = 0, \tag{3.8}$$

with the same suppression of symbols as in the previous equation. The q and r refer to different matrix elements between the same states because the direct product may not be multiplicity free.

We now show that we can choose a noncompletely reducible representation which satisfies the above equations. To do this we assume that the only nonzero reduced matrix element of p is A. Then Eq. (3.8) is clearly satisfied as all terms are zero. (The Hermitian conjugate of A is zero.) Equation (3.7) is also satisfied

because the first two terms are zero, as are the diagonal reduced matrix elements, and since the only values which j' and j can have correspond to the representation with highest weight $\mathbf{w}_j + \mathbf{w}_p$, the last term is zero also. For Eq. (3.5), there is only one value of j'' for which the result is not identically zero, and that gives Eq. (3.6). But this term is zero because of the Clebsch–Gordan coefficients, not because of the reduced matrix elements. So all the commutation relations are satisfied.

If the p's form a reducible representation,Eq. (3.1) must hold for each irreducible part and, with the above choice of matrix elements, it clearly does. In addition, Eq. (3.2) must hold between the various irreducible parts. We now show that we can choose the matrix elements to satisfy the latter requirement.

First, if two irreducible parts are isomorphic we choose the reduced matrix elements to be the same, and so we can now assume that all the irreducible parts of the p's have different highest weights. Consider two irreducible parts p and q and denote the nonzero reduced matrix elements that were discussed above by A and B respectively. Take the commutator of p and q between states 1 and 4. This gives

$$A_{13}B_{34}C_1C_2 - B_{12}A_{24}C_3C_4 = 0, \tag{3.9}$$

with the C's representing Clebsch–Gordan coefficients. Since no conditions have yet been placed on the A's and B's, we can clearly choose them to satisfy this equation.

What we must now show is that they can be chosen consistently to satisfy the set of equations obtained by taking the commutator between all possible states (notice that states 2, 3, and 4 are uniquely determined by state 1). If there are only two irreducible parts, this is immediate, because A_{13}, B_{34}, etc., each appear in only two equations. Different values of state 1 give different subscripts on the A's and B's. Now even assuming that the values of the matrix elements with the smaller subscripts are known, we can still take the other two so that the equation is satisfied. Further, since all the matrix elements are independent we can impose any boundary conditions that we wish, e.g., that certain sets of matrix elements be zero.

If there are more than two irreducible parts (and by the stipulation made above no two have the same highest weight), then the same argument goes through with only a slight modification, because each matrix element appears in more than two equations. We label the matrix elements $A, B, D, E, \cdots$, and let representation 1 be the one with least highest weight in the representation of the inhomogeneous algebra. Then we have equations involving $A_{12}B_{23}$, $A_{12}D_{24}$, etc. We fix A and from these equations calculate B_{15}, D_{16}, etc. If we now consider state 2 we have $B_{23}S_{37}$, etc., which we can use to determine B_{23}. Continuing in this manner we see that we can determine consistently all the matrix elements in order to satisfy all of Eq. (3.2).

We note that the above argument holds even if the homogeneous part is semisimple but not simple, for it is necessary only that representations 1, 2, etc., be unique and be determined by their highest weight. This is true even if they are direct products of representations of different simple algebras.

Because the reduced matrix elements of the p's are all upper triangular, these representations are not completely reduced. The representations can be finite or infinite dimensional because p applied to any semisimple representation will give a representation with a larger highest weight, so that there is no upper limit unless all but a finite set of the A's, B's, etc., are set equal to zero.

An example of an algebra which does not have any completely reducible representations is the fundamental $ISU(2)$, whose two inhomogeneous generators are denoted by p_+ and p_-, with the requirement that these commute. This gives the equation

$$\begin{aligned}|(j+\tfrac{1}{2}|\,p\,|j)|^2\left\{\left[\frac{(j+m+1)^2}{(2j+1)(2j+2)}\right]^{\frac{1}{2}}\right.&\\ \left.+\left[\frac{(j-m+1)^2}{(2j+1)(2j+2)}\right]^{\frac{1}{2}}\right\}&\\ -|(j-\tfrac{1}{2}|\,p\,|j)|^2\left\{\left[\frac{(j-m)^2}{(2j+1)(2j)}\right]^{\frac{1}{2}}\right.&\\ \left.+\left[\frac{(j+m)^2}{(2j+1)(2j)}\right]^{\frac{1}{2}}\right\}=0,&\end{aligned} \tag{3.10}$$

which leads to the condition

$$(j+1)^{\frac{1}{2}}f(j+\tfrac{1}{2}) - (j)^{\frac{1}{2}}f(j) = 0, \tag{3.11}$$

where

$$f(j) = |(j-\tfrac{1}{2}|\,p\,|j)|^2; \tag{3.12}$$

and the solution to this difference equation is

$$f(j) = K[j(2j+1)]^{-\frac{1}{2}}. \tag{3.13}$$

But clearly,

$$f(0) = |(-\tfrac{1}{2}|\,p\,|0)^2 = 0, \tag{3.14}$$

giving $K = 0$. So all matrix elements of the p's are zero and there are no completely reducible representations.

4. THE CONTINUOUS CASE

For those representations for which all the p's are diagonal we present only a very heuristic discussion. There might be representations for which the ordinary

naive approach may not hold, e.g., those which are representations of the algebra but not of the group.[6] These are explicitly excluded from our considerations, as are nonunitary representations (if there are any).

With these limitations we can say that all representations of the group of this class are completely reducible, and that this fact is a simple consequence of the requirement that every group element has an inverse. For this result we need not require that the non-Abelian part of the group be semisimple.

The eigenfunction (and because this discussion is only heuristic we use this term and are not concerned that we are really talking about functionals) of p_m is

$$\exp(ik_m x_m), \tag{4.1}$$

and that of all the p's together is obtained by summing over the index. The p's have been considered as differential operators in the space of the x's. A particular eigenvector is determined by the set of values k_i.

A representation is irreducible if the space over which it acts has no invariant subspaces. To see whether this is true here we must first generate the space over which the representation acts. To do this we take any eigenvector and apply all possible transformations generated by the homogeneous part to it. Each vector thus obtained is written as the sum of the original eigenvector and vectors orthogonal to it. To go from one basis vector to another we transform and then subtract out the first vector to get an orthogonal one which is the new basis vector. This process is then repeated on all the vectors obtained in this manner and continued until no further vectors orthogonal to all the previous ones are obtained. However, in this process of generating the basis of the space there is a significant difference between the presently considered case and that for representations with discrete basis vectors. For here each vector obtained by a transformation is orthogonal to the original vector (unless it is simply proportional to it). Any two vectors of the form given above are orthogonal if their sets of k differ. The second step, of subtracting out the first vector to obtain an orthogonal one, is not necessary here. Unlike the discrete case, any transformation takes a basis vector to a basis vector.

But from this, simple reducibility follows immediately. Since there was a transformation T that took basis vector V_0 to basis vector V_1, the transformation T^{-1} takes V_1 to V_0; by choosing the right transformation, we can take V_1 to any vector of the space. Hence there is no subspace that is invariant and the representation space is either completely reduced or irreducible.

The results for the algebra are somewhat different from those for the group; we restrict ourselves here to semisimple homogeneous parts. The operator of the algebra is a sum of terms of the form $x_i d/dx_j$, except for multiplicative constants and, when applied to an eigenvector such as that given by Eq. (4.1), leads to terms of the form $x_i p_j \exp(ip_m x_m)$. Expanding this in a set of basis vectors gives

$$ip_j \int \delta'(q_i - p_i)\delta(q_1 - p_1) \cdots \exp(iq_m x_m)\, dq, \tag{4.2}$$

so the coefficient is the derivative of a delta function.

In order to interpret this result consider the functional—all the eigenvectors are of course really only symbols for functionals—

$$\begin{aligned} \hat{f}(p + \Delta p) &= \hat{f}(p) + \Delta p \hat{f}'(p) + \cdots \\ &= \int f(x) e^{i(p+\Delta p)x}\, dx \\ &= \int f(x) e^{ipx}\, dx + \Delta p \int f(x) ix e^{ipx}\, dx + \cdots \\ &\to e^{ipx} - i\delta'(p) e^{ipx} \Delta p + \cdots, \end{aligned} \tag{4.3}$$

where Δp is "small." Thus heuristically we can see that this is the "neighboring" functional and that the derivative of the delta function appears in the expansion of it around the original functional.

We can therefore roughly say that the entire space of the eigenvectors of a single representation is covered if we apply the operators of the algebra an infinite number of times. We have shown this, remembering the relation between the group and the algebra, when we considered a finite transformation, i.e., a transformation of the group. However, a single application only gives the neighboring eigenvector and the operators of the algebra cannot take us from one eigenvector to a "finitely different" one. This is the basis of O'Raifeartaigh's theorem.[7,8]

In the sense that any eigenvector can be reached from any other by the application of the algebra operator an infinite number of times, the representation of the algebra is completely reducible. The representation space of the algebra is defined to be that of the group.

5. THE MIXED CASE

For the mixed representations we make only a brief comment. Let the diagonal p's be denoted by q, the nondiagonal by r, the diagonal (in the sense to be

[6] M. Flato and D. Sternheimer, Phys. Rev. Letters **16**, 1185 (1966).

[7] L. O'Raifeartaigh, Phys. Rev. Letters **14**, 575 (1965).

[8] R. Mirman, Phys. Rev. Letters **16**, 58 (1966).

discussed below) E's by F, and the nondiagonal by G. The various subscripts have been suppressed.

By diagonal E's we mean those E's which commute with all the q's. Since they do not commute with each other, clearly they are not all diagonal. The F's form an algebra, for $F_3 = [F_1, F_2]$ commutes with all the q's if the F's in the commutators do. Unfortunately, this algebra need not be semisimple. For example, consider the adjoint $ISU(3)$, with generators E_j^i, H_k, p_n^m, and p_s (corresponding to H_s). If we take p_2 (corresponding to the hypercharge) as the only diagonal p, then the isospin subalgebra and H_2 form the commuting algebra which is semisimple. If however, we take as diagonal p_2^1, the commuting algebra is E_2^1, H_2, E_3^1, E_2^3, which is solvable. The fact that the theory of representations of solvable algebras is much less developed than that of semisimple algebras presents a difficulty in the analysis of this class of representations.

An example of mixed representations has been given elsewhere.[9]

APPENDIX

There are two types of invariants: those containing only p's, and those containing p's and E's. It is useful to note that all invariants of the second type give zero when applied to any function of the p's.

First we prove that the realization

$$E_a = -D_{a\beta}^{\gamma} x_\beta \partial/\partial x_\gamma, \quad p_\beta = \partial/\partial x_\beta \tag{A1}$$

satisfies the commutation relations defining the Lie algebra

$$[E_a, E_b] = C_{ab}^c E_c, \tag{A2}$$

$$[E_a, p_\beta] = D_{a\beta}^{\gamma} p_\gamma, \tag{A3}$$

where the structure constants satisfy the Jacobi identity

$$D_{a\nu}^{\rho} D_{b\mu}^{\nu} + D_{b\nu}^{\rho} D_{\mu a}^{\nu} + C_{ab}^c D_{\mu c}^{\rho} = 0. \tag{A4}$$

It can be seen immediately that with the realization given in Eq. (A1), Eq. (A3) holds and, also, because of the Jacobi identity given in Eq. (A4), that Eq. (A2) holds.

The realization

$$p_\beta, \quad E_a = D_{a\gamma}^{\beta} p_\beta \partial/\partial p_\gamma \tag{A5}$$

also satisfies the commutation relations by the same argument.

Consider the realization in Eq. (A5). We remember that the commutator of two operators is defined by its action on every function including 1. The commutator of an invariant with any product of powers of any p's equals the invariant acting on that product. That is, the invariant is a sum of terms which are products of powers of p's times products of powers of differential operators in the p's, and this acts on the product with which the invariant is commuted. But this commutator is zero. Therefore, the invariant acting on any function which can be expanded in a Taylor series in the p's gives zero.

This result applies only to functions of variables which transform like the p's. It does not apply to all possible representations of the semisimple part. For example, if the p's transform according to the adjoint representation, the result does not apply to functions of the fundamental representations. In general the p's and the basis vectors of the representations of the semisimple part are both polynomials in the basis vectors of the fundamental representations. In analyzing the representations of inhomogeneous groups it is sometimes important to remember this point, especially in the study of mixed representations.

[9] R. Mirman, J. Math. Phys. 8, 57 (1967).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Crossing and Unitarity in a Multichannel Static Model. II

P. O. G. EHRHARDT AND D. B. FAIRLIE
Mathematics Department, University of Durham, Durham City, Great Britain

(Received 28 November 1967)

Constraints on the solution of the unitarity equation for a two-channel scattering problem arising from the requirements of crossing for the inelastic amplitude are shown to imply the relations imposed by crossing under $SU(2)$ for the elastic amplitudes in the limit of vanishing coupling to the inelastic channel. This result is extended to the three-channel case, where at least two distinct classes of internal symmetry crossing relations can exist: the unitarity equations alone simulate the $SU(2)$ case, but not that associated with higher symmetry groups.

In this paper we explore a remarkable feature of the multichannel unitarity equations for scattering processes in the static limit.[1] We discovered the existence of a very close connection between the demands of unitarity and the requirements of crossing under $SU(2)$ in the course of an attempt to find a power-series expansion for the solution of a scattering problem incorporating both two-particle unitarity, and crossing symmetry under an internal symmetry group, which was treated by other methods in the previous paper. We find that the requirements of the off-diagonal unitarity equation for a coupled two-channel process simulate those of $SU(2)$ crossing for the diagonal elements in the limit as the off-diagonal elements become zero. This result is based upon the following theorem: We impose the same analytic requirements on the scattering amplitudes as in Ref. 1 but do *not* impose crossing (c).

If $S_{11}(\omega)$, $S_{12}(\omega)$, $S_{22}(\omega)$ are S-matrix elements for a coupled two-channel scattering process in the Chew–Low approximation and the inelastic element $S_{12}(\omega)$ is supposed an even or odd function of ω, the s-channel energy in the static limit, then $S_{12}(\omega)$, $S_{22}(\omega)$ are completely determined by $S_{11}(\omega)$ apart from inessential phase factors. When this connection between $S_{11}(\omega)$ and $S_{22}(\omega)$ is studied in the limit as $S_{12}(\omega) \to 0$ so that the coupled-channel process becomes diagonal, indicative of the presence of a superselection rule, the relationship between these matrix elements is just that imposed by $SU(2)$ crossing. This analysis is extended to three coupled channels where at least two distinct types of crossing relation can exist for elastic processes: one arising from processes invariant under $SU(2)$ and another from processes in $SU(n)$ $(n > 2)$. (In the two-channel case, only $SU(2)$ crossing can arise for elastic processes.) We find that by a similar procedure, as outlined above for the two-channel case, we recover the $SU(2)$ solution, but not the other, and interpret this result as the emergence of angular momentum constraints.

Consider first the case of two-channel coupled unitarity equations, in static limit, written in terms of the variable z, defined by the transformation

$$z = (2\pi i)^{-1} \log [\omega + (1 - \omega^2)^{\frac{1}{2}}], \tag{1}$$

following Mescheryakov.[2] This is essentially the same transformation which was used by Rothleitner[3] and Cunningham.[4] Unitarity takes the form

$$\sum_{j=1}^{2} S_{ij}(z) S_{jk}(1 - z) = \delta_{ik}, \quad \omega \text{ real} > 1. \tag{2}$$

Now we wish to establish the theorem referred to above and restated here.

Theorem 1: Given the two-channel unitarity equations (2), under the assumption of either symmetry or antisymmetry of the off-diagonal matrix element $S_{12}(z) = S_{21}(z)$; then, apart from arbitrary unitary phase functions $D_i(\omega)$, Eqs. (2) determine $S_{12}(z)$ and $S_{22}(z)$ in terms of $S_{11}(z)$.

Proof: The general solution of Eqs. (2), without any symmetry requirements on $S_{12}(z)$, is as follows:

$$\begin{aligned} S_{22}(z) &= S_{11}(z)[A(z)/A(1 - z)], \\ S_{12}^2(z) &= -S_{11}^2(z)[1 - S_{11}^{-1}(z) S_{11}^{-1}(1 - z)] \\ &\quad \times [A(z)/A(1 - z)], \end{aligned} \tag{3}$$

where $A(z)$ is an arbitrary function of z. The imposition of the symmetry requirement $S_{12}(z) = \pm S_{12}(-z)$ fixes $A(z)/A(1 - z)$ uniquely up to an arbitrary meromorphic unimodular function $D(\omega)$, even in ω. (This is the same type of arbitrariness admitted by the equations in the previous paper.[1]) Suppose $A(z)/A(1 - z)$ is one such solution; then the most general

[1] P. O. G. Ehrhardt and D. B. Fairlie, J. Math. Phys. **9**, 1685 (1968).

[2] V. A. Mescheryakov, Phys. Letters **24B**, 63 (1967).

[3] J. Rothleitner, Z. Physik **177**, 287 (1964).

[4] A. A. Cunningham, J. Math. Phys. **8**, 716 (1967).

solution is

$$A(z)B(z)B(-z)/A(1-z)B(1-z)B(1+z), \quad (4)$$

where

$$B(z) = B(2+z);$$

i.e., $B(z)$ is a periodic function of z in order to satisfy the symmetry requirements on $S_{12}^2(z)$. Thus $B(z)$ admits a Fourier representation of the form

$$B(z) = \sum a_n \sin n\pi z, \quad (5)$$

but this is simply a power series in ω because of the conformal transformation (1), and thus the multiplicative factor in (4) is of the character of a $D(\omega)$ function. Consequently the hypothesis of evenness or oddness of the off-diagonal matrix implies a solution for $S_{22}(z)$ and $S_{12}(z)$ in terms of $S_{11}(z)$ unique up to an energy-dependent phase factor.

This result has some far-reaching consequences, for we expect that for processes connected by a single inelastic channel, the off-diagonal matrix element must be even or odd under crossing. Now suppose the coupling to the inelastic channel is very weak, i.e., $S_{12}(z)$ is proportional to some small parameter ϵ, small enough that higher powers of ϵ may be neglected in the unitarity equation. We then study the solution of (2) in the limit $\epsilon \to 0$, maintaining the constraint imposed by the off-diagonal equation

$$S_{11}(z)S_{12}(1-z) + S_{12}(z)S_{22}(1-z) = 0 \quad (6)$$

in this limit, while taking the case of an antisymmetric $S_{12}(z)$. Upon writing

$$\begin{aligned} S_{11}(z) &= f(z)g(z), \\ S_{12}(z) &= (\epsilon/z)g(z), \\ S_{22}(z) &= h(z)g(z), \end{aligned} \quad (7)$$

with $g(z)$ a symmetric function of z, this equation takes the form

$$f(z) - (1 - z^{-1})h(1-z) = 0.$$

Expanding $f(z)$ and $h(z)$ in a formal power series in z^{-1} and comparing coefficients of like power series of z^{-1}, we find the terminating solution

$$f(z) = 1 + az^{-1}, \quad h(z) = 1 - (1+a)z^{-1}; \quad (8)$$

hence, (7) admits the simple solution

$$\begin{aligned} S_{11}(z) &= (1 + az^{-1})g(z), \\ S_{12}(z) &= (\epsilon/z)g(z), \\ S_{22}(z) &= [1 - (1+a)/z]g(z), \end{aligned} \quad (9)$$

with $g(z)$ a symmetric function of z to be determined through the unitarity equation

$$\left[1 + \frac{a + a^2 + \epsilon^2}{z(1-z)}\right] g(z)g(1-z) = 1. \quad (10)$$

Setting $a + a^2 + \epsilon^2 = 2a_1$, Eq. (10) has the formal power series

$$g(z) = 1 + \frac{a_1}{z^2} + \frac{3a_1^2 - a_1}{2z^4} + \cdots. \quad (11)$$

But this solution in the limit $\epsilon \to 0$ is just that for the elastic scattering of a meson with an $SU(2)$ quantum number a, off a "nucleon" with $SU(2)$ attribute $\frac{1}{2}$; for $g(z)$ is just $zA(z)$, with $a = \lambda$. (See Ref. 1, Eq. 9.) Thus the constraint imposed by two-particle unitarity on the diagonal elements of a two-channel S matrix when the off-diagonal element is odd is just the same as that imposed by crossing under $SU(2)$ in the limit as the inelastic coupling between the channels is switched off, and we have a diagonal S matrix, which is indicative of the existence of a superselection rule. To understand the implications of this result it is necessary to consider three-channel scattering processes, where we know that within the framework of invariance under the special unitary group two essentially different 3×3 elastic crossing matrices can arise; one for the scattering of a particle transforming as a vector off a particle transforming as an arbitary representation of $SU(2)$, and another family whose members include the matrices for the scattering of particles transforming according to the regular representation, off quarks in $SU(n)$. In the case of two-channel scattering, one finds an elastic crossing matrix only in the case of $SU(2)$. It is important to realize that, in the multichannel situation, the unitarity equations to order ϵ are essentially the same as the two-channel ones; i.e.,

$$\begin{aligned} &S_{ii}(z)S_{ii}(1-z) = 1, \\ &S_{ii}(z)S_{ij}(1-z) + S_{ij}(z)S_{jj}(1-z) = 0, \\ &\qquad\qquad \text{no summation.} \end{aligned} \quad (12)$$

To obtain a three-channel solution it is more convenient to consider a four-channel situation first. Then,

$$\begin{aligned} S_{11}(z) &= (1 + az^{-1})(1 + bz^{-1})g(a, z)g(b, z), \\ S_{22}(z) &= [1 - (1+a)z^{-1}](1 + bz^{-1})g(a, z)g(b, z), \\ S_{33}(z) &= [1 - (1+a)z^{-1}] \\ &\quad \times [1 - (1+b)z^{-1}]g(a, z)g(b, z), \\ S_{44}(z) &= (1 + az^{-1})[1 - (1+b)z^{-1}]g(a, z)g(b, z), \\ S_{12}(z) &= \epsilon[g(a, z)g(b, z)]z^{-2}(b^2 - z^2) \\ &\quad \times [(b-1)^2 - z^2] \cdots, \\ S_{13}(z) &= i\epsilon[g(a, z)g(b, z)]z^{-2}. \end{aligned} \quad (13)$$

The function $g(a, z)$ is the same as that in Eq. (11) and is dictated by unitarity. Only two of the off-diagonal elements are given: if b is integral or half-integral, then the infinite product in $S_{12}(z)$ whose symmetry characteristics are indeterminate becomes even or odd, respectively; if further we demand that the process is essentially a three-channel one, then we require

$$S_{44} \equiv 1, \tag{14}$$

i.e.,

$$b = a - 1. \tag{15}$$

Thus we have a single-parameter solution of Eqs. (2). We now let $\epsilon \to 0$ and recover a diagonal S matrix. In this limit the solution satisfies the crossing matrix for the scattering of a spin-1 meson off a particle of spin a. We emphasize that this solution has been found without the *a priori* assumption of an $SU(2)$ crossing matrix, and even some indication of the necessity for a to be integral or half-integral has emerged.[3–5] The condition (15) simplifies the solution (13) in a surprising manner. We find

$$S_{11}(z) = \frac{z + a - 1}{z - a},$$
$$S_{22}(z) = \frac{(z - a - 1)(z + a - 1)}{z^2 - a^2}, \tag{16}$$
$$S_{33}(z) = \frac{z - a - 1}{z + a}.$$

This form of solution has been given by Mescheryakov[2] who, of course, assumes the crossing relation for $SU(2)$ in order to obtain it.

However, there is a two-parameter family of crossing matrices (three if n is a parameter) which is associated with the scattering of a self-adjoint representation with young tableau $[K^{n-1}, K]$ (i.e., a tableau with $2K$ boxes in the first row, K boxes in the next $n - 2$ rows) off a quark in $SU(n)$ with negative eigenvector $(-2/n(n + K - 1), -2/n, 2K/n)$. This matrix is given by

$$\begin{pmatrix} [K^{n-1}, K+1] & [(K-1)^{n-1}K] & [K^{n-1}, 2, K-1] \\ -\dfrac{K}{(n+K-2)(n+2K-1)} & -\dfrac{(n-2)(n+K-1)}{(n+K-2)(K+1)} & \dfrac{(n+2K)(n+K-1)}{(k+1)(n+2K-1)} \\ -\dfrac{n+2K-2}{(n+K-2)(n+2K-1)} & \dfrac{K(n+K-1)}{(n+K-2)(K+1)} & \dfrac{(n+2K)}{(K+1)(n+2K-1)} \\ \dfrac{(n+2K-2)K}{(n+K-2)(n+2K-1)} & \dfrac{(n-2)K}{(n+K-2)(K+1)} & \dfrac{n+K-1}{(K+1)(n-2K-1)} \end{pmatrix} \tag{17}$$

and has no nondegenerate overlap with the previous case. The solution of the unitarity equations corresponding to this crossing matrix can be constructed, since we know its eigenvectors, and we can use unitarity to determine the power series. Here the infinitesimal off-diagonal elements, as determined by Eq. (13), contain both even and odd terms. The fact that this crossing matrix can be easily extended to a three-parameter, four-dimensional matrix which decribes the scattering of a meson belonging to the representation $[K^{n-1}, K]$ off a particle belonging to $[1^h]$ (a single column with h boxes) in $SU(n)$ offers no clue as to the possible origin of its structure in terms of a unitarity requirement.

