Invariant imbedding and the resolvent of Fredholm integral equation with composite symmetric kernel*

Richard Bellman

Departments of Mathematics, Electrical Engineering, and Medicine, University of Southern California, Los Angeles, California 90007

Sueo Ueno

Department of Astronomy, University of Southern California, Los Angeles, California 90007

(Received 10 April 1972)

Consider a Fredholm, integral equation for $f(t,b)$ with a composite displacement kernel $f(t,b) = g(t) + \lambda\int_a^b [E(t,y) + F(t+y)]\, f(y,b)dy$, where $g(t)$ is a given forcing function, the parameter λ is defined over $0 < \lambda \leq 1, 0 \leq a \leq t \leqslant b$, and the E and F functions are together the displacement kernels. Extending the invariant imbedding method given in our preceding paper [cf. Bellman, Kagiwada, Kalaba, and Ueno, J. Math Phys. 9, 906 (1968)], we show how the Bellman–Krein formula provides us with a Cauchy system of the functional equations governing the resolvent and the scattering function. The invariant imbedding equations for the scattering function and the auxiliary function for $a = 0$ reduce to those given in a preceding paper [cf. Casti, Kalaba, and Ueno J. Quant. Spectry. Radiative Transfer 9, 537 (1969)], which treated with the diffuse reflection and transmission of radiation by a finite isotropically scattering atmosphere bounded by a specular reflector at the bottom.

1. INTRODUCTION

In a preceding paper (cf. Bellman, Kagiwada, Kalaba, and Ueno[1]), it was shown how to get the invariant imbedding equations of the X and Y functions and the resolvent kernel of the Fredholm integral equation with a symmetric kernel, reducible to a auxiliary equation. In this paper, extending the procedure to a Fredholm integral equation with a composite displacement kernel, we show that the Bellman–Krein formula is a powerful tool for getting the Cauchy system of the functional equations for the scattering function and the resolvent kernel of the Fredholm integral equation. The invariant imbedding equations for the scattering function and the auxiliary function reduce to those for the scattering function and the source function in a finite isotropically scattering atmosphere bounded by a specular reflector at the bottom (cf. Casti, Kalaba, and Ueno[2]). The resolvent kernel permits us to calculate the internal radiation field in a finite atmosphere with a given distribution of emitting sources and a specular reflector at the bottom.

In a subsequent paper, extending the procedure developed in this paper, it will be shown how the diffuse scattering and transmission functions of radiation in a finite isotropically scattering atmosphere by a Lambert's law reflector at the bottom can also be found in a straightforward manner. So far as we know, the result is new. Up to the present only the total scattering and the diffuse transmission functions have been discussed (cf. Chandrasekhar[3]; van de Hulst[4]; Sobolev[5]; Bellman, Kagiwada, Kalaba, and Ueno[6]; Kagiwada and Kalaba[7]).

2. BELLMAN-KREIN FORMULA FOR RESOLVENT OF FREDHOLM INTEGRAL EQUATION

Consider a Fredholm integral equation with a composite displacement kernel

$$f(t,b) = g(t) + \lambda\int_a^b [E(t,y) + F(t+y)]f(y,b)dy, \tag{1}$$

where $g(t)$ is a given forcing function, a constant parameter λ is defined over $0 < \lambda \leq 1$, and $0 \leq a \leq t \leq b$.

In Equation (1), $E(t,y)$ and $F(t+y)$ are positive displacement kernels, e.g.,

$$E(t,y) = E(|t-y|), \tag{2}$$

$$E(s) = \int_0^1 e^{-s/z}\frac{dz}{z}, \tag{3}$$

and

$$F(s) = \int_0^1 e^{-s/z}A(z)\,\frac{dz}{z}, \tag{4}$$

where $s \geq 0$.

Let the resolvent of Equation (1) be denoted by $K(t,y,b)$, which is symmetric with respect to t and y, because of the symmetric character of the kernels E and F. Then, Eq. (1) yields

$$f(t,b) = g(t) + \int_a^b K(t,y,b)g(y)dy. \tag{5}$$

The resolvent $K(t,y,b)$ is governed by the following integral equations:

$$K(t,y,b) = \lambda G(t,y) + \lambda\int_a^b G(t,z)K(z,y,b)dz, \tag{6}$$

$$K(t,y,b) = \lambda G(t,y) + \lambda\int_a^b K(t,z,b)G(z,y)dz, \tag{7}$$

where

$$G(t,y) = E(t,y) + F(t+y). \tag{8}$$

On differentiating Eq. (6) with respect to b, we have

$$K_b(t,y,b) = \lambda G(t,b)K(b,y,b) + \int_a^b G(t,z)K_b(z,y,b)dz, \tag{9}$$

where the subscript represents partial differentiation with respect to b. As in our preceding paper (Bellman, Kagiwada, Kalaba, and Ueno[1]), putting

$$K_b(t,y,b) = K(t,b,b)K(b,y,b), \tag{10}$$

inserting it into Eq. (9), and allowing for Eq. (6), we get

$$\begin{aligned} K_b(t,y,b) &= K(b,y,b)\Big(\lambda G(t,b) \\ &\qquad + \lambda\int_a^b G(t,z)K(z,b,b)dz\Big) \\ &= K(b,y,b)K(t,b,b). \end{aligned} \tag{11}$$

Equation (10) is the Bellman–Krein formula for the Fredholm resolvent $K(t,y,b)$ (cf. Bellman[8]; Krein[9]). By writing

$$\Phi(t,b) = K(t,b,b) = K(b,t,b), \tag{12}$$

Copyright © 1973 by the American Institute of Physics

Eq. (9) is rewritten in the form

$$K_b(t, y, b) = \Phi(t, b)\Phi(y, b), \tag{13}$$

which is useful for our further procedure.