The interpretation of these results seems to be as the emergence of a spin angular-momentum group, excluding groups of a non-rotational character. Since there is nothing in the static approximation which reveals the dimensionality of space-time of the relativistic theory of which it is a limiting form, we should be surprised to find $SU(2)$ singled out to the exclusion of possible rotation groups in higher dimensions. Indeed the four-channel solution, without the restriction (15) is a solution satisfying $SU(2) \times SU(2)$, which is isomorphic to 0(4), the rotation group in four dimensions.

ACKNOWLEDGMENTS

One of us (P. O. G. Ehrhardt) thanks the Lancashire County Council for the award of a research grant.

The other (D. B. Fairlie) thanks Professor D. Speiser and Professor F. Cerulus for hospitality at the Centre de Physique Nucleaire, University of Louvain, where part of this work was done.

[5] A. W. Martin and W. D. McGlinn, Phys. Rev. **136**, B1515 (1964).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Perturbation Theory of Nonlinear Boundary-Value Problems*

MARTIN H. MILLMAN AND JOSEPH B. KELLER
Courant Institute of Mathematical Sciences, New York University, New York, New York

(Received 5 April 1968)

A systematic perturbation theory is presented for the analysis of nonlinear problems. The lowest-order result is just that obtained by linearizing the problem, and the higher-order terms are the solutions of inhomogeneous linear problems. The essential feature of the method is the procedure for avoiding secular terms, which is based on the Lindstedt–Poincaré technique employed in celestial mechanics. The method is applied to the following nonlinear boundary value problems: (1) temperature distribution due to a nonlinear heat source or sink; (2) self-sustained oscillations of a system with infinitely many degrees of freedom; (3) forced vibrations of a "string" with a nonlinear restoring force; (4) superconductivity in a body of arbitrary shape with external magnetic field; (5) superconductivity in an infinite film with parallel magnetic field; (6) comparison of solutions of the Hartree, Fock, and Schrödinger equations for the helium atom. The results in each case are different both qualitatively and quantitatively from those of the linear theory.

1. INTRODUCTION

If the amplitude of the solution of a nonlinear problem is small enough, the solution of the corresponding linear problem provides a good approximation to it. When the amplitude becomes larger, this approximation becomes inaccurate. To obtain a better approximation, we shall employ perturbation theory. The linear problem will yield the first term in the perturbation expansion of the solution and further terms will also be determined by linear problems. This theory is well known and frequently used for problems involving nonlinear ordinary differential equations and occasionally partial differential equations. It has been applied systematically to periodic vibration problems involving nonlinear partial differential equations by Keller and Ting.[1] There is just one feature of the method which is not obvious. It is based upon the discovery by Lindstedt and Poincaré that to avoid the occurrence of secular terms in applying perturbation theory to periodic motions in celestial mechanics, it is necessary to make a perturbation expansion of the period of the motion. Once the generalization of this idea is used, the method is perfectly straightforward. Despite its simplicity, it leads to interesting qualitative as well as quantitative results. We shall demonstrate this by applying it to the six problems listed in the abstract, which are treated in Secs. 1–6.

Before turning to specific problems, we shall describe the method in general terms. Thus let F denote a nonlinear operator which depends upon a parameter λ and maps some unitary vector space into itself. For each λ we seek a vector $u(\lambda)$ satisfying

$$F(u, \lambda) = 0. \tag{1.1}$$

Suppose that $u = u_0$ is a solution when $\lambda = \lambda_0$,

$$F(u_0, \lambda_0) = 0. \tag{1.2}$$

To find u for $\lambda \neq \lambda_0$ we might try to expand $u(\lambda)$ in a Taylor series in powers of $\lambda - \lambda_0$. The coefficients in this series are the derivatives of u with respect to λ evaluated at λ_0. We seek them by differentiating Eq. (1.1) repeatedly with respect to λ and then setting $\lambda = \lambda_0$. The first differentiation yields

$$F_u(u_0, \lambda_0)u_\lambda(\lambda_0) + F_\lambda(u_0, \lambda_0) = 0. \tag{1.3}$$

If the linear operator $F_u(u_0, \lambda_0)$ is nonsingular, the unique solution of (1.3) for $u_\lambda(\lambda_0)$ is

$$u_\lambda(\lambda_0) = -[F_u(u_0, \lambda_0)]^{-1}F_\lambda(u_0, \lambda_0). \tag{1.4}$$

The higher derivatives can be found in a similar way and the resulting Taylor series is the perturbation expansion of $u(\lambda)$.

When $F_u(u_0, \lambda_0)$ is singular, (1.3) does not generally have a solution for $u_\lambda(\lambda_0)$. It does so only if $F_\lambda(u_0, \lambda_0)$ satisfies an appropriate solvability condition. Therefore the straightforward perturbation method fails in general. To overcome this failure we introduce a new parameter ϵ and express $u(\lambda)$ in the parametric form

$$u = u(\epsilon), \quad \lambda = \lambda(\epsilon). \tag{1.5}$$

In order that the solution u_0 correspond to $\epsilon = 0$ we require $u(0) = u_0$, $\lambda(0) = \lambda_0$. Then we find the derivatives of u and λ with respect to ϵ at $\epsilon = 0$ by successively differentiating (1.1) with respect to ϵ. The derivatives of λ with respect to ϵ are determined in order to satisfy the solvability conditions for the derived equations, which is possible under appropriate conditions. Then the Taylor series for $u(\epsilon)$ and $\lambda(\epsilon)$

* Supported by the U.S. Army Research Office—Durham under Contract No. DA-31-124-ARO-D-361.

[1] J. B. Keller and Lu Ting, Commun. Pure Appl. Math. 19, 371 (1966).

in powers of the parameter ϵ provide the desired perturbation expansion of $u(\lambda)$ in parametric form.

2. TEMPERATURE DISTRIBUTION DUE TO A NONLINEAR HEAT SOURCE OR SINK

A. Formulation

We wish to determine the steady-state temperature distribution $T(\mathbf{r})$ in a region D containing a distributed heat source or sink of magnitude $-\lambda S(T)$ per unit volume. The heat source or sink may be a chemical or nuclear reaction which proceeds at a temperature-dependent rate or some other temperature-dependent mechanism. At the surface B of D we suppose that the heat flux is proportional to $T - T_0$ where T_0 is the temperature outside D. Thus T satisfies the equations

$$\Delta T = \lambda S(T), \qquad \text{in } D, \tag{2.1}$$

$$\partial T/\partial n = \alpha(T - T_0), \quad \text{on } B. \tag{2.2}$$

Here α is a given proportionality constant and the parameter λ determines the source strength. We shall assume that there is no heat production when $T = T_0$ so that

$$S(T_0) = 0. \tag{2.3}$$

Then for any value of λ, a solution of (2.1) and (2.2) is the uniform temperature distribution

$$T(\mathbf{r}) = T_0. \tag{2.4}$$

Our objective is to find solutions of (2.1) and (2.2) other than (2.4).

It is natural to try to use perturbation theory to solve this problem. Thus we shall seek a one-parameter family of solutions $T(\mathbf{r}, \epsilon)$ which depends differentiably on an amplitude parameter ϵ and which reduces to (2.4) at $\epsilon = 0$. Then we shall attempt to represent $T(\mathbf{r}, \epsilon)$ in a Taylor series in ϵ about $\epsilon = 0$. However, this procedure cannot succeed unless we permit some parameter in the problem, such as λ, to depend upon ϵ also. The reason is that when $S(T)$ is nonlinear and λ is constant, there is no solution $T(\mathbf{r}, \epsilon) \not\equiv T_0$ which reduces to (2.4) at $\epsilon = 0$. This fact will be demonstrated by our results. Therefore we shall permit λ to depend upon ϵ and seek a solution $T(\mathbf{r}, \epsilon)$ and a function $\lambda(\epsilon)$ with

$$T(\mathbf{r}, 0) = T_0. \tag{2.5}$$

Both T and λ will be assumed to be sufficiently differentiable with respect to ϵ at $\epsilon = 0$. Then we shall expand both T and λ in Taylor series in ϵ about $\epsilon = 0$. This modified perturbation theory will succeed, as we shall show, provided $S(T)$ is sufficiently differentiable at T_0.

B. Linear Terms

We shall denote by $\dot{T}(\mathbf{r})$, $\dot{\lambda}$, $\ddot{T}(\mathbf{r})$, $\ddot{\lambda}$, etc., the derivatives of $T(\mathbf{r}, \epsilon)$ and $\lambda(\epsilon)$ with respect to ϵ at $\epsilon = 0$. To determine $\dot{T}(\mathbf{r})$ and $\dot{\lambda}(0)$ we differentiate (2.1) and (2.2) with respect to ϵ and set $\epsilon = 0$, obtaining

$$[\Delta - \lambda(0)S'(T_0)]\dot{T} = 0, \tag{2.6}$$

$$\partial\dot{T}/\partial n - \alpha\dot{T} = 0, \quad \text{on } B. \tag{2.7}$$

We assume that $S'(T_0) \neq 0$ and then (2.6) and (2.7) constitute a linear eigenvalue problem for $\lambda(0)$ and $\dot{T}(\mathbf{r})$. This is just the problem obtained by linearizing (2.1). It has a set of eigenvalues λ_n and corresponding orthonormal eigenfunctions $\phi_n(\mathbf{r})$, $n = 1, 2, \cdots$. Thus for some positive integer n and some constant A, we have

$$\dot{\lambda}(0) = \lambda_n, \tag{2.8}$$

$$\dot{T}(\mathbf{r}) = A\phi_n(\mathbf{r}). \tag{2.9}$$

We assume that λ_n is simple (i.e., nondegenerate) so that there is no ambiguity about which ϕ_n occurs in (2.9). The degenerate case is discussed in Sec. 2F.

By defining ϵ appropriately, we can arrange that $A = 1$. A suitable definition of ϵ for this purpose is

$$\epsilon = \int_D \dot{T}(\mathbf{r})[T(\mathbf{r}, \epsilon) - T_0]\, d\mathbf{r}. \tag{2.10}$$

Differentiating (2.10) with respect to ϵ and then setting $\epsilon = 0$ yields

$$1 = \int_D (\dot{T})^2\, d\mathbf{r}. \tag{2.11}$$

From (2.9) and (2.11) it follows that $A^2 = 1$ and thus $A = +1$ or $A = -1$. Since only the product ϵA occurs in the Taylor expansion of T, it suffices to choose $A = +1$ because the other solution is obtained by replacing ϵ by $-\epsilon$. Thus $\dot{T}(\mathbf{r})$ is uniquely determined to be

$$\dot{T}(\mathbf{r}) = \phi_n(\mathbf{r}). \tag{2.12}$$

If we differentiate (2.10) j times with respect to ϵ and set $\epsilon = 0$, we obtain

$$\int_D \dot{T}(\mathbf{r})T^{(j)}(\mathbf{r})\, d\mathbf{r} = 0, \quad j = 2, 3, \cdots. \tag{2.13}$$

As we shall see, these conditions lead to a unique determination of $T^{(j)}(\mathbf{r}), j = 2, 3, \cdots$.

C. Second-Order Terms

To find $\ddot{T}(\mathbf{r})$ and $\ddot{\lambda}$, we differentiate (2.1) and (2.2) twice with respect to ϵ, set $\epsilon = 0$, and use (2.8) and (2.3) to obtain

$$[\Delta - \lambda_n S'(T_0)]\ddot{T} = \lambda_n S''(T_0)(\dot{T})^2 + 2\dot{\lambda}S'(T_0)\dot{T}, \tag{2.14}$$

$$\partial\ddot{T}/\partial n - \alpha\ddot{T} = 0, \quad \text{on } B. \tag{2.15}$$

The problem of solving (2.14) and (2.15) for $\ddot{T}$ is an inhomogeneous form of the problem (2.6), (2.7). Therefore it has a solution only if the right side of (2.14) is orthogonal to ϕ_n, the solution of (2.6), (2.7) when $\lambda(0) = \lambda_n$. [This fact can be proved by multiplying (2.14) by ϕ_n, integrating the resulting equation over D, integrating the left-hand side by parts and using (2.6), (2.7), and (2.15).] The orthogonality condition is, if (2.12) is used for $\dot{T}$,

$$\int_D [\lambda_n S''(T_0)\phi_n^3 + 2\dot{\lambda}S'(T_0)\phi_n^2]\, d\mathbf{r} = 0. \quad (2.16)$$

Equation (2.16) is satisfied if and only if $\dot{\lambda}$ has the value

$$\dot{\lambda}_n = -\frac{\lambda_n S''(T_0)}{2S'(T_0)} \int_D \phi_n^3\, d\mathbf{r}. \quad (2.17)$$

When $\dot{\lambda}$ is given by (2.17), the problem (2.14), (2.15) has a solution $\ddot{T}$ which is unique except for the addition of an arbitrary multiple of ϕ_n. Then (2.13) with $j = 2$ makes it unique. We can express the solution $\ddot{T}$ in terms of the modified Green's function $G(\mathbf{r}, \mathbf{r}')$ of (2.14), (2.15) defined by

$$[\Delta - \lambda_n S'(T_0)]G = \delta(\mathbf{r} - \mathbf{r}') - \phi_n(\mathbf{r})\phi_n(\mathbf{r}'), \quad (2.18)$$

$$\partial G/\partial n - \alpha G = 0, \quad \mathbf{r} \text{ on } B, \quad (2.19)$$

$$\int_D G(\mathbf{r}, \mathbf{r}')\phi_n(\mathbf{r})\, d\mathbf{r} = 0. \quad (2.20)$$

Upon using (2.17) in (2.14), we can write $\ddot{T}$ in the form

$$\ddot{T}(\mathbf{r}) = \lambda_n S''(T_0) \int_D G(\mathbf{r}, \mathbf{r}') \times \left[\phi_n^2(\mathbf{r}') - \phi_n(\mathbf{r}') \int_D \phi_n^3(\mathbf{r}'')\, d\mathbf{r}''\right] d\mathbf{r}'. \quad (2.21)$$

D. Third-Order Terms

Differentiating (2.1) and (2.2) three times with respect to ϵ and setting $\epsilon = 0$ yields

$$[\Delta - \lambda_n S'(T_0)]\dddot{T} = 3\lambda_n S''(T_0)\dot{T}\ddot{T} + \lambda_n S'''(T_0)\dot{T}^3 + 3\ddot{\lambda}S'(T_0)\dot{T} + 3\dot{\lambda}S'(T_0)\ddot{T} + 3\dot{\lambda}S''(T_0)\dot{T}^2, \quad (2.22)$$

$$\partial\dddot{T}/\partial n - \alpha\dddot{T} = 0, \quad \text{on } B. \quad (2.23)$$

This is again an inhomogeneous form of (2.6), (2.7). Proceeding as before, we find from the orthogonality condition that $\ddot{\lambda}$ is given by

$$\ddot{\lambda}_n = \frac{-\lambda_n}{S'(T_0)}\left\{-\frac{[S''(T_0)]^2}{2S'(T_0)}\left(\int_D \phi_n^3\, d\mathbf{r}\right)^2 + \lambda_n[S''(T_0)]^2 \times \iint_D \phi_n^2(\mathbf{r})G(\mathbf{r}, \mathbf{r}') \times \left[\phi_n^2(\mathbf{r}') - \phi_n(\mathbf{r}')\int_D \phi_n^3(\mathbf{r}'')\, d\mathbf{r}''\right] d\mathbf{r}'\, d\mathbf{r} + \frac{S'''(T_0)}{3}\int_D \phi_n^4\, d\mathbf{r}\right\}. \quad (2.24)$$

With this value of $\ddot{\lambda}$, there is a unique solution of (2.22), (2.23), and (2.13) with $j = 3$ for $\dddot{T}(\mathbf{r})$. It is given by the integral of $G(\mathbf{r}, \mathbf{r}')$ multiplied by the right-hand side of (2.22).

E. Results

Let us now collect our results by using them to write the first three terms in the Taylor series of $\lambda(\epsilon)$ and $T(\mathbf{r}, \epsilon)$ about $\epsilon = 0$. For each positive integer n we have found a one-parameter family of solutions of (2.1) and (2.2) with the expansion

$$T(\mathbf{r}, \epsilon) = T_0 + \epsilon\phi_n(\mathbf{r}) + \frac{\epsilon^2}{2}\lambda_n S''(T_0)\int_D G(\mathbf{r}, \mathbf{r}') \times \left[\phi_n^2(\mathbf{r}') - \phi_n(\mathbf{r}')\int_D \phi_n^3(\mathbf{r}'')\, d\mathbf{r}''\right] d\mathbf{r}' + O(\epsilon^3), \quad n = 1, 2, \cdots. \quad (2.25)$$

For each value of the parameter ϵ, λ is given by

$$\lambda(\epsilon) = \lambda_n - \frac{\epsilon\lambda_n S''(T_0)}{2S'(T_0)}\int_D \phi_n^3(\mathbf{r})\, d\mathbf{r} + \frac{\epsilon^2}{2}\ddot{\lambda}_n + O(\epsilon^3), \quad n = 1, 2, \cdots. \quad (2.26)$$

Here $\ddot{\lambda}_n$ is given by (2.24). In deriving this result we have assumed that $S(T)$ has three derivatives at $T = T_0$, that $S'(T_0) \neq 0$, and that the eigenvalue λ_n is simple.

We may consider (2.25) and (2.26) as parametric equations for $T(\mathbf{r}, \lambda)$ with ϵ as the parameter. Since ϵ is the amplitude of the solution of the linearized problem, (2.25) and (2.26) are expansions in powers of the amplitude, which is determined in terms of λ by (2.26). A graph of ϵ versus λ based on (2.26), with $S''(T_0) = 0$, is shown in Fig. 1. The axis $\epsilon = 0$ denotes the trivial solution $T = T_0$, and the points $\lambda = \lambda_n$ on this axis are bifurcation points. At these points the solutions (2.25) split off from the trivial solution. To order ϵ^2, the nth branch is a parabola with its vertex at the bifurcation point $(0, \lambda_n)$. It is concave to the right or left according as $\ddot{\lambda}_n$ is positive or negative and correspondingly it represents nontrivial solutions which exist for values of λ greater than or less than λ_n.

If $S''(T_0) \neq 0$, the graph of ϵ versus λ is as shown in Fig. 2. The bifurcation points are the same as in the previous case and the branches are still parabolas to order ϵ^2. However, the vertices are no longer at the bifurcation points and are not on the line $\epsilon = 0$. Suppose the vertex of the nth branch is at $(\epsilon_{nv}, \lambda_{nv})$ to the left of the nth bifurcation point so that $\lambda_{nv} < \lambda_n$. Let the system be in the state $T = T_0$ corresponding to $\epsilon = 0$ and $\lambda < \lambda_{nv}$. If λ is slowly increased, the system may jump to a state with $\epsilon \neq 0$, given by

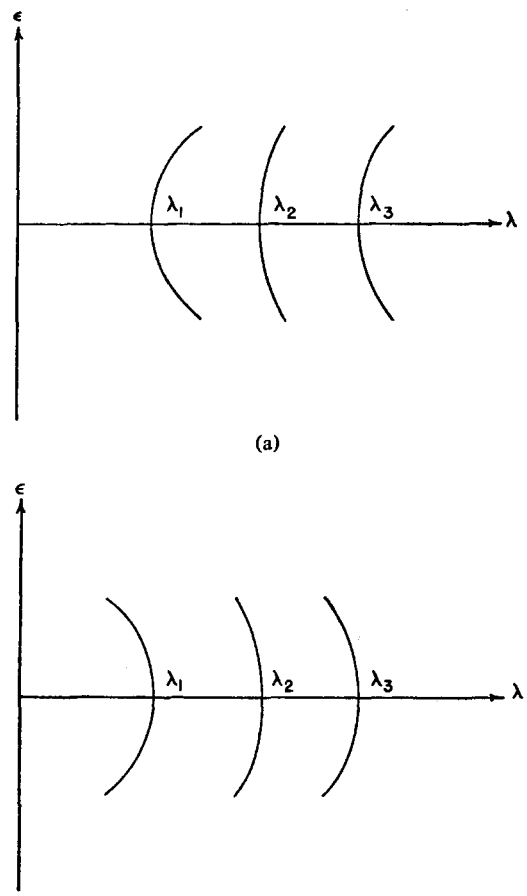


FIG. 1. The amplitude ϵ of the temperature distribution as a function of the source-strength parameter λ for the case $S'(T_0) < 0$ and $S''(T_0) = 0$. In (a), $\ddot{\lambda}_n > 0$ while in (b) $\ddot{\lambda}_n < 0$ for every n. The axis $\epsilon = 0$ represents the trivial solution $T = T_0$ and the bifurcation points are at the eigenvalues λ_n of the linearized problem.

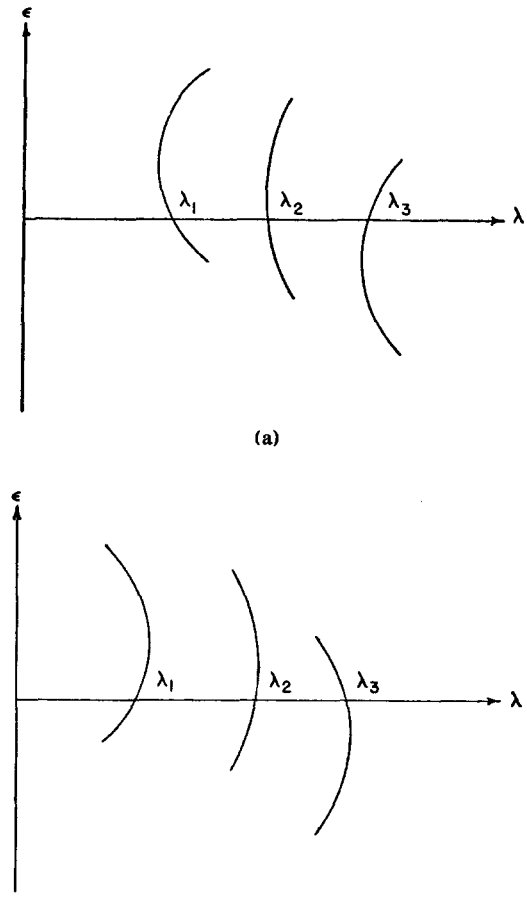


FIG. 2. The amplitude ϵ of the temperature distribution as a function of the source-strength parameter λ for the case $S'(T_0) < 0$ and $S''(T_0) > 0$. In (a), $\ddot{\lambda}_n > 0$ while in (b) $\ddot{\lambda}_n < 0$ for every n. The axis $\epsilon = 0$ represents the trivial solution $T = T_0$ and the bifurcation points are at the eigenvalues λ_n of the linearized problem.

(2.25) and (2.26), when λ reaches or exceeds λ_{nv}. Whether or not this occurs will depend upon the relative stabilities of the three states which exist for $|\epsilon|$ small and $\lambda > \lambda_{nv}$. If $\epsilon \neq 0$ and $\lambda > \lambda_{nv}$ and if λ is decreased slowly, the system must jump back to the state $T = T_0$ corresponding to $\epsilon = 0$ for some $\lambda \geq \lambda_{nv}$.

F. Multiple Eigenvalues

In Sec. B we assumed that λ_n was simple. Let us now consider the case in which it has multiplicity $k \geq 1$. Let $\phi_{n1}, \cdots, \phi_{nk}$ be corresponding orthonormal eigenfunctions. Then instead of (2.9) we have, with the A_j being constants,

$$\dot{T} = \sum_{j=1}^{k} A_j \phi_{nj}. \tag{2.27}$$

By using (2.27) in (2.11) we obtain

$$\sum_{j=1}^{k} A_j^2 = 1. \tag{2.28}$$

We now use (2.27) on the right side of (2.14) and note that the right-hand side of (2.14) must be orthogonal to each of the ϕ_{nj}. This yields the k conditions

$$\lambda_n S''(T_0) \sum_{i,j=1}^{k} a_{mij}^n A_i A_j + 2\dot{\lambda} S'(T_0) A_m = 0, \qquad m = 1, \cdots, k. \tag{2.29}$$

Here

$$a_{mij}^n = \int_D \phi_{nm} \phi_{ni} \phi_{nj} \, d\mathbf{r}. \tag{2.30}$$

Equations (2.28) and (2.29) are $k + 1$ equations for the $k + 1$ quantities $A_1, \cdots, A_k$ and $\dot{\lambda}$. From these equations we see that the solutions occur in pairs, since if $A_1, \cdots, A_k, \dot{\lambda}$ is a solution then so is $-A_1, \cdots, -A_k, -\dot{\lambda}$. Only one solution of each pair need be considered since the other is obtained by replacing ϵ by $-\epsilon$.