3. THE CAUCHY SYSTEM FOR THE AUXILIARY FUNCTION

Introducing a new auxiliary function B, which satisfies

$$B(t, b, v) = \lambda[e^{-(b-t)/v} + A(v)e^{-(b+t)/v}] + \lambda\int_a^b [E(t, y) + F(t + y)]B(y, b, v)dy, \tag{14}$$

and rewriting it in terms of Fredholm resolvent K, we get

$$B(t, b, v) = \lambda[e^{-(b-t)/v} + A(v)e^{-(b+t)/v}] + \lambda\int_a^b [e^{-(b-y)/v} + A(v)e^{-(b+y)/v}]K(y, t, b)dy. \tag{15}$$

On differentiating Eq. (15) with respect to b and recalling Eq. (13), we obtain

$$B_b(t, b, v) = -(B(t, b, v)/v) + \lambda\Phi(t, b)\left(1 + A(v)e^{-2b/v} + \int_a^b [e^{-(b-y)/v} + A(v)e^{-(b+y)/v}]\Phi(y, b)dy\right). \tag{16}$$

Making use of Eq. (15), Eq. (16) reduces to

$$B_b(t, b, v) = -(B(t, b, v)/v) + \Phi(t, b)B(b, b, v), \tag{17}$$

where $\Phi(t, b)$ may be expressed in the form

$$\Phi(t, b) = \int_0^1 B(t, b, v)dv/v. \tag{18}$$

Let the scattering function be denoted by $R(b; v, u)$ expressed in terms of B function

$$R(b; v, u) = \int_a^b B(y, b, u)[e^{-(b-y)/v} + A(v)e^{-(b+y)/v}]dy. \tag{19}$$

The differentiation of Eq. (19) with respect to b yields

$$R_b(b; v, u) = B(b, b, u)[1 + A(v)e^{-2b/v}] - \frac{1}{v}\int_a^b B(y, b, u)[e^{-(b-y)/v} + A(v)e^{-(b+y)/v}]dy + \int_a^b B_b(y, b, u)[e^{-(b-y)/v} + A(v)e^{-(b+y)/v}]dy. \tag{20}$$

On substituting Eq. (17) into Eq. (20), and allowing for Eqs. (14) and (18), after some minor rearrangement of terms, we have

$$R_b(b; v, u) = -\left(\frac{1}{v} + \frac{1}{u}\right)R(b; v, u) + \lambda\left(1 + A(v)e^{-2b/v} + \int_0^1 R(b; v, w)\frac{dw}{w}\right)\left(1 + A(u)e^{-2b/u} + \int_0^1 R(b; z, u)\frac{dz}{z}\right), \tag{21}$$

together with the initial condition

$$\lim_{b=a} R(b; v, u) = 0. \tag{22}$$

Equation (21) is the desired invariant imbedding equation for the scattering function. Furthermore, it is readily proved that the scattering function R is symmetric with respect to v and u, as a consequence of

$$\int_a^b B(t, b, v)B(t, b, u)dt = \lambda R(b; v, u) + \int_a^b B(t, b, u)B(t, b, v)dt - \lambda R(b; u, v). \tag{23}$$

Then, we get

$$R(b; v, u) = R(b; u, v) \tag{24}$$

Equation (21) is similar in form to the scattering function of radiation for a finite isotropically scattering atmosphere bounded by a specular reflector at the the bottom (cf. Casti, Kalaba, and Ueno[2]). Once the scattering function has been given, with the aid of Eqs. (17) and (18), we compute $\Phi(t, b)$. Then, the Bellman–Krein formula (13) yields the required Fredholm resolvent $K(t, y, b)$

4. STATEMENT OF THE CAUCHY SYSTEM

Let us restate the initial-value problem that determines the set of equations for the computation of the scattering function $R(b; v, u)$ and the resolvent $K(t, y, b)$. The Cauchy system for R and B functions are as below:

$$R_b(b; v, u) = -\left(\frac{1}{v} + \frac{1}{u}\right)R(b; v, u) + \lambda\left(1 + A(v)e^{-2b/v} + \int_0^1 R(b; v, w)\frac{dw}{w}\right)\left(1 + A(u)e^{-2b/u} + \int_0^1 R(b; w, u)\frac{dw}{w}\right), \quad b - a > 0. \tag{25}$$

$$B_b(t, b, v) = -\frac{B}{v}(t, b, v) + \lambda\left(1 + A(v)e^{-2b/v} + \int_0^1 R(b; w, v)\frac{dw}{w}\right)\int_0^1 B(t, b, z)dz, \tag{26}$$

$$\Phi(t, b) = \int_0^1 B(t, b, v)\frac{dv}{v}, \tag{27}$$

$$K_b(t, y, b) = \Phi(t, b)\Phi(y, b), \tag{28}$$

where

$$R(b; v, u) = R(b; u, v), \tag{29}$$

$$K(t, y, b) = K(y, t, b). \tag{30}$$

The initial conditions are

$$\lim_{b=a} R(b; v, u) = 0, \tag{31}$$

$$\lim_{t=b} B(t, b, v) = \lambda\left(1 + A(v)e^{-2t/v} + \int_0^1 R(t; w, v)\frac{dw}{w}\right), \tag{32}$$

$$\lim_{b=y} K(t, y, b) = \Phi(t, y). \tag{33}$$

*Supported by the National Science Foundation under Grant No. GP 29049, and by the Atomic Energy Commission under Grant No. AT(40-3)-113 Project No. 19.

[1]R. Bellman, H. Kagiwada, R. Kalaba, and S. Ueno, J. Math. Phys. 9, 906 (1968).
[2]J. Casti, R. Kalaba, and S. Ueno, J. Quant. Spectry. Radiative Transfer 9, 537 (1969).
[3]S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., Oxford, 1950).
[4]H. C. van de Hulst, Astrophys. J. 107, 220 (1948).
[5]V. V. Sobolev, Dokl. Akad. Nauk SSSR 61, No. 5 (1948).
[6]R. Bellman, H. Kagiwada, R. Kalaba, and S. Ueno, Icarus 7, 365 (1967).
[7]H. Kagiwada and R. Kalaba, J. Quant. Spectry. Radiative Transfer 11, 1101 (1971).
[8]R. Bellman, Proc. Amer. Math. Soc. 8, 435 (1957).
[9]M. G. Krein, Dokl. Akad. Nauk SSSR 100, 413 (1955).

Clebsch-Gordan coefficients for nonunitary groups

A. Aviran

Department of Physics, Technion-Israel Institute of Technology, Haifa, Israel

D. B. Litvin

Department of Physics, University of British Columbia, Vancouver, Canada

(Received 29 March 1972)

It is shown that calculating Clebsch–Gordan coefficients of a nonunitary group can be reduced to formulas containing only representations of the unitary subgroup and additional conditions due to the antiunitary symmetry. This is another example demonstrating that, in applications involving corepresentations of nonunitary groups, the analysis can be made mainly in terms of representations of its unitary part.