Corresponding to each solution of (2.28) and (2.29) there is a solution $\dot{T}$ of (2.14) and (2.15) which is unique up to an additive arbitrary linear combination

of $\phi_{n1}, \cdots, \phi_{nk}$. One inhomogeneous relation among the coefficients in this linear combination is provided by (2.13) with $j = 2$. Then k additional inhomogeneous linear relations follow from the orthogonality of ϕ_{nj}, $j = 1, \cdots, k$ and the right side of (2.22), which is necessary for the solvability of (2.22) and (2.23) for $\ddot{T}$. These are $k + 1$ inhomogeneous linear equations for the k coefficients and $\ddot{\lambda}$. If they have a unique solution then $\ddot{T}$ and $\ddot{\lambda}$ are uniquely determined and then so will all the subsequent derivatives of T and λ be uniquely determined. This is so because the same coefficient matrix which occurs in the equations just described will occur in all subsequent sets of equations.

G. Example

Let us apply our result to a domain D which is a rectangular parallelepiped with incommensurable edge lengths L_1, L_2, and L_3 upon the surface of which $T = T_0$. Then $\alpha = \infty$ and the normalized solution of (2.6) and (2.7) is

$$\dot{T}_{nmp} = \left(\frac{8}{L_1L_2L_3}\right)^{\frac{1}{2}} \sin\frac{n\pi x}{L_1} \sin\frac{m\pi y}{L_2} \sin\frac{p\pi z}{L_3}, \quad n, m, p \neq 0, \tag{2.31}$$

$$\lambda_{nmp} = -\frac{\pi^2}{S'(T_0)}\left(\frac{n^2}{L_1^2} + \frac{m^2}{L_2^2} + \frac{p^2}{L_3^2}\right). \tag{2.32}$$

Then (2.17) yields

$$\dot{\lambda}_{nmp} = 0. \tag{2.33}$$

If $S''(T_0) = 0$ then (2.21) yields

$$\ddot{T}(\mathbf{r}) = 0. \tag{2.34}$$

Upon using (2.31)–(2.34) in (2.24) we obtain

$$\ddot{\lambda}_{nmp} = \frac{9\pi^2 S'''(T_0)}{8L_1L_2L_3[S'(T_0)]^2}\left(\frac{n^2}{L_1^2} + \frac{m^2}{L_2^2} + \frac{p^2}{L_3^2}\right). \tag{2.35}$$

Thus $\ddot{\lambda}_{nmp}$ has the same sign as $S'''(T_0)$. Therefore all the curves of ϵ as a function of λ are concave to the right if $S'''(T_0) > 0$, and to the left if $S'''(T_0) < 0$, as is shown in Fig. 1. Upon combining (2.31)–(2.35) we obtain

$$T(\mathbf{r}, \epsilon) = T_0 + \epsilon\left(\frac{8}{L_1L_2L_3}\right)^{\frac{1}{2}} \times \sin\frac{n\pi x}{L_1} \sin\frac{m\pi y}{L_2} \sin\frac{p\pi z}{L_3} + O(\epsilon^3), \tag{2.36}$$

$$\lambda(\epsilon) = \left\{-\frac{\pi^2}{S'(T_0)} + \frac{\epsilon^2 9\pi^2 S'''(T_0)}{16L_1L_2L_3[S'(T_0)]^2} + O(\epsilon^3)\right\} \times \left(\frac{n^2}{L_1^2} + \frac{m^2}{L_1^2} + \frac{p^2}{L_3^2}\right). \tag{2.37}$$

H. Stability of Steady-State Solutions

Now that we have obtained perturbation expansions of the steady-state temperature distributions satisfying (2.1) and (2.2), we shall examine their stability. For this purpose we shall employ the time-dependent forms of (2.1) and (2.2), which are

$$\Delta\theta - \gamma\theta_t = \lambda S(\theta), \quad \text{in } D, \tag{2.38}$$

$$\frac{\partial\theta}{\partial n} = \alpha(\theta - T_0), \quad \text{on } B. \tag{2.39}$$

Here $\theta(\mathbf{r}, t)$ is the temperature at $\mathbf{r}$ at time t, γ is a positive constant related to the thermal conductivity of the material in D, and the other quantities are the same as in Sec. 2A. We shall seek a solution $\theta(\mathbf{r}, t, \epsilon, \eta)$, depending upon the new small parameter η, which reduces to $T_n(\mathbf{r}, \epsilon)$ at $\eta = 0$. Thus we require

$$\theta(\mathbf{r}, t, \epsilon, 0) = T_n(\mathbf{r}, \epsilon). \tag{2.40}$$

The parameter η may be thought of as the amplitude of some initial deviation of the temperature from its steady-state value. Our objective is to determine whether such a deviation will grow or decay in time. If there is any initial deviation which grows we call the steady state T_n unstable, while if all deviations decay we call T_n stable.

We shall attempt to represent θ by a finite Taylor expansion in η. The value of θ for $\eta = 0$ is given by (2.40). To find $\theta_\eta(\mathbf{r}, t, \epsilon, 0)$ we differentiate (2.38) and (2.39) with respect to η and set $\eta = 0$ to obtain

$$\Delta\theta_\eta - \gamma\theta_{\eta t} = \lambda_n S'(T_n)\theta_\eta, \tag{2.41}$$

$$\frac{\partial\theta_\eta}{\partial n} = \alpha\theta_\eta. \tag{2.42}$$

To solve (2.41) and (2.42) we employ separation of variables. Since the coefficients are independent of t, the t dependence of a product solution θ is clearly exponential. Therefore we seek a product solution of the form

$$\theta_\eta(\mathbf{r}, t, \epsilon, 0) = u(\mathbf{r}, \epsilon)e^{\beta(\epsilon)t}. \tag{2.43}$$

Then (2.41) and (2.42) become the following equations for u and β:

$$\Delta u - \{\gamma\beta(\epsilon) + \lambda_n(\epsilon)S'[T_n(\mathbf{r}, \epsilon)]\}u = 0, \tag{2.44}$$

$$\partial u/\partial n = \alpha u. \tag{2.45}$$

To find $u(\mathbf{r}, \epsilon)$ and $\beta(\epsilon)$ we shall represent them as Taylor series in ϵ. Then setting $\epsilon = 0$ in (2.44) yields

$$\Delta u(\mathbf{r}, 0) - [\gamma\beta(0) + \lambda_n(0)S'(T_0)]u(\mathbf{r}, 0) = 0. \tag{2.46}$$

Setting $\epsilon = 0$ in (2.45), or in any derivative of (2.45)

with respect to ϵ, merely replaces u by $u(\mathbf{r}, 0)$ or by the corresponding derivative. We shall refer to any of the resulting equations as (2.45). The solution of (2.46) and (2.45) is $u(\mathbf{r}, 0) \equiv 0$ unless the coefficient of $u(\mathbf{r}, 0)$ equals $S'(T_0)$ times an eigenvalue λ_j. This is the case if $\beta(0)$ has the value

$$\beta(0) = [\lambda_j - \lambda_n]S'(T_0)/\gamma. \tag{2.47}$$

Then if λ_j has multiplicity one, $u(\mathbf{r}, 0)$ is a constant B_j, times the corresponding eigenfunction ϕ_j,

$$u_j(\mathbf{r}, 0) = B_j\phi_j(\mathbf{r}). \tag{2.48}$$

If $S'(T_0) < 0$ then (2.47) shows that $\beta(0) > 0$ if $\lambda_j < \lambda_n$. Therefore $T_n(\mathbf{r}, \epsilon)$ is unstable for $|\epsilon|$ sufficiently small unless λ_n is the smallest eigenvalue. Similarly if $S'(T_0) > 0$, $T_n(\mathbf{r}, \epsilon)$ is unstable for $|\epsilon|$ sufficiently small unless λ_n is the largest eigenvalue. All the eigenvalues have the same sign as $-S'(T_0)$ and there is a smallest one when $S'(T_0) < 0$ and a largest when $S'(T_0) > 0$. We shall call this smallest or largest eigenvalue, which is the eigenvalue of smallest absolute value, λ_1. Then we have shown that for $|\epsilon|$ small, T_n is unstable for $n \neq 1$. For $n = 1$, (2.47) shows that $\beta(0) < 0$ for $j \neq 1$ and $\beta(0) = 0$ for $j = 1$. In this last case we must consider further terms in the Taylor expansion of $\beta(\epsilon)$ to determine its stability.

The stability of the trivial solution $T = T_0$ for arbitrary λ is also determined by the preceding equations with T_n and λ_n replaced by T_0 and λ. Then (2.47) shows that if $S'(T_0) < 0$, then T_0 is stable for $\lambda < \lambda_1$ and unstable for $\lambda > \lambda_1$. However, if $S'(T_0) > 0$ then T_0 is stable for $\lambda > \lambda_1$ and unstable for $\lambda < \lambda_1$.

Let us denote by $u_j(\mathbf{r}, \epsilon)$ and $\beta_j(\epsilon)$ the solution of (2.44) and (2.45) which reduces to $u_j(\mathbf{r}, 0) = B_j\phi_j(\mathbf{r})$ and to $\beta(0)$ given by (2.47) at $\epsilon = 0$. To obtain the derivative $\dot{u}_j$, $\dot{\beta}_j$ of this solution with respect to ϵ we differentiate (2.44) with respect to ϵ and set $\epsilon = 0$, which yields

$$[\Delta - \gamma\beta_j(0) - \lambda_n(0)S'(T_0)]\dot{u}_j = \gamma\dot{\beta}_jB_j\phi_j + \dot{\lambda}_nS'(T_0)B_j\phi_j + \lambda_nS''(T_0)B_j\phi_j\phi_n. \tag{2.49}$$

In order that (2.49) have a solution u_j satisfying (2.45), the right-hand side of (2.49) must be orthogonal to ϕ_j, the solution of the homogeneous problem. Upon solving this orthogonality relation for β_j we obtain

$$\dot{\beta}_j = -\frac{1}{\gamma}\left[\lambda_nS''(T_0)\int\phi_j^2\phi_n\,d\mathbf{r} + \dot{\lambda}_nS'(T_0)\right]. \tag{2.50}$$

By using the expression (2.17) for $\dot{\lambda}_n$ we can write (2.50) as follows for $j = n$:

$$\dot{\beta}_n = \dot{\lambda}_nS'(T_0)/\gamma. \tag{2.51}$$

When (2.50) holds, Eqs. (2.49) and (2.45) have solutions which we shall write in the form

$$\dot{u}_j = B_jv_j(\mathbf{r}) + C_j\phi_j(\mathbf{r}). \tag{2.52}$$

Here $v_j(\mathbf{r})$ is a particular solution of (2.49) and (2.45) with $B_j = 1$, and C_j is an arbitrary constant.

For $j = n$, $\beta_n(0) = 0$. If also $\dot{\beta}_n = 0$ we must find $\ddot{\beta}_n$. To do this we consider the case $j = n$, differentiate (2.44) twice with respect to ϵ at $\epsilon = 0$ and obtain

$$\begin{aligned}[\Delta - \lambda_n(0)S'(T_0)]\ddot{u}_n &= \gamma\ddot{\beta}_nB_n\phi_n + \ddot{\lambda}_nS'B_n\phi_n + 2\dot{\lambda}_nS''B_n\phi_n^2 \\ &\quad + \lambda_nS'''B_n\phi_n + \lambda_nS''B_n\phi_n\ddot{T}_n \\ &\quad + (4\dot{\lambda}_nS' + 2\lambda_nS''\phi_n)(B_nv_n + C_n\phi_n).\end{aligned} \tag{2.53}$$

In order that (2.53) have a solution satisfying (2.45), the right-hand side of (2.53) must be orthogonal to ϕ_n. By using the expression (2.17) for λ_n in this orthogonality condition, we find that the coefficient of C_n vanishes and then B_n cancels. Therefore we can solve the orthogonality condition for $\ddot{\beta}_n$ with the result

$$\begin{aligned}\ddot{\beta}_n = -\frac{1}{\gamma}\Big[&4\dot{\lambda}_nS'\int\phi_nv_n\,d\mathbf{r} \\ &+ 2\lambda_nS''\int\phi_n^2v_n\,d\mathbf{r} + \lambda_nS''\int\phi_n^2\ddot{T}_n\,d\mathbf{r} \\ &+ \lambda_nS'''\int\phi_n^4\,d\mathbf{r} + \ddot{\lambda}_nS' - 4S'(\dot{\lambda}_n)^2/\lambda_n\Big].\end{aligned} \tag{2.54}$$

Here we have used (2.17) to simplify the last term. All quantities in (2.54) are to be evaluated at $\epsilon = 0$ so that S', S'', and S''' are evaluated at T_0.

Let us now collect our results by writing the Taylor series for θ and β in the following forms, in which we add together the product solutions of (2.44) and (2.45) for θ_η:

$$\begin{aligned}\theta(\mathbf{r}, t, \epsilon, \eta) = T_n(\mathbf{r}, \epsilon) + \eta\sum_{j=1}^{\infty}\{&B_j\phi_j(\mathbf{r}) + \epsilon[B_jv_j(\mathbf{r}) \\ &+ C_j\phi_j(\mathbf{r})] + O(\epsilon^2)\}e^{\beta_j(\epsilon)t} + O(\eta^2),\end{aligned} \tag{2.55}$$

$$\begin{aligned}\beta_j(\epsilon) = &\frac{1}{\gamma}(\lambda_j - \lambda_n)S'(T_0) \\ &- \frac{\epsilon}{\gamma}\left[\lambda_nS''(T_0)\int\phi_j^2\phi_n\,d\mathbf{r} + \dot{\lambda}_nS'(T_0)\right] \\ &+ \frac{\epsilon^2}{2}\ddot{\beta}_j + O(\epsilon^3).\end{aligned} \tag{2.56}$$

For $j \neq n$ we have not calculated $\ddot{\beta}_j$, while for $j = n$ it is given by (2.54). We have seen that some $\beta_j(0)$ is positive unless $n = 1$ so T_n is unstable for $n \neq 1$. The stability of $T_1(\mathbf{r}, \epsilon)$ is determined by the sign of $\beta_1(\epsilon)$, which is given by (2.56) with $j = n = 1$.

Let us consider the stability of the solution (2.43) for the example treated in Sec. 2G. Equations (2.33) and (2.51) show that

$$\dot{\beta}_{nmp} = 0. \tag{2.57}$$

Then for $j = n$ the right-hand side of (2.49) is zero, since in Sec. 2G we assumed that $S''(T_0) = 0$. Therefore a particular solution of (2.49) for $j = n$ is

$$v_{nmp}(\mathbf{r}) \equiv 0. \tag{2.58}$$

Upon using (2.33), (2.34), and (2.58) in (2.54) we obtain $\ddot{\beta}_{nmp}$. Simplifying it with the aid of (2.24) yields

$$\ddot{\beta}_{nmp} = -\frac{\lambda_{nmp}S'''(T_0)}{\gamma}\int \phi_{nmp}^4\, d\mathbf{r} - \frac{\ddot{\lambda}_{nmp}S'(T_0)}{\gamma}$$
$$= \frac{2S'(T_0)}{\gamma}\ddot{\lambda}_{nmp}. \tag{2.59}$$

Now using (2.32), (2.57), and (2.59) in (2.56) for the case $n = m = p = 1$, we obtain

$$\beta_{111}(\epsilon) = \frac{\epsilon^2 9\pi^2 S'''(T_0)}{8L_1L_2L_3\gamma S'(T_0)}\left(\frac{1}{L_1^2} + \frac{1}{L_2^2} + \frac{1}{L_3^2}\right) + O(\epsilon^3). \tag{2.60}$$

Thus the lowest mode $n = m = p = 1$ is stable or unstable according as $S'''(T_0)/S'(T_0)$ is negative or positive.

3. SELF-SUSTAINED OSCILLATIONS OF A SYSTEM WITH INFINITELY MANY DEGREES OF FREEDOM

A. Formulation

We wish to find periodic solutions of the nonlinear equation

$$u_{tt} - u_{xx} + u = \epsilon f(u_t), \quad 0 < x < \pi. \tag{3.1}$$

The periodicity and boundary conditions are

$$u(x, t + 2\pi/\omega) = u(x, t), \tag{3.2}$$

$$u(0, t) = u(\pi, t) = 0. \tag{3.3}$$

In (3.2), ω is an undetermined angular frequency. In (3.1) ϵ is a prescribed small parameter and $f(u_t)$ is a given nonlinear "damping" force which is of the same sign as u_t when $|u_t|$ is small, and of the opposite sign when $|u_t|$ is large. Thus it represents real damping for $|u_t|$ small and negative damping for $|u_t|$ large. An example of such a function is

$$f_0(u_t) = u_t - \tfrac{1}{3}(u_t)^3. \tag{3.4}$$

When $f = f_0$ and u is independent of x, (3.1) is just the van der Pol equation. It has one periodic solution which is called a self-sustained oscillation. This suggests that (3.1)–(3.3) will also have at least one solution. This problem is a prototype for the study of self-sustained oscillations in nonlinear systems with infinitely many degrees of freedom.

B. Perturbation Method

Let us introduce $t' = \omega t$ and $u'(x, t') = u(x, t)$ in (3.1)–(3.3) and then omit the primes to obtain (3.3) again and

$$\omega^2 u_{tt} - u_{xx} + u = \epsilon f(\omega u_t), \tag{3.5}$$

$$u(x, t + 2\pi) = u(x, t). \tag{3.6}$$

If (3.3), (3.5), and (3.6) have a solution u with angular frequency ω, then both u and ω depend upon ϵ. We shall seek a solution $u(x, t, \epsilon)$ and a corresponding angular frequency $\omega(\epsilon)$ which are differentiable with respect to ϵ at $\epsilon = 0$. Then we shall represent them by finite Taylor series in ϵ. To obtain equations for these quantities at $\epsilon = 0$ we set $\epsilon = 0$ in (3.3), (3.5), and (3.6), which yields a linear problem. To obtain equations for their derivatives at $\epsilon = 0$ we differentiate (3.3), (3.5), and (3.6) repeatedly with respect to ϵ and then set $\epsilon = 0$. Thus this is just the perturbation method of solution.

C. Zero-Order Terms

Let us set $\epsilon = 0$ in (3.3), (3.5), and (3.6) and write $u(x, t, 0) = u_0(x, t)$, $\omega(0) = \omega_0$. Then we obtain (3.3) and (3.6) with u replaced by u_0 and

$$\omega_0^2 u_{0tt} - u_{0xx} + u_0 = 0. \tag{3.7}$$

For each positive integer n, (3.3), (3.6), and (3.7) have solutions. By choosing the origin of t appropriately we can write these solutions as

$$u_0 = A_n \sin nx \cos t, \tag{3.8}$$

$$\omega_0 = (1 + n^2)^{\frac{1}{2}}, \quad n = 1, 2, \cdots. \tag{3.9}$$

The amplitude A_n is undetermined so far.

D. First-Order Terms

We differentiate (3.3), (3.5), and (3.6) once with respect to ϵ and set $\epsilon = 0$ to obtain

$$\omega_0^2 \dot{u}_{tt} - \dot{u}_{xx} + \dot{u} = -2\omega_0\dot{\omega}u_{0tt} + f(\omega_0 u_{0t}) \tag{3.10}$$

and (3.3) and (3.6) with u replaced by $\dot{u}$. Here $\dot{u} = u_\epsilon(x, t, 0)$ and $\dot{\omega} = \omega_\epsilon(0)$. The problem of solving (3.10) subject to (3.3) and (3.6) is an inhomogeneous form of the problem of solving (3.7) subject to the same conditions. Therefore (3.10) will have a solution only if an appropriate solvability condition is satisfied. We show this and derive the condition by multiplying (3.10) by $w(x, t)$ and integrating with respect to x

from 0 to π and with respect to t from 0 to 2π. Making use of integrations by parts, we obtain

$$\int_0^{2\pi}\int_0^{\pi} \dot{u}(\omega_0^2 w_{tt} - w_{xx} + w)\,dx\,dt + \int_0^{\pi}[\dot{u}_t w - \dot{u}w_t]_{t=0}^{2\pi}\,dx + \int_0^{2\pi}[\dot{u}_x w - \dot{u}w_x]_{x=0}^{\pi}\,dt = -2\omega_0\dot{\omega}\int_0^{2\pi}\int_0^{\pi} w u_{0tt}\,dx\,dt + \int_0^{2\pi}\int_0^{\pi} w f(\omega_0 u_{0t})\,dx\,dt. \tag{3.11}$$

The first term on the left-hand side of (3.11) vanishes for either of the two functions

$$w = \sin nx \sin t, \quad w = \sin nx \cos t. \tag{3.12}$$

For both choices of w the second and third terms also vanish, by virtue of (3.3) and (3.6), and (3.11) yields the two conditions

$$\omega_0\dot{\omega}\pi^2 A_n + \int_0^{2\pi}\int_0^{\pi} \sin nx \cos t \times f(-\omega_0 A_n \sin nx \sin t)\,dx\,dt = 0, \tag{3.13}$$

$$\int_0^{2\pi}\int_0^{\pi} \sin nx \sin t \times f(-\omega_0 A_n \sin nx \sin t)\,dx\,dt = 0. \tag{3.14}$$

Here we have evaluated u_{0t} and u_{0tt} from (3.8) and have performed the integrations indicated in the first term on the right-hand side of (3.11). Condition (3.13) determines $\dot{\omega}$ in terms of A_n and (3.14) determines A_n.

To obtain explicit results, we now take $f = f_0$ given by (3.4). Then (3.14) has the three roots

$$A_n = 0, \quad \pm(4/\sqrt{3})\omega_0. \tag{3.15}$$

We choose $A_n = +(4/\sqrt{3})\omega_0$ since the other sign would just change the phase of the solution and $A_n = 0$ would yield $u = 0$. The integral in (3.13) is zero, so we obtain

$$\dot{\omega} = 0. \tag{3.16}$$

We now insert (3.15) and (3.16) into the right-hand side of (3.10) and find

$$\omega_0^2\dot{u}_{tt} - \dot{u}_{xx} + \dot{u} = \tfrac{4}{27}\sqrt{3}\,(\sin 3nx \sin 3t - 3 \sin 3nx \times \sin t - 3 \sin nx \sin 3t). \tag{3.17}$$

The solution of (3.17), (3.3), and (3.6) is

$$\dot{u} = -(\sqrt{3}/54)\sin 3nx \sin 3t + [\sqrt{3}/18(1 + n^2)]\sin nx \sin 3t - (\sqrt{3}/18n^2)\sin 3nx \sin t + D \sin nx \sin t + E \sin nx \cos t. \tag{3.18}$$

Here D and E are arbitrary constants.

E. Second-Order Terms

We differentiate (3.3), (3.5), and (3.6) twice with respect to ϵ, set $\epsilon = 0$ and use (3.16) to obtain

$$\omega_0^2\ddot{u}_{tt} - \ddot{u}_{xx} + \ddot{u} = -2\omega_0\dot{\omega}\dot{u}_{0tt} + 2\omega_0\dot{u}_t f'(\omega_0 u_{0t}). \tag{3.19}$$

Again (3.3) and (3.6) hold with $\ddot{u}$ in place of u and $\ddot{u} = u_{\epsilon\epsilon}(x, t, 0)$, $\ddot{\omega} = \omega_{\epsilon\epsilon}(0)$. Once more we have an inhomogeneous form of (3.7). As before, we multiply by w and integrate over x and t. Making use of (3.8), (3.9), and (3.18), we obtain

$$E = 0, \tag{3.20}$$

$$\ddot{\omega} = -(1 + n^2)^{\frac{1}{2}}/72[1 - (1/n^2) + 9/(1 + n^2)]. \tag{3.21}$$

F. Results

Let us collect our results and reintroduce the original variable t. Then we have, for each integer n, a mode of vibration given by

$$u_n(x, t, \epsilon) = [4/\sqrt{3}(1 + n^2)^{\frac{1}{2}}]\sin nx \cos \omega_n t + \epsilon[-(\sqrt{3}/54)\sin 3nx \sin 3\omega_n t + [\sqrt{3}/18(1 + n^2)]\sin nx \sin 3\omega_n t - (\sqrt{3}/18n^2)\sin 3nx \sin \omega_n t + D \sin nx \sin \omega_n t] + O(\epsilon^3), \tag{3.22}$$

$$\omega_n(\epsilon) = (1 + n^2)^{\frac{1}{2}} - [\epsilon^2(1 + n^2)^{\frac{1}{2}}/144] \times [1 - (1/n^2) + 9/(1 + n^2)] + O(\epsilon^3). \tag{3.23}$$

We see that as ϵ increases the frequency of each mode decreases owing to the presence of the nonlinear damping force.