I. INTRODUCTION

Nonunitary groups and their corepresentations are of great importance in magnetic materials. In such material an antiunitary element is a product of time reversal and an element of a space group. In nonmagnetic materials time reversal is itself a symmetry element. In every case where an antiunitary element is added to the ordinary space group there is a need to deal with corepresentations. The theory of nonunitary groups and their corepresentations was founded by Wigner,[1] developed by Dimmock and Wheeler,[2–4] Dimmock,[5] and has been reviewed by Bradley and Davis.[6]

The problem often arises of decomposing a reducible corepresentation of a nonunitary group into a sum of irreducible parts. An example of this is in determining selection rules governing transitions in magnetic crystals, where the reducible corepresentation is a direct product of two irreducible corepresentations. Sometimes more detailed information is required, and one must calculate the matrix which transforms the corepresentation into a reduced form. The elements of this matrix are called the Clebsch–Gordan coefficients. Such information is needed, for example, in the Eckart–Wigner theorem.[1,7]

The Eckart–Wigner theorem was originally applied to calculate matrix elements of operators in physical systems of spherical symmetry, and found widespread use in such varied fields as atomic spectra, NMR, and elementary particles. Koster[8] generalized this theorem to make it applicable to other unitary groups, and this generalization takes the form

$$(\psi^i_\alpha | P^k_\sigma | \psi^j_\beta) = a_1 U_{\sigma,(\alpha\beta)} + a_2 U_{\sigma+n_k,(\alpha\beta)} + \cdots,$$

where i, j, and k denote irreducible representations of a unitary group G; $a_1, a_2, \cdots$ are constants called "reduced matrix elements" and U is the matrix of Clebsch–Gordan coefficients.

For physical systems of spherical symmetry the Eckart–Wigner theorem takes on a simple form with only one term on the right-hand side of the above relation. In such a case knowing only the Clebsch–Gordan coefficients one is able to find selection rules and compare transition intensities. For systems of other unitary symmetry one usually needs to know more information about the physical system.

The Eckart–Wigner theorem has been generalized by Aviran and Zak[9] to nonunitary groups. It was shown that the addition of an antiunitary element leads in general to additional connections among the reduced matrix elements.

This paper deals with the problem of finding the Clebsch–Gordan coefficients for nonunitary groups. The method used is one put forward by Aviran and Zak[9,10] based on the method developed by Koster[8] for unitary groups. It is shown that the finding of Clebsch–Gordan coefficients can be reduced to formulas containing only representations of the unitary subgroup of the nonunitary group, and additional conditions due to the antiunitary symmetry.

We review in Sec. II the construction of irreducible corepresentations and the calculation of reduction coefficients for nonunitary groups. We emphasize the role played by the unitary subgroup. In Sec. III a method is derived of finding the Clebsch–Gordan coefficients for nonunitary groups. An example is given in Sec. IV.

II. COREPRESENTATIONS OF NONUNITARY GROUPS

A nonunitary group H contains elements half of which are unitary and half antiunitary. The $N/2$ unitary elements, denoted by u, form an invariant subgroup G of index two and we can write H as $H = G + Ga_0$, where a_0 is a fixed antiunitary element. The irreducible corepresentations D^k of a nonunitary group H are constructed in one of three ways depending on the following classification of the irreducible representations $\Delta^k(a_0^{-1}ua_0)^*$ of the unitary subgroup G of H[1]:

Type I: $\Delta^k(u)$ is equivalent to $\Delta^k(a_0^{-1}ua_0)^*$,

$$\Delta^k(a_0^{-1}ua_0)^* = \beta^{k^{-1}}\,\Delta^k(u)\beta^k \quad \text{and} \quad \beta^k\beta^{k*} = \Delta^k(a_0^2).$$

Type II: $\Delta^k(u)$ is equivalent to $\Delta^k(a_0^{-1}ua_0)^*$,

$$\Delta^k(a_0^{-1}ua_0)^* = \beta^{k^{-1}}\Delta^k(u)\beta^k \quad \text{but} \quad \beta^k\beta^{k*} = -\,\Delta^k(a_0^2).$$

Type III: $\Delta^k(u)$ is not equivalent to $\Delta^k(a_0^{-1}ua_0)^*$.

The three types of irreducible corepresentations of H corresponding to the above classification are[1]

Type I: $D^k(u) = \Delta^k(u), \quad D^k(ua_0) = \Delta^k(u)\beta^k.$

Type II:

$$D^k(u) = \begin{pmatrix} \Delta^k(u) & \\ & \Delta^k(u) \end{pmatrix},$$

$$D^k(ua_0) = \begin{pmatrix} & \Delta^k(u)\beta^k \\ -\,\Delta^k(u)\beta^k & \end{pmatrix}. \tag{1}$$

Copyright ©1973 by the American Institute of Physics

Type III:

$$D^k(u) = \begin{pmatrix} \Delta^k(u) & \\ & \Delta^k(a_0^{-1}ua)^* \end{pmatrix},$$

$$D^k(ua_0) = \begin{pmatrix} & \Delta^k(ua_0^2) \\ \Delta^k(a_0^{-1}ua_0)^* & \end{pmatrix}.$$

The number of times an irreducible corepresentation D^k is contained in a reducible corepresentation D is denoted by C^k and calculated from[11]

$$C^k = \sum_u \chi(D(u))\chi(D^k(u))^* / \sum_u \chi(D^k(u))\chi(D^k(u))^*, \tag{2}$$

where $\chi(D^k(u))$ is the character of $D^k(u)$. The sums in (2) are over the elements u of the unitary subgroup only. When the reducible corepresentation is direct product of two irreducible corepresentations, $D = D^i \times D^j$, eq. (2) takes the form[11]

$$C_{ij}^k = \sum_u \chi(D^i(u))\,\chi(D^j(u))\chi(D^k(u))^* \Big/ \sum_u \chi(D^k(u))\chi(D^k(u))^*. \tag{3}$$

By using the explicit form of the irreducible corepresentations given in (1), the coefficients C_{ij}^k can be written in terms of coefficients d_{ij}^k, the number of times the irreducible representations Δ^k of the unitary subgroup is contained in the reduced form of the direct product $\Delta^i \times \Delta^j$. The explicit form of the relation between the C_{ij}^k and the d_{ij}^k depends on the types of the irreducible corepresentations D^i, D^j, and D^k appearing in (3). The relation between C_{ij}^k and the d_{ij}^k, for all possible cases, has been given by Bradley and Davis.[6]

III. CLEBSCH–GORDAN COEFFICIENTS FOR NONUNITARY GROUPS

The matrix U, whose elements are the Clebsch–Gordan coefficients, transform a corepresentation D into reduced form in the following manner[1]:

$$UD(u)U^{-1} = D_r(u) = \begin{pmatrix} D^k(u) & & & & \\ & \ddots & & & \\ & & D^k(u) & & \\ & & & D^m(u) & \\ & & & & \ddots \end{pmatrix}, \tag{4}$$

$$UD(a)U^{-1*} = D_r(a) = \begin{pmatrix} D^k(a) & & & & \\ & \ddots & & & \\ & & D^k(a) & & \\ & & & D^m(a) & \\ & & & & \ddots \end{pmatrix}, \tag{5}$$

where u is a unitary element, a an antiunitary element, and D_r the reduced form of D.