4. FORCED VIBRATIONS OF A "STRING" WITH A NONLINEAR RESTORING FORCE

A. Formulation

We consider periodic small vibrations of a model equation for a uniform string fixed at one end and harmonically driven at the other, under the action of a nonlinear restoring force. The relevant equations are

$$u_{tt} - u_{xx} = \epsilon F(u), \quad 0 < x < \pi, \tag{4.1}$$

$$u(0, t) = 0, \quad u(\pi, t) = A \cos \omega t, \tag{4.2}$$

$$u(x, t + 2\pi/\omega) = u(x, t), \tag{4.3}$$

$$\frac{1}{2}\int_0^{\pi}\left[u_x^2 + u_t^2 - 2\epsilon\int_0^{u} F(u)\,du\right]_{t=0} dx = E. \tag{4.4}$$

Here u is the analog of the displacement of the string, ϵ is a small parameter, $F(u)$ is a nonlinear restoring force, A is the amplitude, ω is the angular frequency of the forcing term, and E is the energy of

the motion at $t = 0$. In Sec. 4G we treat the case in which the forcing function is in the differential equation, rather than in the boundary condition.

It seems natural in (4.2) to regard ω as prescribed in advance, but it turns out that $u(x, t, \omega, \epsilon)$ is not regular in ϵ at $\epsilon = 0$ for ω at a resonance frequency. Therefore we prescribe the energy E in (4.4) and consider ω and u to be functions of E, A, and ϵ. Then u and ω are both regular in ϵ, as we shall see. Equation (4.3) expresses the requirement that the response have the same period as the excitation. It is convenient to introduce $t' = \omega t$ in (4.1)–(4.4) and then to omit the primes. This yields

$$\omega^2 u_{tt} - u_{xx} = \epsilon F(u), \quad 0 < x < \pi, \tag{4.5}$$

$$u(0, t) = 0, \quad u(\pi, t) = A \cos t, \tag{4.6}$$

$$u(x, t + 2\pi) = u(x, t), \tag{4.7}$$

$$\frac{1}{2}\int_0^\pi \left[u_x^2 + \omega^2 u_t^2 - 2\epsilon \int_0^u F(u)\, du\right]_{t=0} dx = E. \tag{4.8}$$

B. Perturbation Expansion

We represent $u(x, t, \epsilon)$ and $\omega(E, A, \epsilon)$ by their Taylor series in ϵ about $\epsilon = 0$ as follows:

$$u(x, t, \epsilon) = u_0(x, t) + \epsilon u_1(x, t) + \cdots, \tag{4.9}$$

$$\omega^2(E, A, \epsilon) = \omega_0^2(E, A) + \epsilon \omega_1(E, A) + \cdots. \tag{4.10}$$

Equations for the first coefficients in these expansions can be obtained by setting $\epsilon = 0$ in (4.5)–(4.8), for the next coefficients by differentiating with respect to ϵ and putting $\epsilon = 0$, etc.

C. Zero-Order Terms

Setting $\epsilon = 0$ in (4.5)–(4.8) yields

$$\omega_0^2 u_{0tt} - u_{0xx} = 0, \quad 0 < x < \pi, \tag{4.11}$$

$$\frac{1}{2}\int_0^\pi [u_{0x}^2 + \omega_0^2 u_{0t}^2]_{t=0}\, dx = E, \tag{4.12}$$

and (4.6), (4.7) with u_0 in place of u. The solution of (4.11), (4.6), and (4.7) is

$$u_0 = (A/\sin \omega_0 \pi) \sin \omega_0 x \cos t. \tag{4.13}$$

Insertion of (4.13) into (4.12) then yields the following linear response relation between E and ω_0:

$$\frac{4\pi E}{A^2}\left(\frac{\sin \omega_0 \pi}{\omega_0 \pi}\right)^2 = 1 + \frac{\sin 2\omega_0 \pi}{2\omega_0 \pi}. \tag{4.14}$$

Figure 3 shows a graph of E vs ω_0 based on (4.14). It can be seen that for given E there are no roots for ω_0 unless

$$E \geq A^2/2\pi. \tag{4.15}$$

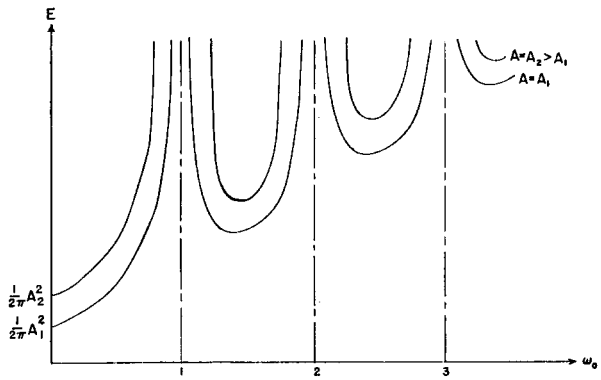


FIG. 3. Response curves of the linear "string" driven at one end. The energy E is shown as a function of the forcing frequency ω_0, for two values of the forcing amplitude A_1 and A_2, based upon Eq. (4.14). Resonances occur at the free vibration frequencies $\omega_0 = 1, 2, 3, \cdots$.

As $E \to \infty$, the number of roots increases monotonically and they draw closer to the values $\omega_0 = 1, 2, \cdots$. Thus for each E satisfying (4.15) there is at least one root for ω_0, which is not an integer, and corresponding to this root there is a finite u_0 given by (4.13).

D. First-Order Terms

We next differentiate (4.5)–(4.8) once with respect to ϵ and set $\epsilon = 0$ to obtain

$$\omega_0^2 u_{1tt} - u_{1xx} = -\omega_1 u_{0tt} + F(u_0), \quad 0 < x < \pi, \tag{4.16}$$

$$\int_0^\pi \left[u_{0x}u_{1x} + \omega_0^2 u_{0t}u_{1t} + 2\omega_1 u_{0t}^2 - \int_0^{u_0} F(u)\, du\right]_{t=0} dx = 0, \tag{4.17}$$

and (4.6), (4.7) with u_1 in place of u. The homogeneous form of (4.16) has no solution satisfying (4.6), so (4.16) has a unique solution. It is readily found to be

$$u_1 = \frac{A\omega_1 \cos t}{\sin \omega_0 \pi} \sum_{k=1}^{\infty} \frac{B_k}{k^2 - \omega_0^2} \sin kx + \sum_{\substack{k=1 \\ j=0}}^{\infty} \frac{C_{jk}}{k^2 - \omega_0^2 j^2} \cos jt \sin kx. \tag{4.18}$$

Here the coefficients B_k and C_{jk} are defined by

$$\sin \omega_0 x = \sum_{k=1}^{\infty} B_k \sin kx, \tag{4.19}$$

$$F(u_0) = \sum_{\substack{k=1 \\ j=0}}^{\infty} C_{jk} \cos jt \sin kx, \quad 0 < x < \pi. \tag{4.20}$$

Insertion of (4.18) into (4.17) then yields

$$\omega_1 = (\sin \omega_0 \pi / A)\left[\frac{2}{\pi}\int_0^\pi \left(\int_0^{u_0} F(u)\, du\right)_{t=0} dx - \sum_{j,k} \frac{kc_{jk}D_k}{k^2 - \omega_0^2 j^2}\right] \Big/ \sum_k \frac{kB_kD_k}{k^2 - \omega_0^2}. \tag{4.21}$$

Here D_k is defined by

$$\frac{A\omega_0}{\sin \omega_0 \pi} \cos \omega_0 x = \sum_{k=0}^{\infty} D_k \cos kx, \quad 0 < x < \pi. \quad (4.22)$$

E. Results

Let us collect our results and reintroduce the original variable t. We have determined the periodic modes of vibration in the form

$$u(x, t, \epsilon) = (A/\sin \omega_0 \pi) \sin \omega_0 x \cos \omega t + \epsilon\left[\frac{A\omega_1}{\sin \omega_0 \pi} \sum_{k=1}^{\infty} \frac{B_k}{k^2 - \omega_0^2} \sin kx \cos \omega t + \sum_{\substack{j=0 \\ k=1}}^{\infty} \frac{C_{jk}}{k^2 - \omega_0^2 j^2} \sin kx \cos j\omega t\right] + O(\epsilon^2), \quad (4.23)$$

$$\omega^2(E, A, \epsilon) = \omega_0^2 + \epsilon\omega_1 + O(\epsilon^2). \quad (4.24)$$

In (4.23) and (4.24), ω_0 and ω_1 are given as functions of E and A by (4.14) and (4.21), respectively. Therefore (4.23) gives $u(x, t, E, A, \epsilon)$ and (4.24) gives $\omega(E, A, \epsilon)$. We may consider these equations to be parametric equations for $u(x, t, \omega, A, \epsilon)$ with E being the parameter. From (4.23) we see that the nonlinearity introduces harmonics of the applied frequency. The relation (4.24) can be viewed as a response relation which determines E, the energy at $t = 2\pi n/\omega$, as a function of ϵ, A, and ω.

F. Example

Let us choose

$$F(u) = -u^3. \quad (4.25)$$

Then for $|\omega_0 - n| \ll 1, n = 1, 2, \cdots$, we can simplify

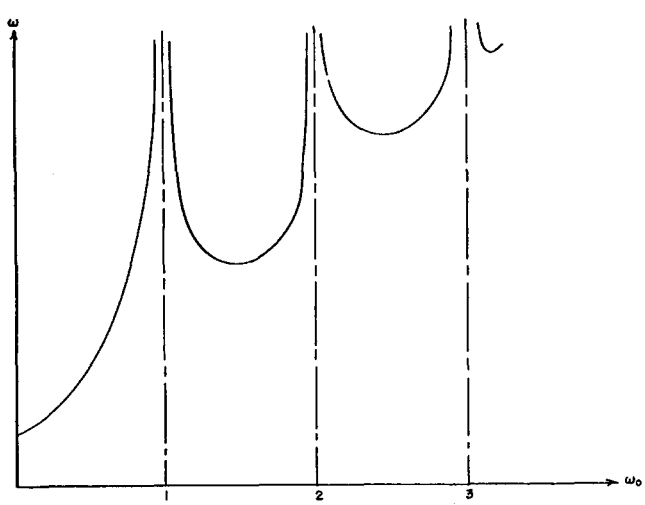


FIG. 4. The frequency ω of the nonlinear "string" driven at one end as a function of ω_0 for $\epsilon > 0$, based on Eq. (4.26).

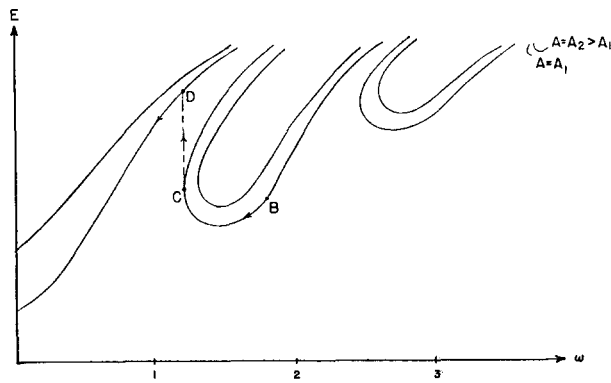


FIG. 5. Response curve of the nonlinear "string" driven at one end. The energy E is shown as a function of the forcing frequency ω for $\epsilon > 0$ for two values of the forcing amplitude A_1 and A_2. The curves are obtained by eliminating ω_0 from Eqs. (4.14) and (4.26) or equivalently from Figs. 3 and 4. If ω is slowly diminished from the value corresponding to the point B, the amplitude will jump from that at C to that at D when the frequency passes through the value corresponding to C.

(4.21) and use the result in (4.24) to obtain

$$\omega^2 \simeq \omega_0^2 + (9\epsilon A^2/16\pi^2)(\omega_0 - n)^{-2} + O(\epsilon^2). \quad (4.26)$$

The qualitative behavior of ω as a function of ω_0 for $\epsilon > 0$ based on (4.26) is indicated in Fig. 4.

We now eliminate ω_0 between (4.14) and (4.26), or between Figs. 3 and 4, to obtain E as a function of ω. The result is the nonlinear response relation shown in Fig. 5. The nonlinearity has bent the linear response curves to the right, just as it does for an ordinary differential equation involving a "hard" spring. We note the following interesting properties of the nonlinear response curves:

(1) Resonance infinities no longer occur.
(2) The possibility of "jump phenomena" arises.

To see the jumps, we consider an experiment in which the amplitude of the excitation is fixed at A_1, and in which the frequency is slowly decreased. If we start at point B in Fig. 5, we see that the energy E gradually decreases as ω decreases and then E slowly increases until point C is reached, where the tangent is vertical. With a further decrease in ω, there occurs a sudden jump in the energy to point D, after which E gradually decreases again. Such jump phenomena are well known in systems governed by ordinary differential equations.

G. Response to a Distributed Force

Let us now modify the problem of Sec. 3A by introducing the distributed forcing function $A \sin x \cos \omega t$ in (4.1) and keeping both end points fixed. Then we

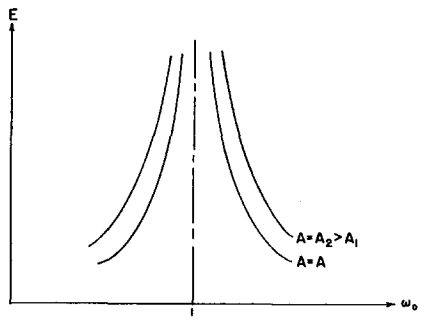


FIG. 6. Response of the linear "string" to a distributed force. The energy E is shown as a function of the forcing frequency ω_0, for two values of the forcing amplitude A_1 and A_2, based upon Eq. (4.30). Resonance occurs at the free-vibration frequency $\omega_0 = 1$.

must change (4.5) and (4.6) to

$$\omega^2 u_{tt} - u_{xx} - \epsilon F(u) = A \sin x \cos t, \quad 0 < x < \pi, \tag{4.27}$$

$$u(0, t) = u(\pi, t) = 0. \tag{4.28}$$

Thus we must solve (4.27), (4.28), (4.7), and (4.8). By setting $\epsilon = 0$ we find for the solution of (4.27), (4.28), and (4.7) the result

$$u_0 = [A/(1 - \omega_0^2)] \sin x \cos t. \tag{4.29}$$

Upon introducing (4.29) into (4.8), we obtain the linearized response relation (see Fig. 6):

$$E = A^2/4(1 - \omega_0^2). \tag{4.30}$$

There is only one resonance because the driving force, being proportional to $\sin x$, excites only the lowest mode of the free system.

We next differentiate (4.27) and (4.8) once with respect to ϵ and then set $\epsilon = 0$ to obtain

$$\omega_0^2 u_{1tt} - u_{1xx} = -\omega_1 u_{0tt} + F(u_0), \tag{4.31}$$

$$\int_0^\pi \left[u_{0x} u_{1x} + \omega_0^2 u_{0t} u_{1t} + 2\omega_1 u_{0t}^2 - \int_0^{u_0} F(u)\, du \right]_{t=0} dx = 0. \tag{4.32}$$

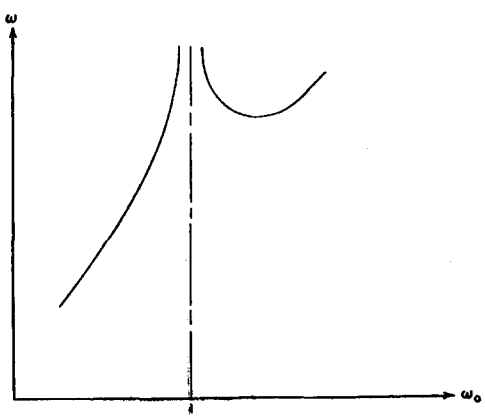


FIG. 7. The frequency ω of the nonlinear "string" driven by a distributed force as a function of ω_0 for $\epsilon > 0$, based on Eq. (4.35).

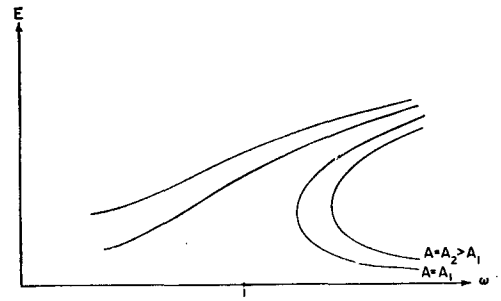


FIG. 8. Response curve of the nonlinear "string" driven by a distributed force. The energy E is shown as a function of the forcing frequency ω for $\epsilon > 0$ for two values of the forcing amplitude A_1 and A_2. The curves are obtained by eliminating ω_0 from Eq. (4.30) and Eq. (4.35) or equivalently from Figs. 6 and 7.

Differentiation of (4.28) and (4.7) at $\epsilon = 0$ merely replaces u by u_1 in them. The solution of (4.31), (4.28), and (4.7) is

$$u_1 = \frac{\omega_1 A}{(1 - \omega_0^2)^2} \sin x \cos t + \sum_{\substack{k=1 \\ j=0}}^{\infty} \frac{C_{jk}}{\omega_0^2 j^2 - k^2} \sin kx \cos jt. \tag{4.33}$$

Here the coefficients C_{jk} are defined by (4.20). Insertion of (4.33) into (4.32) then yields

$$\omega_1 = \frac{(1 - \omega_0^2)^2}{A} \sum_{j=0}^{\infty} \frac{C_{j1}}{1 - \omega_0^2 j^2} - \frac{2(1 - \omega_0^2)^3}{\pi A^2} \int_0^\pi \int_0^\beta F(u)\, du\, dx, \tag{4.34}$$

where the upper limit of integration

$$\beta = [A/(1 - \omega_0^2)] \sin x.$$

Again our result is $u = u_0 + \epsilon u_1 + O(\epsilon^2)$ and $\omega^2 = \omega_0^2 + \epsilon\omega_1 + O(\epsilon^2)$. To exemplify it let us take $F(u) = u^3$. Then after some algebra we can simplify (4.34), for ω_0 near unity, and obtain the approximate result

$$\omega^2 \simeq \omega_0^2 + [9A^2\epsilon/16(1 - \omega_0^2)^2] + \cdots, \quad |\omega_0 - 1| \ll 1. \tag{4.35}$$

The qualitative behavior of ω as a function of ω_0 is shown in Fig. 7.

We now eliminate ω_0 between (4.30) and the relation $\omega^2 = \omega_0^2 + \epsilon\omega_1 + O(\epsilon^2)$, or between Figs. 6 and 7, to obtain E as a function of ω. This is the nonlinear response relation shown in Fig. 8. Again, we find a behavior analogous to that of a "hard" spring.

5. SUPERCONDUCTIVITY IN A BODY OF ARBITRARY SHAPE WITH EXTERNAL MAGNETIC FIELD

A. Formulation

We consider a superconducting body of permeability μ placed in an originally uniform external magnetic field h. If the body occupies a region D bounded by surface S, the Landau–Ginzburg equations together with boundary and other conditions are, in appropriate units,

$$\{(i/k)\nabla + \mathbf{a}\}^2\phi = \phi(1 - |\phi|^2), \quad \text{in } D, \tag{5.1}$$

$$-\nabla \times \nabla \times \mathbf{a} = (i/2k)(\phi^*\nabla\phi - \phi\nabla\phi^*) + \mathbf{a}\,|\phi|^2, \quad \text{in } D, \tag{5.2}$$

$$\nabla \times \nabla \times \mathbf{a} = 0, \quad \text{outside } D, \tag{5.3}$$

$$\hat{\mathbf{n}} \cdot \{(i/k)\nabla + \mathbf{a}\}\phi = 0, \quad \text{on } S, \tag{5.4}$$

$$[\hat{\mathbf{n}} \times \mathbf{a}] = 0, \tag{5.5}$$

$$[(1/\mu)\hat{\mathbf{n}} \times (\nabla \times \mathbf{a})] = 0, \tag{5.6}$$

$$\nabla \times \mathbf{a} \to h\hat{\mathbf{z}}, \quad \text{as} \quad r \to \infty. \tag{5.7}$$

Here ϕ is the "order parameter" which measures the amount of superconductivity, $\mathbf{a}$ is the magnetic vector potential, $\hat{\mathbf{n}}$ is the outward unit normal on S, $\hat{\mathbf{z}}$ is a unit vector in the external field direction, [ ] stands for the jump in the enclosed quantity across S, r is the distance from the origin, and k is a given physical constant. In (5.4) $\mathbf{a}$ denotes the limiting value of the vector potential as S is approached from the interior of D. Equation (5.3) states that there are no currents in the vacuum region. Equations (5.5) and (5.6) are the usual boundary conditions on the magnetic vector potential for an interface between two permeable media. In the vacuum $\mu = 1$. Finally, (5.7) asserts that the magnetic field reduces to the applied field far from the body.

These equations are gauge invariant in the sense that if $(\phi, \mathbf{a})$ is a solution so is $(e^{ikf}\phi, \mathbf{a} + \nabla f)$, for an arbitrary function f. This can be seen by inserting the latter pair into (5.1)–(5.7) and observing that all terms containing f cancel out and the system reduces to the original one for $(\phi, \mathbf{a})$. This freedom allows us to impose the additional gauge condition

$$\nabla \cdot \mathbf{a} = 0, \tag{5.8}$$

which proves convenient in later calculations. Even (5.8) does not fix $\mathbf{a}$ uniquely since $(e^{ikg}\phi, \mathbf{a} + \nabla g)$ is a solution of (5.1)–(5.7) which also satisfies (5.8) when g is any harmonic function. This has no effect on the physical quantities since $\nabla \times \mathbf{a}$ and $|\phi|^2$ are invariant to such a transformation.

A solution of (5.1)–(5.8) which corresponds to the "normal" state is

$$\phi \equiv 0, \quad \mathbf{a} = h\mathbf{a}_N. \tag{5.9}$$

Insertion of (5.9) into (5.1)–(5.8) yields for $\mathbf{a}_N$ the equations

$$\nabla \times \nabla \times \mathbf{a}_N = 0, \quad \text{except on } S, \tag{5.10}$$

$$[\hat{\mathbf{n}} \times \mathbf{a}_N] = 0, \tag{5.11}$$

$$[(1/\mu)\hat{\mathbf{n}} \times (\nabla \times \mathbf{a}_N)] = 0, \tag{5.12}$$

$$\nabla \times \mathbf{a}_N \to \hat{\mathbf{z}}, \quad \text{as} \quad r \to \infty, \tag{5.13}$$

$$\nabla \cdot \mathbf{a}_N = 0. \tag{5.14}$$

Equations (5.10)–(5.14) correspond to the problem of determining the vector potential for a permeable body carrying no current placed in an originally uniform unit external field. Well-known methods of magnetostatics are available for its solution.

B. Perturbation Expansion

We seek a "superconducting" solution of (5.1)–(5.8), i.e., one in which $\phi \not\equiv 0$, which depends continuously on a parameter ϵ and which reduces to (5.9) for $\epsilon = 0$. It is convenient to introduce ϵ, ψ, and $\mathbf{A}$ through the equations

$$\phi = \epsilon^{\frac{1}{2}}\psi, \tag{5.15}$$

$$\mathbf{a} = h\mathbf{a}_N + \epsilon\mathbf{A}, \quad \epsilon \geq 0. \tag{5.16}$$

Insertion of (5.15)–(5.16) into (5.1)–(5.8) yields

$$[(i/k)\nabla + h\mathbf{a}_N]^2\psi - \psi = -2\epsilon[\tfrac{1}{2}|\psi|^2 + h\mathbf{a}_N \cdot \mathbf{A} + (i/k)\mathbf{A} \cdot \nabla + \tfrac{1}{2}\epsilon\mathbf{A}^2]\psi, \quad \text{in } D, \tag{5.17}$$

$$-\nabla \times \nabla \times \mathbf{A} = (i/2k)(\psi^*\nabla\psi - \psi\nabla\psi^*) + |\psi|^2(h\mathbf{a}_N + \epsilon\mathbf{A}), \quad \text{in } D, \tag{5.18}$$

$$\nabla \times \nabla \times \mathbf{A} = 0, \quad \text{outside } D, \tag{5.19}$$

$$\hat{\mathbf{n}} \cdot [(i/k)\nabla + h\mathbf{a}_N]\psi = -\epsilon(\hat{\mathbf{n}} \cdot \mathbf{A})\psi, \quad \text{on } S, \tag{5.20}$$

$$[\hat{\mathbf{n}} \times \mathbf{A}] = 0, \tag{5.21}$$

$$[(1/\mu)\hat{\mathbf{n}} \times (\nabla \times \mathbf{A})] = 0, \tag{5.22}$$

$$\nabla \times \mathbf{A} \to 0, \quad \text{as} \quad r \to \infty, \tag{5.23}$$

$$\nabla \cdot \mathbf{A} = 0. \tag{5.24}$$

We shall regard h as a function of ϵ, to be determined along with ψ and $\mathbf{A}$. All three quantities will be represented as finite Taylor series in ϵ about $\epsilon = 0$ as follows:

$$h = h_0 + \epsilon h_1 + \cdots, \tag{5.25}$$

$$\psi = \psi_0 + \epsilon\psi_1 + \cdots, \tag{5.26}$$

$$\mathbf{A} = \mathbf{A}_0 + \epsilon\mathbf{A}_1 + \cdots. \tag{5.27}$$

We now proceed to determine the coefficients in these expansions.