The matrices $D(u)$, for all u, form a representation of the unitary subgroup G of H. The irreducible corepresentations appearing in $D_r(u)$, for all u, are either irreducible representations or sums of irreducible representations of the unitary group G. Consequently, to find the matrix U from (4) alone can be considered a calculation of a matrix which transforms a representation of a unitary group into reduced form. Such a calculation can be preformed using Koster's method.[8] The matrix U so found is not unique, and requiring that U also satisfies (5) imposes additional conditions on its elements.

The theory of corepresentations is such that a single method applicable simultaneously to all three types of irreducible corepresentations which may appear in D_r is unobtainable. We therefore discuss three cases corresponding to the three types of irreducible corepresentations. In each case we derive from (4), using Koster's method,[8] equations from which the elements of U are calculated, and the additional conditions on these elements imposed by (5).

A. Type I corepresentations

We assume that a Type I irreducible corepresentation D^k of dimension d appears l times in the reduced form D_r. To calculate the dl rows of U corresponding to these corepresentations, we rewrite (4) as $D(u) = U^{-1}D_r(u)U$ take the pqth element, multiply by $D^k(u)^*_{mn} = \Delta^k(u)^*_{mn}$, sum on u:

$$\frac{d}{N/2}\sum_u D(u)_{pq}\Delta^k(u)^*_{mn} = \frac{d}{N/2}\sum_{st} U^*_{sp}U_{tq}\sum_u D_r(u)_{st}\Delta^k(u)^*_{mn}.$$

Using the explicit form of $D_r(u)$ and the orthogonality relations for irreducible representations, we have

$$\frac{d}{N/2}\sum_u D(u)_{pq}\Delta^k(u)^*_{mn} = U^*_{mp}U_{nq} + U^*_{d+m,p}U_{d+n,q} + \cdots + U^*_{(l-1)d+m,p}U_{(l-1)d+n,q}. \tag{6}$$

The elements of U calculated from (6) satisfy (4),[8] but not necessarily (5). We now derive the additional conditions on the elements of U calculated from (6) imposed by (5). We rewrite (5) as $D(a) = U^{-1}D_r(a)U^*$. Every antiunitary element a can be written as $a = ua_0$ and $D(a)$ as $D(u)D(a_0)$. Taking the pqth element, multiplying by $D^k(ua_0)^*_{mn} = (\Delta^k(u)\beta^k)^*_{m,n}$, using the explicit form of $D_r(a)$, and summing over u, we have

$$\frac{d}{N/2}\sum_u (D(u)D(a_0))_{pq}(\Delta^k(u)\beta_k)^*_{m,n} = U^*_{mp}U^*_{nq} + U^*_{d+m,p}U^*_{d+n,q} + \cdots + U^*_{(l-1)d+m,p}U^*_{(l-1)d+n,q}.$$

We rearrange the left-hand side as

$$\sum_{xy} D(a_0)_{xq}\beta^{k*}_{yn}\left(\frac{d}{N/2}\sum_u D(u)_{px}\Delta^k(u)_{my}\right),$$

and, using (6), we have

$$\sum_{xy} D(a_0)_{xq}\beta^{k*}_{yn}[U^*_{mp}U_{yx} + U^*_{d+m,p}U_{d+y,x} + \cdots + U^*_{(l-1)d+m,p}U_{(l-1)d+y,x}] = U^*_{m,p}U^*_{n,q} + U^*_{d+m,p}U^*_{d+n,q} + \cdots + U^*_{(l-1)d+m,p}U^*_{(l-1)d+n,q}.$$

Multiplying by $U_{bd+m,p}$, summing on p, and using the orthogonality relations of the rows of U gives

$$U_{bd+n,q} = \sum_{xy}\beta^k_{yn}U^*_{bd+y,x}D(a_0)^*_{xq}, \tag{7}$$

where $b = 0, 1, \ldots, l-1$. Relation (7) is the additional condition imposed by (5) on the elements of the matrix U calculated from (6).

TABLE I: Corepresentations of the unitary subgroup C_{3v} and θ.

	E	C_3	C_3^2	δ_v	δ_v^{I}	δ_v^{II}	θ
D^1	1	1	1	1	1	1	$e^{i\varphi}$
D^2	1	1	1	-1	-1	-1	$e^{i\zeta}$
D^3	$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	$\begin{pmatrix}\epsilon & \\ & \epsilon^2\end{pmatrix}$	$\begin{pmatrix}\epsilon^2 & \\ & \epsilon\end{pmatrix}$	$\begin{pmatrix} & 1\\ 1 & \end{pmatrix}$	$\begin{pmatrix} & \epsilon^2\\ \epsilon & \end{pmatrix}$	$\begin{pmatrix} & \epsilon\\ \epsilon^2 & \end{pmatrix}$	$\begin{pmatrix} & e^{i\psi}\\ e^{i\psi} & \end{pmatrix}$
D^4	$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	$\begin{pmatrix}-\epsilon & \\ & -\epsilon^2\end{pmatrix}$	$\begin{pmatrix}\epsilon^2 & \\ & \epsilon\end{pmatrix}$	$\begin{pmatrix} & i\\ i & \end{pmatrix}$	$\begin{pmatrix} & \epsilon^2 i\\ \epsilon i & \end{pmatrix}$	$\begin{pmatrix} & \epsilon i\\ \epsilon^2 i & \end{pmatrix}$	$\begin{pmatrix} & e^{i\alpha}\\ -e^{i\alpha} & \end{pmatrix}$
D^5	$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	$\begin{pmatrix}-1 & \\ & -1\end{pmatrix}$	$\begin{pmatrix}1 & \\ & 1\end{pmatrix}$	$\begin{pmatrix}i & \\ & -i\end{pmatrix}$	$\begin{pmatrix}-i & \\ & i\end{pmatrix}$	$\begin{pmatrix}i & \\ & -i\end{pmatrix}$	$\begin{pmatrix} & -1\\ 1 & \end{pmatrix}$

$\epsilon = e^{i2\pi/3}$

B. Type II corepresentations

We assume that a Type II irreducible corepresentation D^k of dimension $2d$ appears l times in D_γ. From (4), the $2ld$ rows of U associated with these corepresentations are calculated from

$$\frac{d}{N/2}\sum_u D(u)_{pq}\Delta^k(u)^*_{mn} = U^*_{m,p}U_{n,q} + U^*_{d+m,p}U_{d+n,q} + \cdots + U^*_{(2l-1)d+m,p}U_{(2l-1)d+n,q}. \tag{8}$$

From (5) one derives the additional conditions on the elements of U calculated from (8) to be

$$U_{(b+1)d+n,q} = \sum_{xy} \beta^k_{yn} U^*_{bd+y,x} D(a_0)^*_{xq}$$

for $b = 0, 2, 4, \ldots, 2l-2$.