C. Zero-Order Terms

Equations governing the zero-order coefficients are obtained by setting $\epsilon = 0$ in (5.17)–(5.24). This yields

$$[(i/k)\nabla + h_0\mathbf{a}_N]^2\psi_0 - \psi_0 = 0, \quad \text{in } D, \tag{5.28}$$

$$-\nabla \times \nabla \times \mathbf{A}_0 = (i/2k)(\psi_0^*\nabla\psi_0 - \psi_0\nabla\psi_0^*) + h_0|\psi_0|^2\mathbf{a}_N, \quad \text{in } D, \tag{5.29}$$

$$\nabla \times \nabla \times \mathbf{A}_0 = 0, \quad \text{outside } D, \tag{5.30}$$

$$\hat{\mathbf{n}} \cdot [(i/k)\nabla + h_0\mathbf{a}_N]\psi_0 = 0, \quad \text{on } S, \tag{5.31}$$

$$[\hat{\mathbf{n}} \times \mathbf{A}_0] = 0, \tag{5.32}$$

$$[(1/\mu)\hat{\mathbf{n}} \times (\nabla \times \mathbf{A}_0)] = 0, \tag{5.33}$$

$$\nabla \times \mathbf{A}_0 \to 0, \quad \text{as} \quad r \to \infty, \tag{5.34}$$

$$\nabla \cdot \mathbf{A}_0 = 0. \tag{5.35}$$

Equations (5.28) and (5.31) determine a discrete set of eigenvalues for h_0, which are independent of the gauge of $\mathbf{a}_N$. We define the "critical field" h_c to be the largest positive one of these and set $\psi_0 = \alpha\chi$ where χ is a corresponding normalized eigenfunction and α is a constant.

We can make $|\alpha| = 1$ by defining ϵ so that ψ satisfies the normalization condition

$$\int_D \psi^*\psi \, d\mathbf{r} = 1. \tag{5.36}$$

Setting $\epsilon = 0$ in (5.36) and using the fact that $\psi_0 = \alpha\chi$ yields $|\alpha|^2 = 1$. We shall fix the gauge of ψ_0 by choosing $\alpha = 1$ so that $\psi_0 = \chi$. Then (5.29), (5.30), and (5.32)–(5.35) is the problem of determining the vector potential $\mathbf{A}_0$ due to a permeable body carrying a specified real current distribution, since the right-hand side of (5.29) is known. Again, well-known methods of solution are available. However, in certain one-dimensional cases no solution exists. We deal with this in Sec. 6.

We have now shown how to determine the critical field h_c and the zero-order approximations to ψ and $\mathbf{A}$.

D. First-Order Terms

We next differentiate (5.17)–(5.24) once with respect to ϵ and then set $\epsilon = 0$ to obtain

$$[(i/k)\nabla + h_c\mathbf{a}_N]^2\psi_1 - \psi_1 = -2[\tfrac{1}{2}|\psi_0|^2 + h_1[h_0a_N^2 + (i/k)\mathbf{a}_N \cdot \nabla] + h_0\mathbf{A}_0 \cdot \mathbf{a}_N + (i/k)\mathbf{A}_0 \cdot \nabla]\psi_0, \quad \text{in } D, \tag{5.37}$$

$$-\nabla \times \nabla \times \mathbf{A}_1 = (i/2k)(\psi_0^*\nabla\psi_1 + \psi_1^*\nabla\psi_0 - \psi_0\nabla\psi_1^* - \psi_1\nabla\psi_0^*) + h_1|\psi_0|^2\mathbf{a}_N + h_0\mathbf{a}_N(\psi_0\psi_1^* + \psi_1\psi_0^*) + |\psi_0|^2\mathbf{A}_0, \quad \text{in } D, \tag{5.38}$$

$$\nabla \times \nabla \times \mathbf{A}_1 = 0, \quad \text{outside } D, \tag{5.39}$$

$$\hat{\mathbf{n}} \cdot [(i/k)\nabla + h_c\mathbf{a}_N]\psi_1 = -\hat{\mathbf{n}} \cdot (\mathbf{A}_0 + h_1\mathbf{a}_N)\psi_0, \quad \text{on } S, \tag{5.40}$$

$$[\hat{\mathbf{n}} \times \mathbf{A}_1] = 0, \tag{5.41}$$

$$[(1/\mu)\hat{\mathbf{n}} \times (\nabla \times \mathbf{A}_1)] = 0, \tag{5.42}$$

$$\nabla \times \mathbf{A}_1 \to 0, \quad \text{as} \quad r \to \infty, \tag{5.43}$$

$$\nabla \cdot \mathbf{A}_1 = 0. \tag{5.44}$$

We consider first the system (5.38)–(5.39) and (5.41)–(5.44) for $\mathbf{A}_1$. Assuming ψ_1 and h_1 were known, these equations would again correspond to the problem of determining the vector potential due to a permeable body carrying a known current distribution. Thus $\mathbf{A}_1$ is fixed, once ψ_1 and h_1 are given. Now (5.37) and (5.40) for ψ_1 are inhomogeneous forms of (5.28) and (5.31) and therefore have a solution only if an appropriate solvability condition is satisfied. We derive this condition by multiplying both sides of (5.37) from the left by ψ_0^* and integrating over D. After considerable manipulation, integration by parts and use of (5.14), (5.28), (5.29), (5.31), (5.35), and (5.40), this yields the (real) result

$$h_1 = \frac{\frac{1}{2}\int |\psi_0|^4 \, d\mathbf{r} + \int \mathbf{A}_0 \cdot [\nabla \times \nabla \times \mathbf{A}_0] \, d\mathbf{r}}{\int \mathbf{a}_N \cdot [\nabla \times \nabla \times \mathbf{A}_0] \, d\mathbf{r}}. \tag{5.45}$$

When h_1 is given by (5.45), then (5.37) and (5.40) have a solution ψ_1. It is unique as a consequence of the equation obtained by differentiating (5.36) with respect to ϵ and then setting $\epsilon = 0$.

E. Results

We have shown how to determine a solution of (5.1)–(5.8) of the form

$$h = h_c + \epsilon h_1 + \cdots, \tag{5.46}$$

$$\phi = \epsilon^{\frac{1}{2}}[\psi_0 + \epsilon\psi_1 + \cdots], \tag{5.47}$$

$$\mathbf{a} = h\mathbf{a}_N + \epsilon[\mathbf{A}_0 + \epsilon\mathbf{A}_1 + \cdots], \tag{5.48}$$

Equation (5.48) leads to a magnetic field

$$\mathbf{B} = h\nabla \times \mathbf{a}_N + \epsilon\nabla \times \mathbf{A}_0 + O(\epsilon^2). \tag{5.49}$$

If $h_1 < 0$, as we shall show is the case in an example, we have a solution for all values of the external field slightly below a certain critical value h_c. If $\hat{\mathbf{z}} \cdot (\nabla \times \mathbf{A}_0) < 0$ in D, as is the case in the example below, the magnetic field within the body is less than its value in the normal state. This is the Meissner effect.

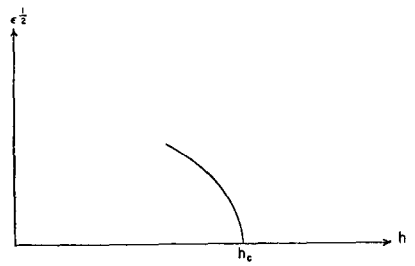


FIG. 9. The amplitude $\epsilon^{\frac{1}{2}}$ of the order parameter in a superconductor in a magnetic field as a function of the external magnetic field strength h based upon Eq. (5.46). The normal state corresponds to $\epsilon^{\frac{1}{2}} = 0$ for any h. Bifurcation of superconducting solutions occurs at the eigenvalues of a linear problem, the largest of which is called h_c. As h decreases below h_c, the amplitude $\epsilon^{\frac{1}{2}}$ increases like $(h_c - h)^{\frac{1}{2}}$.

Since ψ is normalized, $\epsilon^{\frac{1}{2}}$ is the amplitude of ϕ and (5.46) gives the dependence of this amplitude on h. This dependence is shown in Fig. 9.

F. Example

We illustrate the preceding results by applying them to the problem of a superconducting cylinder of infinite length and radius r_0 in a uniform external field parallel to its axis. We employ cylindrical coordinates and assume all quantities to be independent of z. Then a solution of (5.10)–(5.14) is given by

$$\mathbf{a}_N = \begin{cases} \frac{1}{2}\mu r\hat{\boldsymbol{\theta}}, & r < r_0, \\ \frac{1}{2}[r + r_0^2(\mu - 1)/r]\hat{\boldsymbol{\theta}}, & r > r_0. \end{cases} \tag{5.50}$$

Now (5.28) and (5.31) become

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi_0}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \psi_0}{\partial \theta^2} - ih_0 k\mu \frac{\partial \psi_0}{\partial \theta} + k^2(1 - \tfrac{1}{4}h_0^2\mu^2 r^2)\psi_0 = 0, \tag{5.51}$$

$$\frac{\partial \psi_0}{\partial r} = 0, \quad \text{at} \quad r = r_0. \tag{5.52}$$

We seek a solution of (5.51) of the form

$$\psi_0(r, \theta) = R(r)e^{im\theta}. \tag{5.53}$$

Then $R(r)$ must satisfy

$$r^2R'' + rR' + [k^2r^2(1 + m\mu h_0/k - \tfrac{1}{4}h_0^2\mu^2 r^2) - m^2]R = 0, \tag{5.54}$$

$$R'(r_0) = R'(0) = 0. \tag{5.55}$$

We shall now examine the symmetric solutions for which $m = 0$. Then it can be shown that (5.54) and (5.55) yield for h_0^2 a discrete set of eigenvalues which tend to minus infinity and of which at least one is positive. Let us set $\mathbf{A}_0 = A_0(r)\hat{\boldsymbol{\theta}}$. Then (5.29), (5.30), and (5.32)–(5.35) become

$$\frac{d}{dr}\left[\frac{1}{r}\frac{d}{dr}(rA_0)\right] = \begin{cases} \frac{1}{2}h_c\mu r R^2, & r < r_0, \\ 0, & r > r_0, \end{cases} \tag{5.56}$$

$$[A_0] = 0, \quad \text{at} \quad r = r_0, \tag{5.57}$$

$$\left[\frac{1}{\mu}\frac{d}{dr}(rA_0)\right] = 0, \quad \text{at} \quad r = r_0, \tag{5.58}$$

$$\frac{1}{r}\frac{d}{dr}(rA_0) \to 0, \quad \text{as} \quad r \to \infty. \tag{5.59}$$

The solution of (5.56)–(5.59) is

$$A_0(r) = \frac{-1}{r}\int_0^r r''\int_{r''}^{r_0} \tfrac{1}{2}h_c\mu r' R^2(r')\,dr'\,dr'', \quad r < r_0, \tag{5.60}$$

$$A_0(r) = \frac{-1}{r}\int_0^{r_0} r''\int_{r''}^{r_0} \tfrac{1}{2}h_c\mu r' R^2(r')\,dr'\,dr'', \quad r > r_0. \tag{5.61}$$

From (5.60) the magnetic field is given by

$$\nabla \times \mathbf{A}_0 = \frac{1}{r}\frac{d}{dr}(rA_0)\hat{\mathbf{z}} = -\tfrac{1}{2}\mu h_c\hat{\mathbf{z}}\int_r^{r_0} r' R^2(r')\,dr', \quad r < r_0. \tag{5.62}$$

Thus $\hat{\mathbf{z}} \cdot (\nabla \times \mathbf{A}_0) < 0$ for $r < r_0$, so the Meissner effect occurs.

From (5.45) and (5.60) we obtain

$$h_1 = -\left(\int R^4\,dS - \mu h_c\int rR^2A_0\,dS\right)\Big/\tfrac{1}{2}h_c\mu^2\int r^2R^2\,dS. \tag{5.63}$$

Here the integrations are extended over the cross section $r < r_0$. From (5.60) we see that $A_0 < 0$ for $r < r_0$ and then (5.63) yields $h_1 < 0$. Thus the superconducting solution which we have found exists for all values of the external field slightly less than h_c.

6. SUPERCONDUCTIVITY IN AN INFINITE FILM WITH PARALLEL MAGNETIC FIELD

A. Formulation

We consider a superconducting film with plane parallel faces which occupies the region $-d \leq x \leq d$ in a uniform external magnetic field h directed along the z axis. The analysis of Sec. 5 is not applicable to this case because the film is not finite, so that equations (5.29), (5.30), and (5.32)–(5.35) have no solution. Therefore we analyze this case separately. If all field quantities depend only on x, the Landau–Ginzburg

equations of Sec. 5 can be written in the form

$$k^{-2}\phi'' = \phi(\phi^2 - 1 + h^2B^2), \quad -d < x < d, \tag{6.1}$$

$$B'' = \phi^2 B, \quad -d < x < d, \tag{6.2}$$

$$\phi' = 0, \quad \text{at } x = \pm d, \tag{6.3}$$

$$B' = 1, \quad \text{at } x = \pm d. \tag{6.4}$$

Here ϕ is the "order parameter," h is the magnitude of the external field, hB is the magnetic vector potential, $H = hB'$ is the magnetic field, and k is a given physical constant.

For any h, a trivial solution of (6.1)–(6.4) which corresponds to the "normal" state is

$$\phi = 0, \tag{6.5}$$

$$B = x + c. \tag{6.6}$$

Here c is an arbitrary constant. Odeh[2] has proved the existence of a nontrivial solution of (6.1)–(6.4) for $h < h_c$, where h_c is a critical field which we shall determine. His method of proof for h just slightly less then h_c is similar to our procedure of perturbation calculation.

B. Perturbation Expansion

We seek a solution of (6.1)–(6.4) depending on a parameter ϵ which reduces to (6.5) and (6.6), with special values of c and h, when $\epsilon = 0$. It is convenient to introduce ϵ, ψ, and A through the defining equations

$$\phi = \epsilon^{\frac{1}{2}}\psi, \quad B = x + c + \epsilon A. \tag{6.7}$$

$$\int_{-d}^{d} \psi^*\psi \, dx = 1. \tag{6.8}$$

Insertion of (6.7) into (6.1)–(6.4) yields

$$k^{-2}\psi'' + [1 - h^2(x + c)^2]\psi = \epsilon\{\psi^3 + h^2\psi[2(x + c)A + \epsilon A^2]\}, \tag{6.9}$$

$$A'' = (x + c)\psi^2 + \epsilon\psi^2 A, \tag{6.10}$$

$$\psi'(\pm d) = 0, \tag{6.11}$$

$$A'(\pm d) = 0. \tag{6.12}$$

We shall attempt to represent ψ, A, c, and h^2 in powers of ϵ:

$$\psi = \psi_0 + \epsilon\psi_1 + \cdots, \tag{6.13}$$

$$A = A_0 + \epsilon A_1 + \cdots, \tag{6.14}$$

$$h^2 = h_0^2 + \epsilon h_1 + \cdots, \tag{6.15}$$

$$c = c_0 + \epsilon c_1 + \cdots. \tag{6.16}$$

[2] F. Odeh, J. Math. Phys. 8, 2351 (1968).

C. Zero-Order Terms

Equations for the zero-order coefficients in (6.13)–(6.16) can be obtained by putting $\epsilon = 0$ in (6.9)–(6.12). This yields

$$k^{-2}\psi_0'' + [1 - h_0^2(x + c_0)^2]\psi_0 = 0, \tag{6.17}$$

$$A_0'' = (x + c_0)\psi_0^2, \tag{6.18}$$

$$\psi_0'(\pm d) = 0, \tag{6.19}$$

$$A_0'(\pm d) = 0. \tag{6.20}$$

The change of variable

$$w = (4k^2h_0^2)^{\frac{1}{4}}(x + c_0) \tag{6.21}$$

converts (6.17) and (6.18) into

$$\frac{d^2}{dw^2}\psi_0(w) + \left[\frac{k}{2h_0} - \tfrac{1}{4}w^2\right]\psi_0(w) = 0, \tag{6.22}$$

$$\frac{d}{dw}\psi_0(w) = 0, \quad \text{at} \quad w = (4k^2h_0^2)^{\frac{1}{4}}(c_0 \pm d). \tag{6.23}$$

Equation (6.22) is Weber's equation of index, $(k/2h_0) - \frac{1}{2}$. For fixed c_0, (6.22) and (6.23) determine a set of eigenvalues for h_0 and a corresponding set of normalized eigenfunctions. Hereafter we let $h_0^2(c_0) = h_c^2$ denote the largest positive eigenvalue and let $\chi(c_0)$ be a corresponding normalized eigenfunction. Then $\psi_0 = \alpha\chi$ where α is a constant and from (6.8) with $\epsilon = 0$ it follows that $|\alpha| = 1$. We fix the gauge by setting $\alpha = 1$ so that $\psi_0(c_0) = \chi(c_0)$.

We now insert $\psi_0(c_0)$ on the right-hand side of (6.18) and integrate once to obtain

$$A_0' = \int_{-d}^{x} (x' + c_0)\psi_0^2(c_0)\, dx'. \tag{6.24}$$

Equation (6.20) at $x = d$ requires that we choose c_0 to satisfy

$$\int_{-d}^{d} (x + c_0)\psi_0^2(c_0)\, dx = 0. \tag{6.25}$$

Another integration then yields

$$A_0 = \int_{-d}^{x} (x - x')(x' + c_0)\psi_0^2\, dx' + k_1. \tag{6.26}$$

Here k_1 is an arbitrary constant.

We have now determined $\psi_0(c_0)$ as the normalized eigenfunction corresponding to the largest positive eigenvalue h_c^2. A_0 is given by (6.26) and c_0 is a solution of (6.25). Saint-James and De Gennes[3] have solved (6.17) and (6.18) for the half-space $x > 0$ with the conditions (6.19) and (6.20) at $x = 0$.

[3] D. Saint-James and P. De Gennes, Phys. Letters 7, 306 (1963).

D. First-Order Terms

Equations for the first-order coefficients in (6.13)–(6.16) can be obtained by differentiating (6.9)–(6.12) once with respect to ϵ and then setting $\epsilon = 0$. This yields

$$k^{-2}\psi_1'' + [1 - h_0^2(x + c_0)^2]\psi_1 = h_1(x + c_0)^2\psi_0 + 2h_0^2c_1(x + c_0)\psi_0 + \psi_0^3 + 2h_0^2(x + c_0)\psi_0A_0, \quad (6.27)$$

$$A_1'' = 2(x + c_0)\psi_0\psi_1 + A_0\psi_0^2 + c_1\psi_0^2. \quad (6.28)$$

$$\psi_1'(\pm d) = 0, \quad (6.29)$$

$$A_1'(\pm d) = 0. \quad (6.30)$$

The problem of solving (6.27) subject to (6.29) is an inhomogeneous form of the problem of solving (6.17) subject to the same condition. Therefore (6.27) has a solution only if an appropriate solvability condition is satisfied. We show this and derive the condition by multiplying (6.27) by ψ_0 and integrating over x. Upon integrating by parts and using (6.17), (6.19), (6.25), and (6.29), we find that the right-hand side vanishes and we obtain

$$h_1 = -\frac{\int_{-d}^{d}\psi_0^4\,dx + 2h_0^2\int_{-d}^{d}(x + c_0)A_0\psi_0^2\,dx}{\int_{-d}^{d}(x + c_0)^2\psi_0^2\,dx}. \quad (6.31)$$

In view of (6.25), the constant k_1 in A_0 will make no contribution to the integral in (6.31), so that h_1 is uniquely determined. When (6.31) holds, (6.27) can be solved for ψ_1. The solution is unique as a consequence of the equation obtained by differentiating (6.8) and setting $\epsilon = 0$. We note that ψ_1 is linear in c_1, since the right-hand side of (6.27) is.

By proceeding as in the case of (6.18) and (6.20) above, we find that (6.28) and (6.30) have a solution only if c_1 satisfies the linear equation

$$\int_{-d}^{d}[2(x + c_0)\psi_0\psi_1 + A_0\psi_0^2 + c_1\psi_0^2]\,dx = 0. \quad (6.32)$$

When (6.32) holds, (6.28) can be solved for A_1. Thus h_1 has been found, and equations for the determination of ψ_1, c_1, and A_1 have been obtained.

E. Results for the Symmetric State

By inspection, $c_0 = 0$ is a solution of (6.25), since (6.17) and (6.19) show that ψ_0 is an even function of x when $c_0 = 0$. Then (6.24) shows that

$$A_0' < 0 \quad \text{for} \quad -d < x < d. \quad (6.33)$$

By definition the magnetic field inside the film is

$$H = hB' = h + \epsilon hA_0' + \cdots. \quad (6.34)$$

Therefore, for sufficiently small $\epsilon > 0$ the field is less inside the film than outside. This is the Meissner effect.

From (6.26), $A_0 - k_1$ is odd, so (6.31) shows that

$$h_1 < 0. \quad (6.35)$$

From (6.15), the square of the external field is

$$h^2 = h_c^2 + \epsilon h_1 + \cdots. \quad (6.36)$$

Thus (6.35) and (6.36) show that, in the superconducting solution, the external field h is below the critical value h_c, in agreement with experiment.

The order parameter is $\phi = \epsilon^{\frac{1}{2}}\psi$ and ψ is normalized, so $\epsilon^{\frac{1}{2}}$ is the amplitude of ϕ. Thus (6.36) gives the relation between the external field strength and the amplitude of the order parameter. This relation is shown in Fig. 9.

7. COMPARISON OF SOLUTIONS OF THE HARTREE, FOCK, AND SCHRÖDINGER EQUATIONS FOR THE HELIUM ATOM

A. Formulation

We wish to compare perturbation solutions of the nonlinear Hartree and Fock equations with the standard perturbation solution of the Schrödinger equation for the helium atom. To do so we ignore spin-dependent forces and take the Hamiltonian of the atom to be

$$H = -\frac{h^2}{2m}(\nabla_1^2 + \nabla_2^2) - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{\lambda e^2}{r_{12}}. \quad (7.1)$$

Here $\mathbf{r}_1$ and $\mathbf{r}_2$ are the coordinates of the electrons relative to the nucleus, $r_{12} = r_{21}$ is the distance between electrons, and λ is a parameter which measures the strength of the Coulomb repulsion between the electrons. We shall seek expansions in powers of λ, and evaluate them at $\lambda = 1$.

The equations of the Hartree approximation corresponding to (7.1) are

$$\left(-\frac{h^2}{2m}\nabla_1^2 - \frac{2e^2}{r_1} + \lambda\int\frac{e^2}{r_{12}}|u_2|^2\,d\mathbf{r}_2 - \epsilon_1\right)u_1(\mathbf{r}_1) = 0, \quad (7.2)$$

$$\left(-\frac{h^2}{2m}\nabla_2^2 - \frac{2e^2}{r_2} + \lambda\int\frac{e^2}{r_{21}}|u_1|^2\,d\mathbf{r}_1 - \epsilon_2\right)u_2(\mathbf{r}_2) = 0, \quad (7.3)$$

$$\int|u_k|^2\,d\mathbf{r}_k = 1, \quad k = 1, 2, \quad (7.4)$$

$$\phi_H(\mathbf{r}_1, \mathbf{r}_2) = u_1(\mathbf{r}_1)u_2(\mathbf{r}_2). \quad (7.5)$$

Here u_1 and u_2 are normalized one-particle states with energy eigenvalues ϵ_1 and ϵ_2 respectively, and ϕ_H is the Hartree wavefunction of the atom. By combining (7.1), (7.4), and (7.5), we obtain for the total energy E of the atom

$$E = \int \phi_H^* H \phi_H \, d\mathbf{r}_1 \, d\mathbf{r}_2 = \epsilon_1 + \epsilon_2 - \lambda \int \frac{e^2}{r_{12}} |u_1|^2 \, |u_2|^2 \, d\mathbf{r}_1 \, d\mathbf{r}_2. \quad (7.6)$$

In the Fock approximation, ϕ_H given by (7.5) is replaced by ϕ_F, which is antisymmetric and is given by

$$\phi_F(\mathbf{r}_1, \mathbf{r}_2) = u_1(\mathbf{r}_1)u_2(\mathbf{r}_2) - u_1(\mathbf{r}_2)u_2(\mathbf{r}_1). \quad (7.5')$$

As a consequence (7.2), (7.3), and (7.6) are modified by the addition of certain exchange terms, while (7.4) is unchanged. However for the ground state of helium, the extra terms vanish so that the Fock equations are identical with the Hartree equations in this case, and then ϕ_F is the antisymmetric part of ϕ_H.