C. Type III corepresentations

We assume that a Type III irreducible corepresentation D^k of dimension $2d$ appears l times in D_γ. From (4), the $2ld$ rows of U associated with these corepresentations are calculated from

$$\frac{d}{N/2}\sum_u D(u)_{pq}\Delta^k(u)^*_{mn} = U^*_{m,p}U_{n,q} + U^*_{2d+m,p}U_{2d+n,q} + \cdots + U^*_{(2l-2)d+m,p}U_{(2l-2)d+n,q} \tag{9a}$$

and

$$\frac{d}{N/2}\sum_u D(u)_{pq}\Delta^k(a_0^{-1}ua_0)_{mn} = U^*_{d+m,p}U_{d+n,q} + U^*_{3d+m,p}U_{3d+n,q} + \cdots + U^*_{(2l-1)d+m,p}U_{(2l-1)d+n,q} \tag{9b}$$

From (5) one derives the additional conditions imposed on the elements of U calculated from (9a), (9b) to be

$$U_{bd+n,q} = \sum_x U^*_{(b+1)d+n,x} D(a_0)^*_{xq}, \tag{10}$$

where $b = 0, 2, 4, \ldots, 2l-2$.

IV. EXAMPLE: C_{3v} WITH TIME REVERSAL

We calculate the matrix U which transforms into reduced form the direct product $D^3 \times D^3 \times D^5$ of irreducible corepresentations of the nonunitary group $H = C_{3v} + C_{3v}\theta$, where $a_0 = \theta$, i.e., time reversal. The irreducible corepresentations of this nonunitary group are given in Table I. The corepresentations are all of Type I with the exception of D^5 which is of Type III.

Using relation (2), one finds that the reduced form contains the irreducible corepresentations D^4 and D^5 each two times, i.e., $D^3 \times D^3 \times D^5 = D^4 + D^4 + D^5 + D^5$. To calculate U, we first use (4). Specifically, to calculate the first four rows of U corresponding to the two $D^4(u)$ appearing on the right-hand side of (4), we use relation (6), and for the last four rows, corresponding to the two $D^5(u)$, we use relations (9a), (9b). The matrix U so derived is

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & b_{12}e^{i\Delta} & b_{11}e^{i\delta} \\
b_{12}e^{i\Delta} & -b_{11}e^{i\delta} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & b_{22}e^{i\Delta} & b_{21}e^{i\delta} \\
b_{22}e^{i\Delta} & -b_{21}e^{i\delta} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & (c_{11}/\sqrt{2})e^{i\tau} & (c_{12}/\sqrt{2})e^{i\Omega} & (c_{11}/\sqrt{2})e^{i\tau} & -(c_{12}/\sqrt{2})e^{i\Omega} & 0 & 0 \\
0 & 0 & (d_{11}/\sqrt{2})e^{i\gamma} & (d_{12}/\sqrt{2})e^{i\nu} & -(d_{11}/\sqrt{2})e^{i\gamma} & (d_{12}/\sqrt{2})e^{i\nu} & 0 & 0 \\
0 & 0 & (c_{21}/\sqrt{2})e^{i\tau} & (c_{22}/\sqrt{2})e^{i\Omega} & (c_{21}/\sqrt{2})e^{i\tau} & -(c_{22}/\sqrt{2})e^{i\Omega} & 0 & 0 \\
0 & 0 & (d_{21}/\sqrt{2})e^{i\gamma} & (d_{22}/\sqrt{2})e^{i\nu} & -(d_{21}/\sqrt{2})e^{i\gamma} & (d_{22}/\sqrt{2})e^{i\nu} & 0 & 0
\end{pmatrix},$$

where $\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}$, $\begin{pmatrix}c_{11} & c_{12}\\ c_{21} & c_{22}\end{pmatrix}$, and $\begin{pmatrix}d_{11} & d_{12}\\ d_{21} & d_{22}\end{pmatrix}$, are unitary matrices,[8] and $\Delta, \delta, \tau, \Omega, \gamma$, and ν are arbitrary phases.

The additional conditions imposed by (5) due to the antiunitary symmetry for the rows of the matrix U corresponding to the corepresentations D^4 are derived from relation (7), and for the rows corresponding to the D^5 from relation (10). From (7) we derive the additional conditions

$$b_{j2}e^{i\Delta} = b_{j1}^* e^{i(\alpha-2\psi-\delta)}, \tag{11}$$

where $j = 1, 2$. From (10) we derive the additional conditions

$$c_{j1}e^{i\tau} = d_{j2}^* e^{-i(2\psi+\nu)},$$

$$c_{j2}e^{i\Omega} = d_{j1}^* e^{-i(2\psi+\gamma)}, \tag{12}$$

where $j = 1, 2$. In addition, from the unitarity of the matrix $\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ and (11), one derives that $b_{11} = 1/\sqrt{2}\, e^{i\xi}$ and $b_{12} = \pm i b_{11}$, where ξ is an arbitrary phase factor. By using conditions (11) and (12), and writing $\beta = \xi + \delta$, the matrix U takes the form

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & e^{i(\alpha-2\psi-\beta)} & -e^{i\beta} \\ e^{i(\alpha-2\psi-\beta)} & -e^{i\beta} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mp ie^{i(\alpha-2\psi-\beta)} & \pm ie^{i\beta} \\ \mp ie^{i(\alpha-2\psi-\beta)} & \mp ie^{i\beta} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_{12}^*e^{-i(2\psi+\nu)} & d_{11}^*e^{-i(2\psi+\gamma)} & d_{12}^*e^{-i(2\psi+\nu)} & -d_{11}^*e^{-i(2\psi+\gamma)} & 0 & 0 \\ 0 & 0 & d_{11}e^{i\gamma} & d_{12}e^{i\nu} & -d_{11}e^{i\gamma} & d_{12}e^{i\nu} & 0 & 0 \\ 0 & 0 & d_{22}^*e^{-i(2\psi+\nu)} & d_{21}^*e^{-i(2\psi+\gamma)} & d_{22}^*e^{-i(2\psi+\nu)} & -d_{21}^*e^{-i(2\psi+\gamma)} & 0 & 0 \\ 0 & 0 & d_{21}e^{i\gamma} & d_{22}e^{i\nu} & -d_{21}e^{i\gamma} & d_{22}e^{i\nu} & 0 & 0 \end{pmatrix},$$

where $\begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$ is a unitary matrix and β, ν, and γ are arbitrary phases.

By using the additional conditions imposed by (5), the ambiguity of the matrix U calculated from (4) has been greatly reduced. From three two-dimensional unitary matrices and six arbitrary phases, we have now only a single unitary matrix and three arbitrary phases.

ACKNOWLEDGMENT

The authors would like to thank Professor J. Zak for helpful discussions.

1. E. P. Wigner, *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
2. J. O. Dimmock and R. G. Wheeler, *The Mathematics of Physics and Chemistry* (Van Nostrand, New York, 1964), Vol. 2, Chap. 12.
3. J. O. Dimmock and R. G. Wheeler, J. Phys. Chem. Solids 23, 729 (1962).
4. J. O. Dimmock and R. G. Wheeler, Phys. Rev. **127**, 391 (1962).
5. J. O. Dimmock, J. Math. Phys. **4**, 1307 (1963).
6. C. J. Bradley and B. L. Davis, Rev. Mod. Phys. **40**, 359 (1968).
7. C. Eckart, Rev. Mod. Phys. **2**, 305 (1930).
8. G. F. Koster, Phys. Rev. **109**, 227 (1958).
9. A. Aviran and J. Zak, J. Math. Phys. **9**, 2138 (1968).
10. A. Aviran, M.Sc. thesis (Technion-Israel Institute of Technology, 1967)
11. G. F. Karavaev, Sov. Phys. Solid State **6**, 2943 (1965).

Some properties of matrices occuring in $SL(2,C)$ invariant wave equations

M. Lorente*, P. L. Huddleston, and P. Roman

Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 15 December 1972; revised manuscript received 4 May 1973)

A general expression is found for the anticommutator of the L^μ matrices that occur in $SL(2,C)$ invariant wave equations. The formula is valid for a broad class of representations which includes, besides the Dirac and the Majorana representations, many other infinite and finite dimensional representations. Several related properties of the L^μ are studied and some applications are given.

1. INTRODUCTION

As the fruit of pioneering work by Dirac, Kemmer, Bhabha, Majorana, Gelfand, Yaglom, and many others, today we possess a deep understanding of the algebraic structure of $SL(2,C)$ invariant wave equations. The salient features of the theory are beautifully summarized in Chap. 11 of Ref. 1. An equation of the form[2]

$$(L^\mu p_\mu + \kappa)\psi(p) = 0,$$

where the L^μ and κ are finite–or infinite–dimensional matrices,[3] is found to be covariant under the $SL(2,C)$ group provided

$$[J^{\mu\sigma}, L^\nu] = i(g^{\sigma\nu}L^\mu - g^{\mu\nu}L^\sigma) \tag{1.1}$$

and also $[J^{\mu\sigma},\kappa] = 0$. Here the $J^{\mu\sigma}$ denote the generators of $SL(2,C)$ which obey the familiar Lie algebra.

Condition (1.1) implies[4] that the irreducible components which are contained in the $SL(2,C)$ representation space to which the wave function ψ belongs, must be interlocking. Let us recall that two irreducible representations[5] $\tau = (l_0, l_1)$ and $\tau' = (l_0', l_1')$ of $SL(2,C)$ are said to be interlocking if either

$$(l_0', l_1') = (l_0 \pm 1, l_1) \tag{1.2a}$$

or

$$(l_0', l_1') = (l_0, l_1 \pm 1). \tag{1.2b}$$

If one now introduces the additional requirement[6] that *the wave function space should be precisely a* (generally speaking reducible) *representation of the $SL(2,C)$ group* [i.e., that the set of basis functions should be completely labeled by the $SL(2,C)$ canonical labels], then a new feature emerges: the Lie algebra of the $J^{\mu\sigma}$ and L^μ operators must close and becomes exactly that of $Sp(4,R)$. This means that in addition to (1.1) and the familiar $[J,J] \sim J$ commutators, we also have[7,8]

$$[L^\mu, L^\sigma] = -iJ^{\mu\sigma}. \tag{1.3}$$

We note that, denoting the generators of $Sp(4,R)$ by J^{ab} $(a,b = 0,1,2,3,4)$ and setting

$$L^\mu = J^{4\mu}, \tag{1.4a}$$

Eqs. (1.1), (1.3), and the $SL(2,C)$ commutators can be condensed to read

$$[J^{ab}, J^{cd}] = i(g^{bc}J^{ad} - g^{ac}J^{bd} - g^{bd}J^{ac} + g^{ad}J^{bc}), \tag{1.4b}$$

where $g^{44} = g^{00} = -g^{kk} = 1$, $g^{ab} = 0$ for $a \neq b$. The major consequence is that $J^{ab}J_{ab}$ is an invariant, i.e., it is a *multiple of the identity*. [To see this, we must also recall that, by assumption, the components of ψ are uniquely labeled by $SL(2,C)$ indices, so that the $Sp(4,R)$ representation must be irreducible.]

In this paper we concentrate only on the situation when the said requirement on the ψ space is met, so that Eq. (1.3) is valid and $J_{ab}J^{ab}$ is a multiple of the identity. In passing, we note that, except for a few cases (such as the Dirac, Kemmer, 20-component Bhabha equations), Eq. (1.3) is incompatible with $(p^2 - \kappa^2)\psi = 0$, so that, in general, we will have a mass spectrum.

In all physical applications the L^μ operators play a crucial role. It is, therefore, most desirable to establish as many algebraic properties for them as possible. Unfortunately, apart from the fundamental relations (1.1) and (1.3), no general statements can be made. Of course, it is true that (Cf. Ref. 1) for any given definite choice of interlocking representations, the L^μ can be calculated and specific cases treated. Thus, for example, one finds for the Dirac case [corresponding to the pair $(1/2, 3/2) \leftrightarrow (-1/2, 3/2)$] the familiar Dirac algebra[9] $\{L^\mu, L^\nu\} = (1/2)g^{\mu\nu}$; for the Kemmer case one obtains the Kemmer-Duffin algebra; and for the infinite dimensional Majorana case one finds[10]

$$(L^\mu p_\mu)^2 = (1/4)p^2 - W^2,$$

where W_ν is the Pauli–Lubanski vector [so that for $p^2 > 0$, one has $(L^\mu p_\mu)^2 = p^2(l + 1/2)^2$].