We seek solutions of (7.2)–(7.4) of the form

$$u_k(\mathbf{r}_k, \lambda) = u_{k0}(\mathbf{r}_k) + \lambda \dot{u}_k(\mathbf{r}_k) + O(\lambda^2), \quad k = 1, 2, \quad (7.7)$$

$$\epsilon_k(\lambda) = \epsilon_{k0} + \lambda \dot{\epsilon}_k + O(\lambda^2), \quad k = 1, 2. \quad (7.8)$$

From (7.5)–(7.8) we then obtain for the two-particle states and the total energy

$$\phi(\mathbf{r}_1, \mathbf{r}_2, \lambda) = u_{10}(\mathbf{r}_1)u_{20}(\mathbf{r}_2) + \lambda[u_{10}(\mathbf{r}_1)\dot{u}_2(\mathbf{r}_2) + \dot{u}_1(\mathbf{r}_1)u_{20}(\mathbf{r}_2)] + O(\lambda^2), \quad (7.9)$$

$$E(\lambda) = (\epsilon_{10} + \epsilon_{20}) + \lambda\left[\dot{\epsilon}_1 + \dot{\epsilon}_2 - \int \frac{e^2}{r_{12}} |u_{10}|^2 \, |u_{20}|^2 \, d\mathbf{r}_1 \, d\mathbf{r}_2\right] + O(\lambda^2). \quad (7.10)$$

We now proceed to determine the coefficients in these expansions.

B. Zero-Order Terms

Putting $\lambda = 0$ in (7.2)–(7.4) yields

$$[-(h^2/2m)\nabla_k^2 - 2e^2/r_k - \epsilon_{k0}]u_{k0} = 0, \quad k = 1, 2, \quad (7.11)$$

$$\int |u_{k0}|^2 \, d\mathbf{r}_k = 1, \quad k = 1, 2. \quad (7.12)$$

The solutions of (7.11) are

$$u_{10}(\mathbf{r}_1) = A_{lm}\psi_{\eta lm}(\mathbf{r}_1), \quad \epsilon_{10} = E_\eta, \quad \eta = 1, 2, \cdots, \quad (7.13)$$

$$u_{20}(\mathbf{r}_2) = B_{lm}\psi_{\rho lm}(\mathbf{r}_2), \quad \epsilon_{20} = E_\rho, \quad \rho = 1, 2, \cdots. \quad (7.14)$$

Here $\psi_{\rho lm}$ represents a normalized hydrogenic wavefunction for nuclear charge $2e$ and E_ρ is the corresponding ρth Bohr energy level. A_{lm} and B_{lm} are undetermined constants and the summation convention is employed, m ranging from $-l$ to l and l ranging from 0 to $\eta - 1$ for A_{lm} and from 0 to $\rho - 1$ for B_{lm}. Upon using (7.13) and (7.14) in (7.12) and taking account of the orthonormality of the $\psi_{\rho lm}$, we obtain the following conditions on the A_{lm} and B_{lm}:

$$A_{lm}A_{lm}^* = 1, \quad (7.15)$$

$$B_{lm}B_{lm}^* = 1. \quad (7.16)$$

C. First-Order Terms

Differentiating (7.2)–(7.4) once with respect to λ and then setting $\lambda = 0$, we obtain

$$\left(-\frac{h^2}{2m}\nabla_1^2 - \frac{2e^2}{r_1} - E_\eta\right)\dot{u}_1 = \left(\dot{\epsilon}_1 - \int \frac{e^2}{r_{12}} |u_{20}|^2 \, d\mathbf{r}_2\right)u_{10}, \quad (7.17)$$

$$\left(-\frac{h^2}{2m}\nabla_2^2 - \frac{2e^2}{r_2} - E_\rho\right)\dot{u}_2 = \left(\dot{\epsilon}_2 - \int \frac{e^2}{r_{21}} |u_{10}|^2 \, d\mathbf{r}_1\right)u_{20}, \quad (7.18)$$

$$\mathrm{Re}\int u_{k0}^* \dot{u}_k \, d\mathbf{r}_k = 0, \quad k = 1, 2. \quad (7.19)$$

Equations (7.17)–(7.19) for u_1 and u_2 are inhomogeneous forms of (7.11) and therefore have solutions only if appropriate solvability conditions are satisfied. We derive these conditions by first multiplying (7.17) from the left-hand side by $\psi_{\eta pq}^*(\mathbf{r}_1)$ and integrating over all $\mathbf{r}_1$. If we integrate by parts and use (7.11), the left-hand side vanishes. We then employ (7.13), (7.14), and the orthogonality of the $\psi_{\eta lm}$ to obtain

$$\dot{\epsilon}_1 A_{pq} = A_{l'm'}B_{lm}^* B_{st} J^{\eta\rho}_{l'm',pq,st,lm}. \quad (7.20)$$

Similarly, multiplication of (7.18) by $\psi_{\rho pq}^*(\mathbf{r}_2)$ and integration over $\mathbf{r}_2$ yields

$$\dot{\epsilon}_2 B_{pq} = B_{l'm'}A_{lm}^* A_{st} J^{\eta\rho}_{st,lm,l'm',pq}. \quad (7.21)$$

Here we have defined

$$J^{\alpha\beta}_{ij,kl,mn,uv} = \int \frac{e^2}{r_{12}} \psi_{\alpha ij}(\mathbf{r}_1)\psi_{\alpha kl}^*(\mathbf{r}_1)\psi_{\beta mn}(\mathbf{r}_2)\psi_{\beta uv}^*(\mathbf{r}_2) \, d\mathbf{r}_2 \, d\mathbf{r}_2. \quad (7.22)$$

For given η and ρ, (7.15), (7.16), (7.20), and (7.21) are $\eta^2 + \rho^2 + 2$ inhomogeneous nonlinear equations for the $\eta^2 + \rho^2 + 2$ quantities A_{pq}, B_{pq}, $\dot{\epsilon}_1$ and $\dot{\epsilon}_2$. A solution of these equations determines the zero order wavefunctions u_{10} and u_{20} given by (7.13) and (7.14) and also determines the first order corrections $\dot{\epsilon}_1$ and $\dot{\epsilon}_2$ to the zero order eigenvalues ϵ_{10} and ϵ_{20}.

The determination of all these quantities is analogous to the procedure in the corresponding linear problem in which a perturbing potential splits a degenerate state. The difference is that here the equations (7.20) and (7.21) are nonlinear in the A_{lm} and B_{lm}.

Once these equations are solved, we insert the values of ϵ_1, ϵ_2, A_{pq}, and B_{pq} into the right-hand sides of (7.17) and (7.18). Then we solve these equations for $\dot{u}_1$ and $\dot{u}_2$ by series expansion in the ψ_{pqs} with the results

$$\dot{u}_1 = (E_\eta - E_p)^{-1} A_{lm} B^*_{ij} B_{tv} H^{\rho\eta p}_{lm,qs,tv,ij}\, \psi_{pqs}(\mathbf{r}_1), \quad (7.23)$$

$$\dot{u}_2 = (E_\rho - E_p)^{-1} B_{lm} A^*_{ij} A_{tv} H^{\eta\rho p}_{tv,ij,lm,qs}\, \psi_{pqs}(\mathbf{r}_2). \quad (7.24)$$

Here there is no sum over η and ρ, the terms with $p = \eta$ and $p = \rho$ are omitted, and we have defined

$$H^{\alpha\beta\gamma}_{ij,kl,mn,uv} = \int \frac{e^2}{r_{12}}\, \psi_{\alpha ij}(\mathbf{r}_1)\psi^*_{\beta kl}(\mathbf{r}_1)\psi_{\gamma mn}(\mathbf{r}_2)\psi^*_{\gamma uv}(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2. \quad (7.25)$$

D. Results

We have obtained solutions of (7.2)–(7.4) of the form

$$u_1(\mathbf{r}_1, \lambda) = A_{lm}\psi_{\eta lm}(\mathbf{r}_1) + \frac{\lambda}{E_\eta - E_p} A_{lm} B^*_{ij} B_{tv} H^{\rho\eta p}_{lm,qs,tv,ij}\, \psi_{pqs}(\mathbf{r}_1) + O(\lambda^2), \quad (7.26)$$

$$u_2(\mathbf{r}_2, \lambda) = B_{lm}\psi_{\rho lm}(\mathbf{r}_2) + \frac{\lambda}{E_\rho - E_p} B_{lm} A^*_{ij} A_{tv} H^{\eta\rho p}_{tv,ij,lm,qs}\, \psi_{pqs}(\mathbf{r}_2) + O(\lambda^2), \quad (7.27)$$

$$\epsilon_1(\lambda) = E_\eta + \lambda\epsilon_1 + O(\lambda^2), \quad (7.28)$$

$$\epsilon_2(\lambda) = E_\rho + \lambda\epsilon_2 + O(\lambda^2). \quad (7.29)$$

From (7.5), (7.5′), (7.26), and (7.27), the two-particle Hartree wavefunctions are given by

$$\phi_H(\mathbf{r}_1, \mathbf{r}_2, \lambda) = A_{lm}B_{l'm'}\psi_{\eta lm}(\mathbf{r}_1)\psi_{\rho l'm'}(\mathbf{r}_2) + \lambda\Big[\frac{1}{E_\rho - E_p} B_{lm}A^*_{ij}A_{tv}A_{l'm'}H^{\eta\rho p}_{tv,ij,lm,qs} \times \psi_{\eta l'm'}(\mathbf{r}_1)\psi_{pqs}(\mathbf{r}_2) + \frac{1}{E_\eta - E_p} A_{lm}B^*_{ij}B_{tv}B_{l'm'}H^{\rho\eta p}_{lm,qs,tv,ij} \times \psi_{pqs}(\mathbf{r}_1)\psi_{\rho l'm'}(\mathbf{r}_2)\Big] + O(\lambda^2). \quad (7.30)$$

In (7.30) there is no summation on η or ρ. For the ground state the Fock wavefunction is

$$\phi_F(\mathbf{r}_1, \mathbf{r}_2, \lambda) = \phi_H(\mathbf{r}_1, \mathbf{r}_2, \lambda) - \phi_H(\mathbf{r}_2, \mathbf{r}_1, \lambda). \quad (7.30')$$

From (7.6), (7.26), and (7.27) the total energy is

$$E(\lambda) = (E_\eta + E_\rho) + \lambda(\epsilon_1 + \epsilon_2 - A^*_{lm}A_{pq}B^*_{l'm'}B_{p'q'}J^{\eta p}_{pq,lm,p'q',l'm'}) + O(\lambda^2). \quad (7.31)$$

E. Examples

For the normalized ground state we have

$$u_{10} = \psi_{100}(\mathbf{r}_1), \quad u_{20} = \psi_{100}(\mathbf{r}_2), \quad (7.32)$$

$$\epsilon_{10} = \epsilon_{20} = E_1. \quad (7.33)$$

Thus

$$A_{ij} = B_{ij} = \delta_{i0}\delta_{j0}. \quad (7.34)$$

Then (20) and (21) yield

$$\epsilon_1 = \epsilon_2 = J^{11}_{00,00,00,00} = 5me^4/8h^2. \quad (7.35)$$

Inserting (7.33)–(7.35) into (7.30) and (7.31), we obtain

$$\phi_H(\mathbf{r}_1, \mathbf{r}_2, \lambda) = \psi_{100}(\mathbf{r}_1)\psi_{100}(\mathbf{r}_2) + \lambda(E_1 - E_p)^{-1}H^{11p}_{00,00,00,qs} \times [\psi_{100}(\mathbf{r}_1)\psi_{pqs}(\mathbf{r}_2) + \psi_{100}(\mathbf{r}_2)\psi_{pqs}(\mathbf{r}_1)] + O(\lambda^2), \quad (7.36)$$

$$E(\lambda) = 2E_1 + (5me^4/8h^2)\lambda + O(\lambda^2). \quad (7.37)$$

The Fock wavefunction is given by (7.30′) and (7.36).

In the first excited states we have

$$u_{10} = \psi_{100}(\mathbf{r}_1), \quad u_{20} = B_{lm}\psi_{2lm}(\mathbf{r}_2), \quad (7.38)$$

$$\epsilon_{10} = E_1, \quad \epsilon_{20} = E_2. \quad (7.39)$$

Thus $A_{ij} = \delta_{i0}\delta_{j0}$, so (7.20) and (7.21) become

$$\epsilon_1 = B^*_{lm}B_{st}J^{12}_{00,00,st,lm}, \quad (7.40)$$

$$\epsilon_2 B_{pq} = B_{lm}J^{12}_{00,00,lm,pq}. \quad (7.41)$$

We see that (7.41) is a linear eigenvalue problem for B_{pq} and the eigenvalue ϵ_2. The magnitude of B_{pq} is determined by (7.16) and then (7.40) yields ϵ_1.

The necessary and sufficient condition that a solution of (7.41) exist is

$$\begin{vmatrix} J_{11,11} - \epsilon_2 & J_{10,11} & J_{1,-1,11} & J_{00,11} \\ J_{11,10} & J_{10,10} - \epsilon_2 & J_{1,-1,10} & J_{00,10} \\ J_{11,1-1} & J_{10,1-1} & J_{1-1,1-1} - \epsilon_2 & J_{00,1,-1} \\ J_{11,00} & J_{10,00} & J_{1,-1,00} & J_{00,00} - \epsilon_2 \end{vmatrix} = 0. \quad (7.42)$$

For brevity we have suppressed the first four subscripts and the two superscripts on the J's here. Each of the four solutions ϵ_2 of (7.42) gives rise to a set of B_{pq}'s via (7.41) and to a value of ϵ_1 via (7.40). We see from (7.40) that the first-order energy of electron 1 depends upon the choice of u_{20}, the zero-order state of electron 2.

From (7.30), (7.31), (7.38), and (7.39), the wavefunctions and energy for the first excited states are

$$\phi_H(\mathbf{r}_1, \mathbf{r}_2, \lambda) = B_{lm}\psi_{100}(\mathbf{r}_1)\psi_{2lm}(\mathbf{r}_2) + \lambda[(E_2 - E_p)^{-1}B_{lm}H^{12p}_{00,00,lm,qs}\,\psi_{100}(\mathbf{r}_1)\psi_{pqs}(\mathbf{r}_2) + (E_1 - E_p)^{-1}B^*_{ij}B_{tv}B_{lm}H^{21p}_{00,qs,tv,ij} \times \psi_{pqs}(\mathbf{r}_1)\psi_{2lm}(\mathbf{r}_2)] + O(\lambda^2). \quad (7.43)$$

$$E(\lambda) = E_1 + E_2 + \lambda\epsilon_2 + O(\lambda^2). \quad (7.44)$$

In obtaining (7.44) from (7.31) we noticed that the last term in the parenthesis in (7.31) cancels ϵ_1.

We note that one of the nonlinear systems (7.20) or (7.21) reduces to a linear system, as in the above example, whenever one electron is in the one-particle ground state ψ_{100}, and only then. Therefore we can deal with any state of this type as we did in the example.

F. Comparison with Perturbation Solution of the Schrödinger Equation

We first present briefly the standard perturbation results for the Schrödinger equation for helium in our notation. The two-particle state and total energy are governed by

$$[H - E(\lambda)]\phi(\mathbf{r}_1, \mathbf{r}_2, \lambda) = 0, \quad (7.45)$$

$$\int |\phi(\mathbf{r}_1, \mathbf{r}_2, \lambda)|^2\, d\mathbf{r}_1\, d\mathbf{r}_2 = 1. \quad (7.46)$$

Here H is given by (7.1). We write the solution in the form

$$\phi(\mathbf{r}_1, \mathbf{r}_2, \lambda) = \phi_0(\mathbf{r}_1, \mathbf{r}_2) + \lambda\dot{\phi}(\mathbf{r}_1, \mathbf{r}_2) + O(\lambda^2), \quad (7.47)$$

$$E(\lambda) = E(0) + \lambda\dot{E} + O(\lambda^2). \quad (7.48)$$

Putting $\lambda = 0$ in (7.45) and (7.46) yields

$$\left[-\frac{h^2}{2m}(\nabla_1^2 + \nabla_2^2) - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} - E(0)\right]\phi_0(\mathbf{r}_1, \mathbf{r}_2) = 0. \quad (7.49)$$

The solution of (7.49) is

$$\phi_0(\mathbf{r}_1, \mathbf{r}_2) = C_{lm,l'm'}\psi_{\eta lm}(\mathbf{r}_1)\psi_{\rho l'm'}(\mathbf{r}_2), \quad (7.50)$$

$$E(0) = E_\eta + E_\rho, \quad \eta, \rho = 1, 2, \cdots. \quad (7.51)$$

The $C_{lm,l'm'}$ are undetermined constants. By setting $\lambda = 0$ in (7.46) and using (7.50), we find

$$C_{lm,l'm'}C^*_{lm,l'm'} = 1. \quad (7.52)$$

We next differentiate (7.45) and (7.46) once with respect to λ, and then set $\lambda = 0$ to obtain

$$\left[-\frac{h^2}{2m}(\nabla_1^2 + \nabla_2^2) - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} - (E_\eta + E_\rho)\right]\dot{\phi} = \left(\dot{E} - \frac{e^2}{r_{12}}\right)\phi_0, \quad (7.53)$$

$$\mathrm{Re}\int \phi_0^*\dot{\phi}\, d\mathbf{r}_1\, d\mathbf{r}_2 = 0. \quad (7.54)$$

Equation (7.53) is an inhomogeneous form of (7.49) and therefore has a solution only if an appropriate solvability condition is satisfied. We derive this condition by multiplying (7.53) by $\psi^*_{\eta pq}(\mathbf{r}_1)\psi^*_{\rho st}(\mathbf{r}_2)$ from the left and integrating over all $\mathbf{r}_1$, $\mathbf{r}_2$. Integrating by parts and using (7.49) and the orthogonality of the $\psi_{\rho st}$, we obtain the system

$$\dot{E}C_{pq,st} = J^{\eta\rho}_{lm,pq,l'm',st}C_{lm,l'm'}. \quad (7.55)$$

Equations (7.55) and (7.52) are $\eta^2 + \rho^2 + 1$ equations for the $\eta^2 + \rho^2 + 1$ quantities $\dot{E}$ and $C_{pq,st}$. When they are satisfied, (7.53) can be solved for $\dot{\phi}$ with the result

$$\dot{\phi}(\mathbf{r}_1, \mathbf{r}_2) = \frac{C_{lm,l'm'}K^{\rho u\eta r}_{lm,st,l'm',vw}}{(E_r + E_u) - (E_\eta + E_\rho)}\psi_{rst}(\mathbf{r}_1)\psi_{uvw}(\mathbf{r}_2). \quad (7.56)$$

Here η, ρ are not summed over and the terms for which $r = \eta$ and/or $u = \rho$ are omitted. We have defined

$$K^{\alpha\beta\gamma\delta}_{ij,kl,mn,qr} = \int \frac{e^2}{r_{12}}\psi_{\alpha ij}(\mathbf{r}_1)\psi^*_{\beta kl}(\mathbf{r}_1)\psi_{\gamma mn}(\mathbf{r}_2)\psi^*_{\delta qr}(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2. \quad (7.57)$$

We shall now compare the results of this section with those of Sec. 7E. For the ground state, (7.50) and (7.51) are

$$\phi_0(\mathbf{r}_1, \mathbf{r}_2) = \psi_{100}(\mathbf{r}_1)\psi_{100}(\mathbf{r}_2), \quad (7.58)$$

$$E(0) = 2E_1. \quad (7.59)$$

Thus $C_{lm,l'm'} = \delta_{l0}\delta_{l'0}$, so (7.55) and (7.56) become

$$\dot{E} = J^{11}_{00,00,00,00} = \frac{5me^4}{8h^2}, \quad (7.60)$$

$$\dot{\phi}(\mathbf{r}_1, \mathbf{r}_2) = \frac{K^{1u1r}_{00,st,00,vw}}{E_r + E_u - 2E_1}\psi_{rst}(\mathbf{r}_1)\psi_{uvw}(\mathbf{r}_2). \quad (7.61)$$

A comparison of (7.58)–(7.61) with (7.36) and (7.37) shows that the energies agree to first order but the wavefunctions agree only to zero order.

For the first excited state there are four degenerate zero-order wavefunctions given by $\psi_{100}(\mathbf{r}_1)$ multiplied by $\psi_{211}(\mathbf{r}_2)$, $\psi_{210}(\mathbf{r}_2)$, $\psi_{21,-1}(\mathbf{r}_2)$, or $\psi_{200}(\mathbf{r}_2)$. To make a comparison with the Hartree theory, we do not antisymmetrize these or count the four exchange degenerate states obtained from them as new states. We first obtain from (7.51)

$$E(0) = E_1 + E_2. \quad (7.62)$$

This is the same as $E(0)$ of the Hartree method, which

is given by (7.44) with $\lambda = 0$. To compare further terms in $E(\lambda)$ we note that the only nonvanishing $C_{lm,l'm'}$ are those of the form $C_{00,l'm'}$. Thus Eqs. (7.55) for $\dot{E}$ and $C_{00,l'm'}$ become identical with Eqs. (7.41), which are satisfied by ϵ_2 and $B_{l',m'}$. Hence the values of $\dot{E}$ obtained from (7.55) are the same as the values of ϵ_2 obtained from (7.42) and the corresponding $C_{00,l'm'}$ are the same as the $B_{l'm'}$. Therefore the zero-order wavefunction (7.50) is the same as that of the Hartree method, which is given by (7.43) with $\lambda = 0$. The energy $E(\lambda)$ given by (7.48) agrees with the result (7.44) of the Hartree method through terms of order λ since $\epsilon_2 = \dot{E}$. However, the term of order λ in the wavefunction, given by (7.56), clearly disagrees with the term of order λ in the Hartree wavefunction (7.43). Therefore the energy given by the Hartree method will disagree with that of the Schrödinger equation in the term of order λ^2.

For any two-particle state with one electron in the one-particle ground state, the Hartree wavefunctions will agree with the exact wavefunction only to terms of order zero in λ. The Hartree energy will agree with the exact energy through terms of first order in λ. However, for two-particle states with no electron in the one-particle ground state, the Hartree energy will agree with the exact energy only to zero order in λ while the Hartree wavefunction will not agree with the exact wavefunction even for $\lambda = 0$. This is because the Eqs. (7.20) and (7.21), which determine ϵ_1, ϵ_2, A_{pq}, and B_{pq} will be nonlinear while (7.55), which determines $\dot{E}$ and $C_{pq,st}$, is linear. Only by accident could the solutions of these two sets of equations agree.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Experimental Confirmation of the Applicability of the Fokker–Planck Equation to a Nonlinear Oscillator*

J. B. MORTON†‡ AND S. CORRSIN
Department of Mechanics, The Johns Hopkins University, Baltimore, Maryland

(Received 20 August 1968)

A traditional derivation of the Fokker–Planck equation is examined with emphasis on assumptions, especially sufficient inequalities between excitation and response characteristic times in a "real" system, i.e., a system with finite correlation time of the excitation. Comparison with experiment for an electronic "cubic spring" oscillator gives good agreement in both second and fourth moments of the response.

I. INTRODUCTION

The Fokker–Planck equation, expressing the balance of probability density, has been used extensively in the theoretical study of both linear and nonlinear systems driven by random excitation. The major success has been in the linear problem of Brownian motion.[1] In this problem there has also been a successful connection established with experiment. In spite of the considerable use of the Fokker–Planck method for nonlinear systems, it is difficult to find published experimental confirmations of its success.[2]

The two purposes of this paper are (a) to indicate more explicitly than is customary how a realistic system can approximate the classical requirements for Fokker–Planck behavior; and (b) to find out experimentally whether a noise-driven nonlinear oscillator has low-order moments consistent with those calculated from its Fokker–Planck equation. The emphasis of the first part of this paper is on the detailed assumptions needed to connect a real system with a Fokker–Planck equation. The steady-state solution of the Fokker–Planck equation for the linear and cubic spring oscillators is then compared with the results of an analog computer experiment.

II. ASSUMPTIONS NEEDED TO DERIVE THE FOKKER–PLANCK EQUATION

We follow the procedure of Wang and Uhlenbeck[3] briefly, and then review the assumptions in order to

* Work supported by U.S. Office of Naval Research, Fluid Dynamics Branch.

† Present address: Department of Aerospace Engineering and Engineering Physics, University of Virginia, Charlottesville, Virginia.

‡ National Aeronautics and Space Administration trainee.

[1] For a recent mathematical critical review, see E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton University Press, Princeton, N.J., 1967).

[2] H. A. Kramers, Physica 7, 284 (1940); R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon & Breach Science Publishers, New York, 1963; trans. by R. A. Silverman); T. K. Caughey, J. Acoust. Soc. Am. 35, 1683 (1963).

[3] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, 323 (1945) [reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed. (Dover Publications, Inc., New York, 1954)].

is given by (7.44) with $\lambda = 0$. To compare further terms in $E(\lambda)$ we note that the only nonvanishing $C_{lm,l'm'}$ are those of the form $C_{00,l'm'}$. Thus Eqs. (7.55) for $\dot{E}$ and $C_{00,l'm'}$ become identical with Eqs. (7.41), which are satisfied by ϵ_2 and $B_{l',m'}$. Hence the values of $\dot{E}$ obtained from (7.55) are the same as the values of ϵ_2 obtained from (7.42) and the corresponding $C_{00,l'm'}$ are the same as the $B_{l'm'}$. Therefore the zero-order wavefunction (7.50) is the same as that of the Hartree method, which is given by (7.43) with $\lambda = 0$. The energy $E(\lambda)$ given by (7.48) agrees with the result (7.44) of the Hartree method through terms of order λ since $\epsilon_2 = \dot{E}$. However, the term of order λ in the wavefunction, given by (7.56), clearly disagrees with the term of order λ in the Hartree wavefunction (7.43). Therefore the energy given by the Hartree method will disagree with that of the Schrödinger equation in the term of order λ^2.