The primary purpose of this paper is to derive, for a broad and important class of representations, *a general relation for the anticommutator* $\{L^\mu, L^\nu\}$. The class of representations for which our result holds will be called the class of strongly interlocking representations and is defined as follows:

Definition: A (reducible) representation R of $SL(2,C)$ is called strongly interlocking if (a) *it consists of interlocking irreducible representations and* (b) *for each participating irreducible component the value of $l_0^2 + l_1^2$ is the same.*

Observe that condition (b) implies that all irreducible components of R have the same first Casimir invariant $l_0^2 + l_1^2 - 1$, hence $(1/2)J_{\mu\nu}J^{\mu\nu}$ *is represented in R by a multiple of the identity.* Furthermore, from Eqs. (1.2a, b) it follows that in strongly interlocking representation there must exist either a component $(1/2, l_1)$(l_1 arbitrary) which then interlocks via the scheme (1.2a), or there must exist a component $(l_0, 1/2)$(l_0 arbitrary) which then interlocks via scheme (1.2b). If we make the additional restriction that the strongly interlocking R contains only one pair of irreducible components, then the following representations qualify:

$$(1/2, n/2) \leftrightarrow (-1/2, n/2) \quad (n \geq 3, \text{odd integer}), \tag{1.5a}$$

$$(1/2, l_1) \leftrightarrow (-1/2, l_1) \quad (l_1 \text{ arbitrary complex}), \tag{1.5b}$$

Copyright © 1973 by the American Institute of Physics

$$(l_0, 1/2) \leftrightarrow (l_0, -1/2) \quad (l_0 \text{ integral or half-integral}), \tag{1.5c}$$

$$(0, 1/2) \leftrightarrow (0, -1/2) \tag{1.5d}$$

$$(1/2, 0) \leftrightarrow (-1/2, 0). \tag{1.5e}$$

Here (1.5a) describes a finite-dimensional representation with spin content $l = 1/2, 3/2 \ldots (n-2)/2$. Case (1.5b) and (1.5c) are infinite-dimensional representations [used for describing particle towers with positive definite energy (l_1 imaginary) or positive definite charge, respectively]. Finally, (1.5d), (1.5e) are the celebrated infinite-dimensional Majorana representations for integral (half-integral) spin towers of particles (which have both positive definite charge and energy).[11]

If we do not impose (the unnecessary) restriction that R contains only one pair of components, then additional strongly interlocking representations can be found, for example,[12]

$$(1/2, 1/2) \leftrightarrow (1/2, -1/2) \leftrightarrow (-1/2, -1/2) \tag{1.5f}$$

or

$$\begin{array}{ccc} (1/2, 1/2) & \leftrightarrow & (1/2, -1/2), \\ \updownarrow & & \updownarrow \\ (-1/2, 1/2) & \leftrightarrow & (-1/2), -1/2). \end{array} \tag{1.5g}$$

We are now prepared to formulate our main result:

Theorem: In any strongly interlocking representation,

$$\{L^\mu, L^\nu\} = (2/3) g^{\mu\nu} J_{\sigma\tau} J^{\sigma\tau} - g_{\sigma\tau}\{J^{\mu\sigma}, J^{\nu\tau}\}, \tag{1.6}$$

or alternatively[13]

$$\{L^\mu, L^\nu\} = (4/3) g^{\mu\nu}(l_0^2 + l_1^2 - 1) - g_{\sigma\tau}\{J^{\mu\sigma}, J^{\nu\tau}\}, \tag{1.6'}$$

where l_0, l_1 are the labels of any participating irreducible component.

The proof of this theorem crucially depends on the following:

Lemma: In any strongly interlocking representation $L_\rho L^\rho$ is a multiple of the identity.

Proof of Lemma: We write the $Sp(4,R)$ invariant in the form $(\alpha, \beta = 0, 1, 2, 3)$

$$\begin{aligned} \text{inv} &= J_{ab} J^{ab} = J_{\alpha b} J^{\alpha b} + J_{4b} J^{4b} \\ &= J_{\alpha\beta} J^{\alpha\beta} + J_{\alpha 4} J^{\alpha 4} + J_{4\beta} J^{4\beta}. \end{aligned}$$

In view of (1.4a), the second and third terms are precisely $L_\rho L^\rho$. Since, in a strongly interlocking representation, $J_{\alpha\beta} J^{\alpha\beta}$ is an invariant (i.e., a multiple of the identity), it follows that $L_\rho L^\rho$ has the same property. QED

The obvious importance of our lemma is that, for strongly interlocking representations, the scalar $L_\rho L^\rho$ commutes not only with $J_{\mu\nu}$; it also *commutes with all* L_μ.

An immediate, useful consequence of this is that

$$L^\mu J_{\nu\mu} + J_{\nu\mu} L^\mu = 0, \tag{1.7}$$

which relation follows from (1.3) and from $[L^\rho L_\rho, L^\mu] = 0$. With the help of (1.1) this relation yields

$$J_{\nu\mu} L^\mu = (3/2) i L_\nu \quad \text{and} \quad L^\mu J_{\nu\mu} = -(3/2) i L_\nu. \tag{1.8}$$

2. PROOF OF MAIN THEOREM

Inserting (1.3) into (1.1) and then multiplying from the right (left) by L^τ, we get the two equations

$$\begin{aligned} L^\mu L^\sigma L^\nu L^\tau - L^\sigma L^\mu L^\nu L^\tau - L^\nu L^\mu L^\sigma L^\tau + L^\nu L^\sigma L^\mu L^\tau \\ = g^{\sigma\nu} L^\mu L^\tau - g^{\mu\nu} L^\sigma L^\tau, \end{aligned}$$

$$\begin{aligned} L^\tau L^\mu L^\sigma L^\nu - L^\tau L^\sigma L^\mu L^\nu - L^\tau L^\nu L^\mu L^\sigma + L^\tau L^\nu L^\sigma L^\mu \\ = g^{\sigma\nu} L^\tau L^\mu - g^{\mu\nu} L^\tau L^\sigma. \end{aligned}$$