For any two-particle state with one electron in the one-particle ground state, the Hartree wavefunctions will agree with the exact wavefunction only to terms of order zero in λ. The Hartree energy will agree with the exact energy through terms of first order in λ. However, for two-particle states with no electron in the one-particle ground state, the Hartree energy will agree with the exact energy only to zero order in λ while the Hartree wavefunction will not agree with the exact wavefunction even for $\lambda = 0$. This is because the Eqs. (7.20) and (7.21), which determine ϵ_1, ϵ_2, A_{pq}, and B_{pq} will be nonlinear while (7.55), which determines $\dot{E}$ and $C_{pq,st}$, is linear. Only by accident could the solutions of these two sets of equations agree.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Experimental Confirmation of the Applicability of the Fokker–Planck Equation to a Nonlinear Oscillator*

J. B. MORTON†‡ AND S. CORRSIN
Department of Mechanics, The Johns Hopkins University, Baltimore, Maryland

(Received 20 August 1968)

A traditional derivation of the Fokker–Planck equation is examined with emphasis on assumptions, especially sufficient inequalities between excitation and response characteristic times in a "real" system, i.e., a system with finite correlation time of the excitation. Comparison with experiment for an electronic "cubic spring" oscillator gives good agreement in both second and fourth moments of the response.

I. INTRODUCTION

The Fokker–Planck equation, expressing the balance of probability density, has been used extensively in the theoretical study of both linear and nonlinear systems driven by random excitation. The major success has been in the linear problem of Brownian motion.[1] In this problem there has also been a successful connection established with experiment. In spite of the considerable use of the Fokker–Planck method for nonlinear systems, it is difficult to find published experimental confirmations of its success.[2]

The two purposes of this paper are (a) to indicate more explicitly than is customary how a realistic system can approximate the classical requirements for Fokker–Planck behavior; and (b) to find out experimentally whether a noise-driven nonlinear oscillator has low-order moments consistent with those calculated from its Fokker–Planck equation. The emphasis of the first part of this paper is on the detailed assumptions needed to connect a real system with a Fokker–Planck equation. The steady-state solution of the Fokker–Planck equation for the linear and cubic spring oscillators is then compared with the results of an analog computer experiment.

II. ASSUMPTIONS NEEDED TO DERIVE THE FOKKER–PLANCK EQUATION

We follow the procedure of Wang and Uhlenbeck[3] briefly, and then review the assumptions in order to

* Work supported by U.S. Office of Naval Research, Fluid Dynamics Branch.

† Present address: Department of Aerospace Engineering and Engineering Physics, University of Virginia, Charlottesville, Virginia.

‡ National Aeronautics and Space Administration trainee.

[1] For a recent mathematical critical review, see E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton University Press, Princeton, N.J., 1967).

[2] H. A. Kramers, Physica 7, 284 (1940); R. L. Stratonovich, *Topics in the Theory of Random Noise* (Gordon & Breach Science Publishers, New York, 1963; trans. by R. A. Silverman); T. K. Caughey, J. Acoust. Soc. Am. 35, 1683 (1963).

[3] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, 323 (1945) [reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Ed. (Dover Publications, Inc., New York, 1954)].

establish realistic, sufficient conditions that they be well approximated by a physical system.

The principal object of study is a system (usually identified by a differential equation) driven by a random forcing function. We are given any required statistical properties of the forcing function and would like to calculate the statistical properties of the system response. In the Fokker–Planck approach, the particular response property sought is the conditional probability density function $P(x \mid y, t)$, defined by the statement that if the system is at x at time zero, $P(x \mid y, t)\, dy$ is the probability that it will be between y and $y + dy$ at time t.

The first assumption on the system and excitation is that the conditional probability of the response satisfies the Smoluchowski–Chapman–Kolmogorov equation:

$$P(x \mid y, t) = \int_{-\infty}^{\infty} dz\, P(x \mid z, t) P(z \mid y, \Delta t). \tag{1}$$

Wang and Uhlenbeck imply that restriction to a Markov process is necessary [sentence following their Eq. (37)] and sufficient [sentence following their Eq. (6)]. However, Parzen,[4] for example, points out that the Markovian restriction is sufficient, but not necessary.

The second, third, and fourth assumptions basic to the derivation deal with the *conditional moments*

$$a^{(m)}(z, \Delta t) = \int_{-\infty}^{\infty} dy\, (y - z)^m P(z \mid y, \Delta t). \tag{2}$$

We assume that these moments exist (i.e., are not infinite), and that $a^{(1)}$ and $a^{(2)}$ are proportional to Δt plus terms with higher powers of Δt, while all higher moments are proportional to higher powers of Δt. Thus,

$$A(z) = \lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} a^{(1)}(z, \Delta t) \right\} \quad \text{is finite,} \tag{3}$$

$$B(z) = \lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} a^{(2)}(z, \Delta t) \right\} \quad \text{is finite,} \tag{4}$$

and, for $m > 2$,

$$\lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} a^{(m)}(z, \Delta t) \right\} = 0. \tag{5}$$

Pawula[5] has shown that if

$$\lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} a^{(m)} \right\}$$

is finite for any $m > 2$, then it must be finite for all m.

[4] E. Parzen, *Stochastic Processes* (Holden-Day, Inc., San Francisco, 1962), p. 203.

[5] R. F. Pawula, IEEE Tran. Inform. Theory **13**, 33 (1967).

With these assumptions, it is easy to show[3] that $P(x \mid y, t)$ satisfies

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial y}(AP) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(BP), \tag{6}$$

which is the Fokker–Planck equation.

For an n-dimensional process, this has been generalized to

$$\frac{\partial P}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial y_i}(A_i P) + \tfrac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2}{\partial y_j \partial y_k}(B_{jk} P), \tag{7}$$

where the A_i and the B_{jk} are defined in a manner analogous to (3) and (4).

To relate this to a particular "real" system, we must show that condition (1) holds and that conditions analogous to the limiting conditions (3), (4), and (5) are satisfied.

III. APPROXIMATE CONDITIONS FOR A "REAL" SYSTEM

Consider a system characterized by the following equation:

$$\ddot{x}(t) + \alpha \dot{x}(t) + \psi_x(t) = f(t), \tag{8}$$

where α is a constant, $\psi_x(t)$ is an arbitrary "smooth" function of $x(t)$, and f is a random forcing function. For ψ_x a linear function, Doob[6] has proven that $(x, \dot{x})$ forms a second-order Markov process, provided $f(t)$ is Gaussian white noise, hence Dirac-correlated. For the nonlinear case, no such theorem exists, although it seems reasonable to expect that for Gaussian white noise, $(x, \dot{x})$ is still Markovian.

We write

$$\dot{x} \equiv s, \tag{9}$$

so that (8) is

$$\dot{s} + \alpha s + \psi_x = f(t). \tag{10}$$

This replacement of (8) by two first-order equations is convenient for establishing conditions appropriate for relations like (3), (4), and (5). The strategy is to integrate (9) and (10) over a time Δt which will later be assumed short compared with the *shortest* characteristic times in $x(t)$ and $s(t)$, yet very long compared with the *longest* characteristic time in $f(t)$. The existence of such a doubly asymptotic time is crucial to the standard Fokker–Planck formulation. For example, the integration of (9) is

$$\int_t^{t+\Delta t} \dot{x}(t_1)\, dt_1 = \int_t^{t+\Delta t} s(t_1)\, dt_1.$$

[6] J. L. Doob, *Stochastic Processes* (John Wiley & Sons, Inc., New York, 1953).

Expanding $s(t_1)$ in Taylor series,

$$x(t+\Delta t) - x(t) = \int_t^{t+\Delta t} \left[s(t) + \dot{s}(t)(t_1 - t) + \ddot{s}(t)\frac{(t_1-t)^2}{2} + \cdots \right] dt_1,$$

$$\therefore \quad x(t+\Delta t) - x(t) = s(t)\Delta t + \dot{s}(t)\frac{(\Delta t)^2}{2} + \cdots. \tag{11}$$

The corresponding integration of (10) is[7]

$$\begin{aligned} s(t+\Delta t) - s(t) = &-[\alpha s(t) + \psi_x(t)]\Delta t \\ &- [\alpha \dot{s}(t) + \psi_x'(t)s(t)](\Delta t)^2/2 \\ &+ \cdots + \int_t^{t+\Delta t} f(t_1)\, dt_1. \end{aligned} \tag{12}$$

We next evaluate the conditional moments $a_x^{(1)}$, $a_s^{(1)}$, $a_{xx}^{(2)}$, $a_{ss}^{(2)}$, $a_{xs}^{(2)}$:

$$a_x^{(1)}(z, k, \Delta t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - z) P_{xs}(z, k \mid y, j, \Delta t)\, dy\, dj, \tag{13}$$

$$a_s^{(1)}(z, k, \Delta t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (j - k) P_{xs}(z, k \mid y, j, \Delta t)\, dy\, dj. \tag{14}$$

These are direct generalizations of (2).

$$P_{xs}(d, b \mid c, a, t)\, dc\, da$$

is the probability that, if

$$\begin{Bmatrix} x = d \\ s = b \end{Bmatrix} \text{ at } t = 0, \text{ then } \begin{Bmatrix} c \le x \le c + dc \\ a \le s \le a + da \end{Bmatrix} \text{ at } t = t.$$

Thus (13) is the conditional average of $x(t + \Delta t) - x(t)$ and (14) is that of $\dot{x}(t + \Delta t) - \dot{x}(t)$. From (11) and (12), then, we see that

$$a_x^{(1)} = s(t)\Delta t + \langle \dot{s}(t) \rangle_c (\Delta t)^2/2 + \cdots, \tag{15}$$

$$\begin{aligned} a_s^{(1)} = &-[\alpha s(t) + \psi_x(t)]\Delta t \\ &- [\alpha \langle \dot{s}(t) \rangle_c + \psi_x'(t)s(t)](\Delta t)^2/2 \\ &+ \cdots + \left\langle \int_t^{t+\Delta t} f(t_1)\, dt_1 \right\rangle_c. \end{aligned} \tag{16}$$

Since $\langle\ \rangle_c$ is the average *conditional* on the occurrence of the actual values of $x(t)$ and $s(t)$ in configuration ("physical") space,[8]

$$\langle x(t) \rangle_c = x(t), \quad \langle s(t) \rangle_c = s(t). \tag{17}$$

In general, however,

$$\langle \dot{s}(t) \rangle_c \neq \dot{s}(t). \tag{18}$$

In a stationary process, $\dot{s}(t)$ cannot be statistically independent of $x(t)$ and $s(t)$, since

$$\langle x\dot{s} \rangle = -\langle \dot{x}^2 \rangle \neq 0 \tag{19}$$

in general. In a nonstationary process, s and $\dot{s}$ may be correlated[9]:

$$\langle s\dot{s} \rangle = \frac{1}{2}\left\langle \frac{d}{dt}(s^2) \right\rangle = \frac{1}{2}\frac{d}{dt}\langle s^2 \rangle \neq 0, \tag{20}$$

in which case they cannot be statistically independent.

In analogy to (3) we write[10]

$$A_x^{(1)} = \left[\frac{1}{\Delta t} a_x^{(1)} \right]_{\Delta t \to \theta}, \tag{21}$$

where θ is a value of Δt which is very small compared with the smallest characteristic time in the configuration-space response function $\dot{x}(t) \equiv s(t)$. To obtain the desired 1-term approximation to (17), no direct reference to the characteristic times in $x(t)$ itself need be made. Since $\dot{x}$ can only have a richer high-frequency spectrum than x, θ will be much smaller than the smallest characteristic time in x, too.

By the mean-value theorem it follows that

$$a_x^{(1)} - s(t)\Delta t \le \langle \dot{s}(t^*) \rangle_c (\Delta t)^2/2,$$

where $t \le t^* \le t + \Delta t$. Operationally then, θ must be so small that

$$\theta\, |\langle \dot{s} \rangle_c| \ll 2\, |s(t)|, \tag{22}$$

permitting us to neglect all terms but the first in (15). Note that (22) includes properties of the velocity s and acceleration $\dot{s}$.

If $x(t)$ has a *narrow-band spectrum* at frequency ω_0 rad/sec, then s and $\dot{s}$ also have narrow-band spectra in that neighborhood, and their smallest characteristic times will all be

$$\tau_{\min} \approx 1/\omega_0. \tag{23}$$

Furthermore, if x is of order x_1 (say), $|s|$ is of order $\omega_0 x_1$ and $|\dot{s}|$ is of order $\omega_0^2 x_1$. $|\langle \dot{s} \rangle_c|$ is, presumably, considerably smaller than $|\dot{s}|$, so a conservative condition sufficient to fulfill (22) is

$$\theta \ll 2/\omega_0. \tag{24}$$

[7] Prime indicates differentiation with respect to x.

[8] Although $x(t)$ and $s(t)$ here are written in configuration-space symbols, these are actually *values* in *probability space*. To keep the number of symbols to a minimum, we use the same symbol for the random variable and its values, where there is no chance for confusion.

[9] We can imagine a nonstationary process, $\langle x^2 \rangle \neq$ const, in which $s \equiv \dot{x}$ is stationary, especially if $f(t)$ is stationary. On the other hand, if we make $f(t)$ nonstationary, it seems likely that s and $\dot{s}$ are inevitably nonstationary too.

[10] θ cannot equal zero because of the need to preserve the driving-force integral (39) in $A_{ss}^{(2)}$. For a system well suited to Fokker-Planck approximation, the results are insensitive to the exact value of θ within a viable range.

If, on the other hand, $x(t)$ has a *broad spectrum function* which decays at only a moderate rate as $\omega \to \infty$, $|\dot{s}| \equiv |\ddot{x}|$ may take on very large values, putting a tighter bound on θ.

It should also be noted that many systems of interest have

$$\langle \dot{s} \rangle_c = 0,$$

in which case we need worry only about the next term [proportional to $(\Delta t)^3$] in series (15).

With (22) fulfilled,

$$A_x^{(1)} = s, \tag{25}$$

which is the Fokker–Planck form.

A similar examination of (16) requires consideration of the last term

$$\left\langle \int_t^{t+\Delta t} f(t_1)\, dt_1 \right\rangle_c. \tag{26}$$

Since the driving force is assumed to be unaffected by the response, the average is really unconditional. Furthermore, since these are ensemble averages, we can interchange the sequence of averaging and integration:

$$\left\langle \int_t^{t+\Delta t} f\, dt_1 \right\rangle = \int_t^{t+\Delta t} \langle f \rangle\, dt_1 = 0, \tag{27}$$

if we have made the usual choice of $f(t)$ as a function with zero average value.

Then (16) becomes

$$a_s^{(1)} = -[\alpha s(t) + \psi_x(t)]\Delta t - [\alpha \langle \dot{s}(t) \rangle_c + \psi_x'(t) s(t)](\Delta t)^2/2 + \cdots. \tag{28}$$

To obtain the Fokker–Planck result, θ must be small enough to permit approximating (28) by its Δt term. We need require, further, only that

$$\theta\, |\alpha \langle \dot{s} \rangle_c + s\psi_x'| \ll 2\, |\alpha s + \psi_x|. \tag{29}$$

Then,

$$a_s^{(1)} \doteq -[\alpha s(t) + \psi_x(t)]\Delta t \tag{30}$$

and

$$A_s^{(1)} \doteq -[\alpha s + \psi_x]. \tag{31}$$

Next we turn to the second moments, $a_{xx}^{(2)}$, $a_{xs}^{(2)}$, and $a_{ss}^{(2)}$:

$$a_{xx}^{(2)} = s^2(t)(\Delta t)^2 + \cdots, \tag{32}$$

$$a_{xs}^{(2)} = -[\alpha s^2 + \psi_x s](\Delta t)^2 + \cdots + \left\langle s(t) \int_t^{t+\Delta t} f\, dt_1 \right\rangle_c \Delta t + \left\langle \dot{s}(t) \int_t^{t+\Delta t} f\, dt_1 \right\rangle_c \frac{(\Delta t)^2}{2} + \cdots, \tag{33}$$

$$a_{ss}^{(2)} = [s + \psi_x]^2(\Delta t)^2 + \cdots + 2\left\langle [\alpha s + \psi_x] \int_t^{t+\Delta t} f\, dt_1 \right\rangle_c \Delta t + \cdots + \left\langle \left(\int_t^{t+\Delta t} f\, dt_1 \right)^2 \right\rangle_c. \tag{34}$$

In arriving at these, we have again replaced conditional averages (of variables used to specify these conditions) by the values of the variables themselves:

$$\langle s^2(t) \rangle_c = s^2(t), \quad \langle x^3(t) s(t) \rangle_c = x^3(t) s(t), \text{ etc.} \tag{35}$$

The Fokker–Planck theory requires that from these three expressions, (32), (33), and (34), only the final term in (34) be kept. For brevity, we focus primarily on this term which we call Λ. Writing the squared integral as a double integral, and interchanging the sequence of averaging and integration, we find

$$\Lambda = \int_t^{t+\Delta t} \int_t^{t+\Delta t} \langle f(t_1) f(t_2) \rangle_c\, dt_1\, dt_2. \tag{36}$$

Since we have assumed that *the statistical properties of $f(t)$ are independent of the system response*, the conditional averages can be replaced by unconditional ones.

Then we restrict to statistically stationary $f(t)$, and introduce the autocorrelation function

$$R_f(\tau) \equiv \langle f(t) f(t+\tau) \rangle / \langle f^2 \rangle. \tag{37}$$

Equation (36) can be written as

$$\Lambda(\Delta t) = \langle f^2 \rangle \int_0^{\Delta t} \int_0^{\Delta t} R_f(t_1 - t_2)\, dt_1\, dt_2, \tag{38}$$

a form possibly due to Taylor[11] in his "theory of diffusion by continuous movements." Kampé de Fériet[12] integrated by parts to get the alternative form

$$\Lambda(\Delta t) = 2\langle f^2 \rangle \Delta t \int_0^{\Delta t} \left[1 - \frac{\tau}{\Delta t}\right] R_f(\tau)\, d\tau. \tag{39}$$

We require that there exist a value θ of Δt which is

(a) sufficiently small that each of the $(\Delta t)^n$ terms in (32), (33), and (34) be negligible compared with $\Lambda(\theta)$ for $n \geq 2$, and

(b) sufficiently large that

$$\Lambda(\theta)/\theta \doteq \lim_{\Delta t \to \infty} [\Lambda(\Delta t)/\Delta t], \tag{40}$$

i.e., θ must be much larger than the largest characteristic time in $f(t)$.

If $f(t)$ possesses an "integral scale"[11]

$$T_f \equiv \int_0^{\infty} R_f(\tau)\, d\tau, \tag{41}$$

which is neither 0 nor ∞, then T_f is a plausible candidate for "largest characteristic time." On the other hand, since $T_f = 0$ signals passed through ac coupled

[11] G. I. Taylor, Proc. London Math. Soc. **A20**, 196 (1921). Taylor's problem is a smooth random walk, equivalent to a system equation $\dot{x} = f(t)$ instead of (8).

[12] J. Kampé de Fériet, Ann. Soc. Sci. Bruxelles **59**, 145 (1939).

circuits, we may need to invoke an alternative such as

$$\int_0^\infty |R_f|\, d\tau, \tag{42}$$

if it exists, or perhaps the value of τ beyond which $R_f(\tau) = 0$ if such an interval exists.

We note in passing that for Δt (or θ) $\to 0$, Eq. (39) indicates that

$$\Lambda \to \langle f^2 \rangle (\Delta t)^2. \tag{43}$$

The final terms to be discussed in the second moments are the integral terms in (33). If the forcing function is statistically independent of the response, these are zero. With the conditions outlined above,

$$a_{xx}^{(2)} \doteq 0, \qquad A_{xx}^{(2)} \doteq 0, \tag{44}$$

$$a_{xs}^{(2)} \doteq 0, \qquad A_{xs}^{(2)} \doteq 0, \tag{45}$$

$$a_{ss}^{(2)} \doteq D\Delta t, \quad A_{ss}^{(2)} \doteq D, \tag{46}$$

where $D \equiv \Lambda(\theta)\theta^{-1}$, and the appropriate Fokker-Planck equation is

$$\frac{\partial P_{xs}(u, v)}{\partial t} = -\frac{\partial}{\partial u}(vP_{xs}) + \frac{\partial}{\partial v}[\{\alpha v + \psi_x(u)\}P_{xs}] + \frac{D}{2}\frac{\partial^2 P_{xs}}{\partial v^2}. \tag{47}$$

The foregoing discussion has been aimed explicitly at identifying conditions which permit the moments, $a_x^{(1)}$, $a_{xx}^{(2)}$, etc., to be replaced by their Fokker-Planck approximations. At what step can we justify the applicability of the Smoluchowski-Chapman-Kolmogorov limitation, the two-variable version of (1)?

The linear approximations to the moments $a_x^{(1)}$ and $a_s^{(1)}$ are consistent with instantaneous linear approximations in configuration space:

$$x(t + \Delta t) \doteq x(t) + s(t)\Delta t \tag{48}$$

and

$$s(t + \Delta t) \doteq s(t) - [\alpha s(t) + \psi_x(t)]\Delta t + \int_t^{t+\Delta t} f\, dt_1. \tag{49}$$

Within these approximations, $x(t + \Delta t)$ and $s(t + \Delta t)$ depend on only $x(t)$ and $s(t)$ [and, of course, $f(t_1)$ in the interval], and not on x and s at prior times, nor on higher derivatives at time t. It seems reasonable to assume that this is sufficient to give Markovian behavior, thus to make applicable the two-variable Smoluchowski-Chapman-Kolmogorov equation.

The collection of conditions sufficient for the Fokker-Planck approximation can be summarized as follows:

(1) The forcing function $f(t)$ is a stationary, normal random variable with zero average value and noninfinite integral scale.

(2) $f(t)$ is statistically independent of the response of the system $x(t)$, or at least uncorrelated with some specific functions of x.

(3) The largest statistically characteristic time of $f(t)$, $T_{\max}$, say, must be so much smaller than the smallest characteristic times of $x(t)$ and $\dot{x}(t)$, δ_x and δ_s, say, that there can exist a time θ which is very much larger than the former and very much smaller than both of the latter:

$$\delta_x, \delta_s \gg \theta \gg T_{\max}. \tag{50}$$

A possible choice of $T_{\max}$ is mentioned in (41). possible choices for δ_x and δ_s are what might be called the "Taylor microscales," [5]

$$\delta_x \equiv \{\langle x^2 \rangle / \langle \dot{x}^2 \rangle\}^{\frac{1}{2}}, \tag{51}$$

$$\delta_s \equiv \{\langle s^2 \rangle / \langle \dot{s}^2 \rangle\}^{\frac{1}{2}}. \tag{52}$$

These are the abcissa intercepts of the vertex-osculating parabolas of the $x(t)$ and $\dot{x}(t)$ autocorrelation functions, respectively.

In fact, the keeping of only the linear Δt term from (15) is equivalent to perfect autocorrelation of s over the time interval $\Delta t = \theta$:

$$R_s(\theta) \doteq 1. \tag{53}$$

In terms of instantaneous behavior, (48) means that x is approximated in the interval Δt by its tangent at t.

We cannot immediately write an identical expression for $R_{\dot{s}}(\theta)$ because of the f term in (10). The Δt term in (16) can be written as

$$[\langle \dot{s}(t) \rangle_c - \langle f(t) \rangle_c]\Delta t. \tag{54}$$

Since f is presumed independent of the response, $\langle f \rangle_c = \langle f \rangle = 0$, and the Fokker-Planck approximation in question is

$$a_s^{(1)} \doteq \langle \dot{s}(t) \rangle_c \Delta t. \tag{55}$$

Since $\langle \dot{s}(t) \rangle_c \neq \dot{s}(t)$, (55) does *not* automatically imply

$$R_{\dot{s}}(\theta) \doteq 1. \tag{56}$$

A complementary view is given by the instantaneous linear approximation, Eq. (49). This can be written as

$$s(t + \Delta t) \doteq s(t) + \dot{s}(t)\Delta t + \int_t^{t+\Delta t} [f(t_1) - f(t)]\, dt_1. \tag{57}$$

IV. COMPARISON BETWEEN FOKKER-PLANCK SOLUTION AND ANALOG-COMPUTER RESULTS

It is not obvious that the conditions cited as sufficient for the Fokker-Planck approximation are less likely to be fulfilled by a nonlinear system than by a linear one with the same orders of "sluggishness"

or "stiffness." A possible exception is the tendency toward Markovian behavior, mentioned after Eq. (49). This is not a formal proof and Doob's theorem[6] is reassuring only for linear systems. Also, the nonlinearity may generate higher harmonics, thereby reducing the range of validity. In any case, it is interesting to look at an experimental test.

The steady-state solution to (47) with $\psi_x = x + \beta x^3$ is[2]

$$P_{xs}(u, v) = C \exp\left\{-\left[\frac{v^2}{D} + \frac{2\alpha}{D}\left(\frac{u^2}{2} + \frac{\beta u^4}{4}\right)\right]\right\}, \quad (58)$$

where C is chosen to normalize P_{xs}. Because P_{xs} is not integrable in tabulated functions, we cannot write down an explicit expression for C.

We can see directly from (58) that the time derivative of the response $s \equiv \dot{x}$ is Gaussian, although x itself is not. The probability density function for x follows by integrating v from $-\infty$ to ∞:

$$P_x(u') = c_1 \exp\{-(2\alpha/\beta D)(u'^2 + u'^4)\}, \quad (59)$$

where

$$u' = (\beta/2)^{\frac{1}{2}}u.$$

Thus we can easily evaluate numerically[13]

$$\langle x^2 \rangle = \frac{2}{\beta}\phi_2\left(\frac{2\alpha}{\beta D}\right), \quad (60)$$

$$\frac{\langle x^4 \rangle}{\langle x^2 \rangle^2} = \phi_4\left(\frac{2\alpha}{\beta D}\right). \quad (61)$$

These two functions are plotted in Figs. 1 and 2.

Equation (8) was simulated, using operational amplifiers and 4-quadrant multipliers, as indicated in Fig. 3. The multipliers were of the diode chain[14] variety with no noticeable phase shift or attenuation below about 3 kHz. Since the analog simulation[15] was scaled such that all of the interesting regions in the spectrum were below about 300 Hz, there were no difficulties introduced by the multipliers.

The operational amplifiers had a flat response and no noticeable phase shift from dc to about 2 kHz.

The noise source[16] had a spectrum flat from dc to about 2 kHz as shown in Fig. 4. Its amplitude probability density was Gaussian to over 99.9% of its distribution as seen in Fig. 5. The derivative[17] of its output was also very nearly Gaussian, Fig. 6.

[13] J. B. Morton, Ph.D. thesis, The Johns Hopkins University, 1967.

[14] One was a Philbrick Q3MIP and the other a GPS.

[15] To facilitate measurements, the analog computer experiment was run with the time scaled 100 times faster than "real time." Thus, to compare experiment and theory, all frequencies in the experiment must be divided by 100.

[16] Solartron Random Signal Generator; Weston, Boonshaft, and Fuchs, Hatboro, Pennsylvania.

[17] The response of the differentiation circuit to a sinusoid yielded the correct amplitude (within ±1 db) and phase shift (±1°) from dc to about 2 KHz.

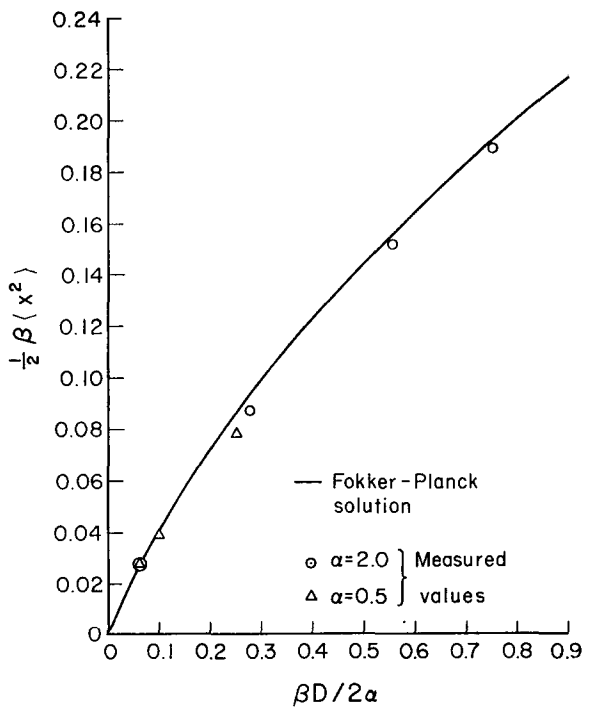


FIG. 1. Comparison of the measured mean-square response and the mean-square response predicted by the Fokker–Planck equation.

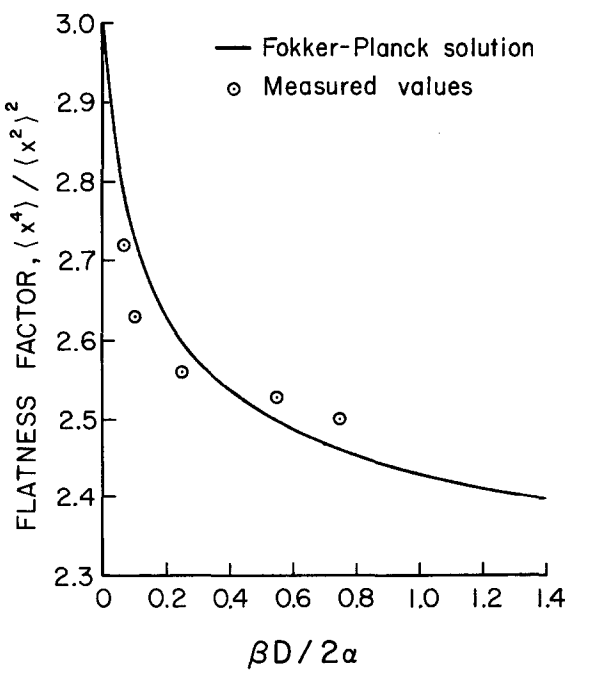


FIG. 2. Comparison of the measured and predicted flatness factor.

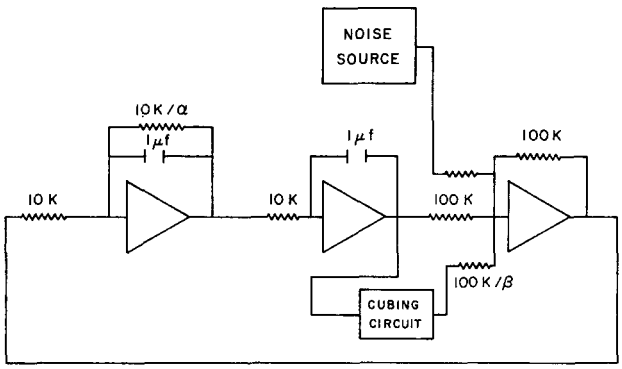


FIG. 3. Operational amplifiers and 4-quadrant multipliers for simulating Eq. (8).

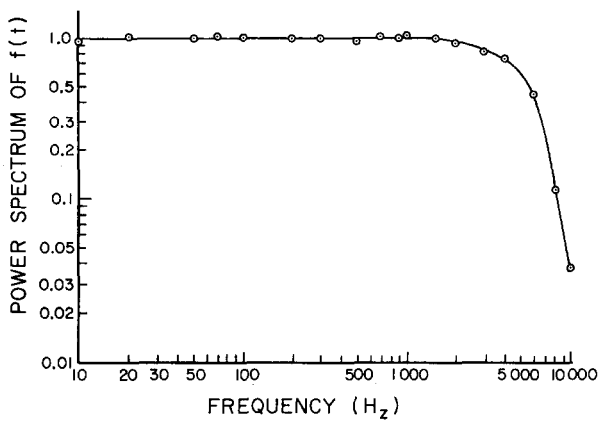


FIG. 4. Power spectrum of the output of the noise generator.

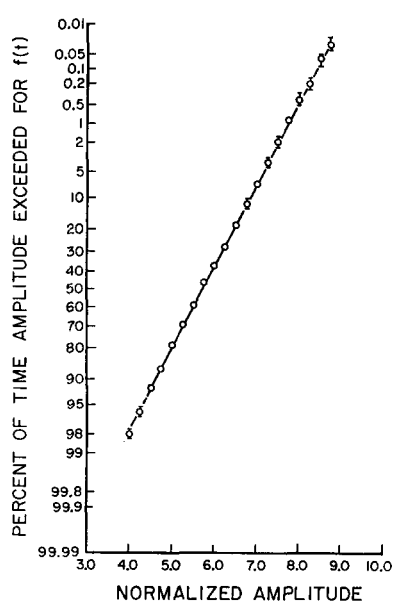


Fig. 5. Distribution function of the output of the noise generator.

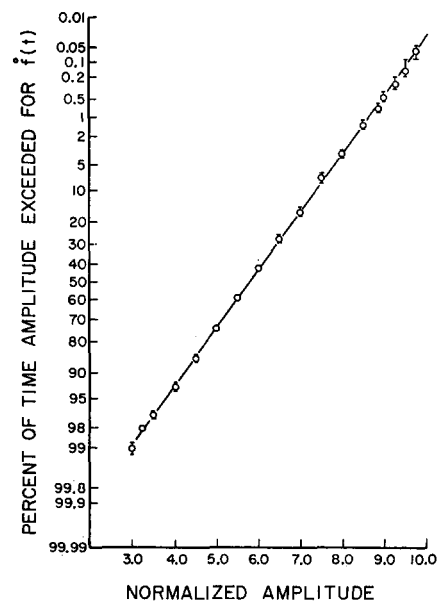


FIG. 6. Distribution function of the time derivative of the output of the noise generator.

The random signal in this device is produced by sampling 7 thyratrons at the same instant of time to produce one 7-bit binary number. Then the amplitude distribution is shaped by processing these random binary numbers. Finally, a square pulse is produced with an amplitude proportional to the processed random numbers. Each square pulse is independent of all others. Thus the autocorrelation function (the Fourier transform of Fig. 4) should be a straight line; cf. Fig. 7.

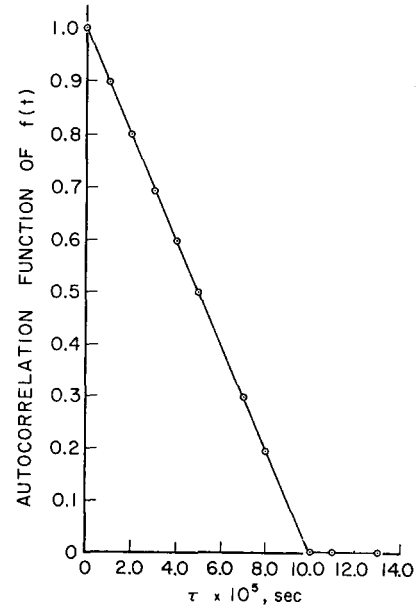


FIG. 7. Correlation function of the output of the noise generator.

From the foregoing description, we can see that $f(t)$ complies well with the Fokker–Planck requirement that it be independent of the response $x(t)$.

The measured results for mean-square response $\langle x^2 \rangle$ and kurtosis or "flatness factor" $\langle x^4 \rangle / \langle x^2 \rangle^2$ are shown in Figs. 1 and 2. The abcissa $\beta D/2\alpha$ is, in part, a measure of the importance of the nonlinear term in (8). The results computed from the Fokker–Planck solution (solid lines) agree with the experiments within the experimental uncertainty of the latter.

For this forcing function, the integral time scale[15] is $T_f = 5 \times 10^{-3}$ sec. The simple form of $R_f(\tau)$ (Fig. 7) offers an obvious alternative choice for "largest characteristic time" in f: $T_f' = 10^{-2}$ sec, where R_f hits zero.

The values of α, β, and D selected for these experiments ($\alpha = 0.5, 2.0$; $\beta = 0.10, 0.25, 1.00, 2.00, 3.00$; $D = 1.0$) were such that the oscillator was sometimes less than critically damped, sometimes more. In the worst case, we can estimate δ_x (cf. Eq. 51) from measured values. $\delta_x \approx 0.7$ sec. From the shape of the response spectrum[13] and from Eq. (8), the "microscale" for s is roughly one-half of δ_x;$\delta_s \approx 0.3$ sec.

The Fokker–Planck condition expressed in (50) now requires that there exist a time interval θ such that

$$0.005 \ll \theta \ll 0.3.$$

Thus (50) is roughly satisfied. It would be interesting to extend this study over a range of relative values of T_f and δ_s.

In situations where (50) is not fulfilled, the Fokker–Planck approach can be rescued by considering not the probability of the response alone, but rather the joint probability of response functions and forcing function.[2] The appropriate formalism is relatively simple if $f(t)$ can itself be treated as the response of another system to another random forcing function $g(t)$ of suitably small correlation time.

As a simple example, consider the system

$$\dot{x} + \alpha x = f(t). \tag{62}$$

If $x(t)$ is not sufficiently sluggish to the frequencies dominant in $f(t)$, x cannot be treated alone by a Fokker–Planck equation.

If, however, f can be identified as the response of another system, one which *does* fulfill the Fokker–Planck requirements,

$$\gamma^{-1}\dot{f} + f = \gamma^{-1}g(t), \tag{63}$$

then the conditonal joint probability of x and f obeys a Fokker–Planck equation. The probability of x is then the integral of the solution over all values of f:

$$P_x(a) = (2\pi D')^{-\frac{1}{2}} e^{-a^2/2D'}, \tag{64}$$

where

$$D' = \frac{D}{\alpha\gamma(\alpha + \gamma)} = \frac{D}{\gamma^2}\left\{\frac{1}{\alpha(1 + \alpha/\gamma)}\right\}.$$

In the limit $D/\gamma^2 = \text{const}$, as $\gamma \to \infty$, the x process becomes simply Fokker–Planck because $f(t)$ becomes a much more rapidly varying function. If $\gamma = \alpha$, $D' = (1/2\gamma)(D/\gamma^2)$, just one-half the 1-variable Fokker–Planck limit.

ACKNOWLEDGMENT

We should like to thank James Riley for his helpful comments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 10, NUMBER 2 FEBRUARY 1969

Generalization of Green's Theorem

JOSHUA N. GOLDBERG*
Syracuse University, Syracuse, New York
AND
EZRA T. NEWMAN†
University of Pittsburgh, Pittsburgh, Pennsylvania

(Received 13 August 1968)

For a system of field equations which is derivable from a Lagrangian whose density is (i) homogeneous quadratic in the first derivatives of the field variables $y_{A,\mu}$ and (ii) homogeneous of degree n in the undifferentiated field variables y_A, one has the identity

$$(n+1)z_A L^A(y) - y_A M^A(y,z) \equiv t^\rho{}_{,\rho},$$

where $M^A(y,z)$ is the first-order change in the field equations $L^A(y)=0$ under the mapping $y_A \to y_A + z_A$. The specific example of general relativity is discussed.

I. INTRODUCTION

In order to discuss conservation laws without explicitly introducing the Lagrangian formalism, Goldberg[1] has used Green's identity for a system of self-adjoint field equations to establish the Noether equation[2]

$$\delta y_A L^A \equiv t^\rho{}_{,\rho}, \tag{1}$$

where $y_A(x)$ $(A=1,\cdots,N)$ are the field variables and δy_A is an infinitesimal mapping of solutions of the field equations $L^A=0$ into other (possibly equivalent) solutions. For self-adjoint linear field equations the general statement of Green's identity may be written

$$z_A L^A(y) - y_A L^A(z) \equiv t^\rho{}_{,\rho}. \tag{2}$$

If $y_A L^A(z)$ can itself be written as a divergence for a particular choice of z_A, or if $L^A(z)=0$, then (2) reduces to the Noether equation (1). A generalization of this result to nonlinear equations would be very useful and we show how this can be achieved for a certain class of nonlinear equations. The result is then applied to the Einstein field equations for general relativity.

II. GENERALIZED GREEN'S THEOREM

Theorem: Consider a system of field equations $L^A(y)=0$ $(A=1,\cdots,N)$ derivable from a Lagrangian density which is homogeneous quadratic in the first derivatives of the field variables $y_{A,\mu}$:

$$\mathfrak{L} = \tfrac{1}{2}\Lambda^{A\mu B\nu} y_{A,\mu} y_{B,\nu} \tag{3}$$

and, furthermore, where the coefficients $\Lambda^{A\mu B\nu} = \Lambda^{B\nu A\mu}$ themselves are homogeneous of degree n in the undifferentiated field variables y_A. Define $M^A(y,z)$, linear in z_A, as

$$M^A(y,z) \equiv L^A(y+z) - L^A(y) + 0(z^2),$$

then

$$(n+1)z_A L^A(y) - y_A M^A(y,z) \equiv t^\rho{}_{,\rho}, \tag{4}$$

with

$$\begin{aligned} t^\mu \equiv{}& (n+1)z_A \Lambda^{A\mu B\nu} y_{B,\nu} \\ & - y_A[\Lambda^{A\mu B\nu} z_{B,\nu} + z_C \partial^C \Lambda^{A\mu B\nu} y_{B,\nu}], \\ & \partial^C \Lambda^{A\mu B\nu} \equiv \partial \Lambda^{A\mu B\nu}/\partial y_C. \end{aligned} \tag{5}$$

The proof is straightforward using the following:

$$L^A(y) \equiv (\Lambda^{A\mu B\nu} y_{B,\nu})_{,\mu} - \tfrac{1}{2}\partial^A \Lambda^{C\mu B\nu} y_{C,\mu} y_{B,\nu}, \tag{6}$$

$$\begin{aligned} M^A(y,z) \equiv{}& (\Lambda^{A\mu B\nu} z_{B,\nu} + z_C \partial^C \Lambda^{A\mu B\nu} y_{B,\nu})_{,\mu} \\ & - \partial^A \Lambda^{C\mu B\nu} y_{C,\mu} z_{B,\nu} \\ & - \tfrac{1}{2} z_D \partial^D \partial^A \Lambda^{C\mu B\nu} y_{C,\mu} y_{B,\nu}, \end{aligned} \tag{7}$$

and, from Euler's theorem on homogeneous functions,

$$y_C \partial^D \partial^C \Lambda^{A\mu B\nu} = (n-1)\partial^D \Lambda^{A\mu B\nu}, \tag{8}$$

$$y_C \partial^C \Lambda^{A\mu B\nu} = n\Lambda^{A\mu B\nu}. \tag{9}$$

Note that when $n=-1$ we have the interesting result that

$$y_A M^A(y,z) \equiv t^\mu{}_{,\mu}. \tag{10}$$

However, if $\Lambda^{A\mu B\nu}$ is homogeneous of degree n with respect to y_A, then it is always possible to introduce new variables Y_A such that

$$y_A = y_A(Y) \tag{11}$$

is homogeneous of degree m in Y_A. Then

$$\Lambda'^{A\mu B\nu} = \Lambda^{C\mu D\nu} \frac{\partial y_C}{\partial Y_A} \frac{\partial y_D}{\partial Y_B} \tag{12}$$

* Research supported in part by the National Science Foundation.
† Research supported in part by the Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force, and the Office of Scientific Research.

[1] J. N. Goldberg, J. Math. Phys. 9, 674 (1968).
[2] A. Trautman, *Gravitation*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), pp. 169–198.

will be homogeneous of degree $l = mn + 2(m-1)$. As we see in the next section, for general relativity, the degree of homogeneity is $n = -1$, with $\mathfrak{g}^{\mu\nu} = \sqrt{-g}\, g^{\mu\nu}$ as the basic field variables, and $n = -3$ if $\mathfrak{g}_{\mu\nu}$ is the basic field variables.

III. GENERAL RELATIVITY

The Lagrangian of general relativity, written in terms of the contravariant metric density, takes the very simple form[3]

$$\mathfrak{L} = \tfrac{1}{8}\{2\mathfrak{g}^{\rho\sigma}\mathfrak{g}_{\lambda\iota}\mathfrak{g}_{\kappa\tau} - \mathfrak{g}^{\rho\sigma}\mathfrak{g}_{\iota\kappa}\mathfrak{g}_{\lambda\tau} - 4\delta^{\sigma}_{\kappa}\delta^{\rho}_{\lambda}\mathfrak{g}_{\iota\kappa}\}\mathfrak{g}^{\iota\kappa}{}_{,\rho}\mathfrak{g}^{\lambda\tau}{}_{,\sigma}$$
$$= \tfrac{1}{2}\Lambda_{(\iota\kappa)}{}^{\rho}{}_{(\lambda\tau)}{}^{\sigma}\,\mathfrak{g}^{\iota\kappa}{}_{,\rho}\mathfrak{g}^{\lambda\tau}{}_{,\sigma}. \tag{13}$$

Clearly, $\Lambda_{(\iota\kappa)}{}^{\rho}{}_{(\lambda\tau)}{}^{\sigma}$ is homogeneous of degree -1 in $\mathfrak{g}^{\iota\kappa}$, and therefore we are faced with the situation described in Eq. (10). The field equations are simply

$$L_{(\mu\nu)} = R_{\mu\nu} \equiv \left\{\begin{matrix}\rho\\ \mu\rho\end{matrix}\right\}_{,\nu} - \left\{\begin{matrix}\rho\\ \mu\nu\end{matrix}\right\}_{,\rho} - \left\{\begin{matrix}\rho\\ \sigma\rho\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \mu\nu\end{matrix}\right\} + \left\{\begin{matrix}\rho\\ \sigma\nu\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \mu\rho\end{matrix}\right\}. \tag{14}$$

Since the difference between two affine connections is a tensor,[4] it is easy to show that

$$M_{(\mu\nu)} \equiv \left\{\delta^{\rho}_{\nu}\left(\bar{\delta}\left\{\begin{matrix}\sigma\\ \mu\sigma\end{matrix}\right\}\right) - \left(\bar{\delta}\left\{\begin{matrix}\rho\\ \mu\nu\end{matrix}\right\}\right)\right\}_{;\rho}. \tag{15}$$

Clearly, then,

$$t^{\rho} \equiv \mathfrak{g}^{\mu\nu}\left(\bar{\delta}\left\{\begin{matrix}\rho\\ \mu\nu\end{matrix}\right\}\right) - \mathfrak{g}^{\rho\mu}\left(\bar{\delta}\left\{\begin{matrix}\sigma\\ \mu\sigma\end{matrix}\right\}\right). \tag{16}$$

With $\mathfrak{g}_{\mu\nu}$ as the basic field variable, $\Lambda'^{(\iota\kappa)\rho(\lambda\tau)\sigma}$ is of degree -3. The field equations are

$$L^{(\mu\nu)} = gR^{\mu\nu} = -\mathfrak{g}^{\mu\iota}\mathfrak{g}^{\nu\kappa}R_{\iota\kappa} \tag{17}$$

and

$$M^{(\mu\nu)} = -\mathfrak{g}^{\mu\iota}\mathfrak{g}^{\nu\kappa}M_{(\iota\kappa)} - R_{\iota\kappa}(\bar{\delta}\mathfrak{g}^{\mu\iota}\mathfrak{g}^{\nu\kappa} + \mathfrak{g}^{\mu\iota}\bar{\delta}\mathfrak{g}^{\nu\kappa}). \tag{18}$$

Then,

$$-2\bar{\delta}\mathfrak{g}_{\mu\nu}gR^{\mu\nu} - \mathfrak{g}_{\mu\nu}M^{(\mu\nu)} \equiv -t^{\mu}{}_{,\mu}. \tag{19}$$

Similar results hold, of course, for other choices of the variables.

[3] J. N. Goldberg, Phys. Rev. **111**, 315 (1958).

[4] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Inc., New York, 1942), pp. 191–2.

IV. APPLICATION TO CONSERVATION LAWS

Equation (4) leads immediately to a generalization of Noether's theorem. From the manner in which M^A is defined, it is clear that if y_A satisfies $L^A(y) = 0$, and z_A satisfies $M^A(y, z) = 0$, then, to first order in z_A, $y'_A = y_A + z_A$ also satisfies $L^A(y') = 0$; i.e., z_A is then an infinitesimal invariant transformation. $t^{\mu}{}_{,\mu} = 0$ is then the associated conservation law. In the usual form of Noether's theorem, the invariant transformation z_A is considered as a local or point function of the variables y_A. (This is true in the usual discussions of coordinate invariance, gauge invariance, etc.) In our case the z_A, being solutions of $M^A(y, z) = 0$ are, in general, functionals of y_B, hence nonlocal.

An interesting variation on the above discussion occurs when the function $\Lambda^{A\mu B\nu}$ is homogeneous of order -1. Equation (10), rather than (4), then applies. It is seen that, in this case, we have a conservation law $t^{\mu}{}_{,\mu} = 0$ when $M^A = 0$, without requiring that $L^A = 0$. In this case the significance of the differential conservation law is certainly obscure. However, if a known invariance group exists, such as the coordinate invariance of general relativity, then for z_A defined by this invariance group $M^A(y, z) \equiv f^A{}_B L^B$, that is, $M^A = 0$ modulo L^A. If one works this out explicitly in the case of general relativity, Eq. (10) yields the Noether equation again.

For linear theories $n = 0$ always. Therefore one always has a correspondence to the Noether equation as indicated in Ref. 1.

Note Added in Proof: Arthur Komar has pointed out that H. Streudel has previously obtained these results for linear theories: H. Streudel, Nuovo Cimento **39**, 395 (1965); Z. Naturf. **21a**, 1826 (1965).

ACKNOWLEDGMENTS

We are indebted to Professor A. Janis for many helpful discussions. The authors also wish to thank the Physics Division of the Aspen Institute, where the basis of this work was discussed, for their hospitality.