We contract both equations with $g_{\sigma\tau}$ and use the fact that $L_\rho L^\rho$ commutes with all L^σ. In this manner (using also the symmetry of $g_{\sigma\tau}$) we obtain, upon adding the two resulting equations,

$$g_{\sigma\tau}(J^{\mu\sigma} J^{\nu\tau} + J^{\nu\tau} J^{\mu\sigma}) = g_{\sigma\tau} g^{\sigma\nu}\{L^\mu, L^\tau\} - g^{\mu\nu} g_{\sigma\tau}\{L^\sigma, L^\tau\},$$

where, on the left-hand side, we also used Eq. (1.3). Elementary manipulation gives

$$-g_{\sigma\tau}\{J^{\mu\sigma}, J^{\nu\tau}\} = \{L^\mu, L^\nu\} - 2g^{\mu\nu} L_\sigma L^\sigma. \tag{2.1}$$

Contracting with $g_{\mu\nu}$ we obtain

$$L_\mu L^\mu = (1/3) J_{\sigma\tau} J^{\sigma\tau} = (2/3)(l_0^2 + l_1^2 - 1). \tag{2.2}$$

Combining (2.2) and (2.1) we finally find precisely Eq. (1.6). QED

In passing, we point out that Eq. (2.2) itself is very useful: it fixes the numerical value of the invariant $L_\mu L^\mu$. For example, in the Dirac representation[9] ($L_\mu = 1/2\gamma_\mu$), it gives the familiar value $\gamma_\mu \gamma^\mu = 4$, and in the Majorana representations ($L_\mu = \Gamma_\mu$), it gives [14] $\Gamma_\mu \Gamma^\mu = -1/2$.

3. SOME APPLICATIONS

A. Relation between p^2 and W^2; the mass spectrum

It is clear that the generic result (1.6a) will reproduce particular formulas that have been previously obtained by elaborate *ad hoc* calculations. We already pointed out at the end of Sec. 2 that we now have an easy method to evaluate $L_\mu L^\mu$.

Similarly, from our theorem we can calculate the expression $(L \cdot p)^2 \equiv (L^\mu p_\mu)(L^\nu p_\nu)$ without knowing any explicit representation for the L^μ matrices. Using (1.6) and (1.3) and noting that by symmetry $J^{\mu\nu} p_\mu p_\nu = 0$, we get

$$(Lp)^2 = L^\mu L^\nu p_\mu p_\nu = (1/3) J_{\rho\sigma} J^{\rho\sigma} p^2 - J^{\mu\rho} J_{\nu\rho} p_\mu p^\nu.$$

Recalling that

$$W^2 = -(1/2) J_{\rho\sigma} J^{\rho\sigma} p^2 + J^{\mu\rho} J_{\nu\rho} p^\nu,$$

we can write more conveniently,

$$\begin{aligned} (Lp)^2 &= -(1/6) J_{\rho\sigma} J^{\rho\sigma} p^2 - W^2 \\ &= -(1/3)(l_0^2 + l_1^2 - 1) p^2 - W^2. \end{aligned} \tag{3.1}$$

This is the *generalization* of the *Takabayashi–Stoyanov–Todorov formula* which holds in every strongly interlocking representation. In particular, for the Majorana representations $l_0^2 + l_1^2 - 1 = -3/4$, so that we get back the specific result of Ref. 10.

The primary importance of (3.1) is that it allows the determination of the *mass spectrum* in any strongly interlocking representation. Applying $(L_\nu p^\nu - \kappa)$ onto the wave equation and using (3.1) [assuming $p^2 \equiv m^2 > 0$, so that $W^2\psi = -m^2 l(l+1)$], we have

$$m^2 = \kappa^2/[l(l+1) - (1/3)(l_0^2 + l_1^2 - 1)]. \qquad (3.1')$$

In passing we note that this result does not necessarily mean a descending spectrum: if, for example, $\kappa^2 = (ap^2 + b)^2$ (cf. footnote 3), solution of (3.1′) for m gives an ascending branch.

B. Properties of L_0

From (1.6) we obtain

$$(L_0)^2 = (1/3)J_{\sigma\tau}J^{\sigma\tau} + J_{0k}J_{0k}.$$

Introducing the notations

$$\mathbf{J} = (J_{23}, J_{31}, J_{12}), \qquad \mathbf{N} = (J_{01}, J_{02}, J_{03}) \qquad (3.2)$$

and noting that $(1/2)J_{\rho\sigma}J^{\rho\sigma} = \mathbf{J}^2 - \mathbf{N}^2$, we obtain

$$(L_0)^2 = -(1/3)(l_0^2 + l_1^2 - 1) + l(l+1), \qquad (3.3)$$

where $l(l+1)$ are the eigenvalues of $\mathbf{J}^2$. Thus, we have the interesting result that for any strongly interlocking representation L_0^2 *is diagonal* in the canonical basis, although it is not a multiple of the identity (except for the Dirac representation which has only one value for l).

Moreover, we may obtain information on *the spectrum of* L_0, and observe that the magnitudes of its eigenvalues are given by[15]

$$|L_0'| = [-(1/3)(l_0^2 + l_1^2 - 1) + l(l+1)]^{1/2}. \qquad (3.4)$$

We remind the reader that the knowledge of L_0 or its spectrum is needed in the calculation of the charge operator or of the parity operator.[16]

C. Relations between tensor operators

The irreducible parts of $L^\mu L^\nu$ can be isolated and we have

$$L^\mu L^\nu = 1/2[L^\mu, L^\nu] + (1/2\{L^\mu, L^\nu\} - 1/4 g^{\mu\nu} L^\rho L_\rho) + (1/4)g^{\mu\nu} L^\rho L_\rho. \qquad (3.5)$$

The first term is the antisymmetrical tensor $(-i/2)J^{\mu\nu}$. The last term is the scalar whose explicit form is given by (2.2). The middle term can be rewritten with (1.6) and (2.2) to yield the traceless symmetrical tensor

$$\begin{aligned} S^{\mu\nu} &= (1/4)g^{\mu\nu} J_{\sigma\tau}J^{\sigma\tau} - (1/2)g_{\sigma\tau}\{J^{\mu\sigma}, J^{\nu\tau}\} \\ &= (1/2)(l_0^2 + l_1^2 - 1) - (1/2)g_{\sigma\tau}\{J^{\mu\sigma}, J^{\nu\tau}\}. \end{aligned} \qquad (3.6)$$

It is not difficult to systematically analyze higher order tensor operators in a similar fashion.

D. Bilinear covariants of the extended Lorentz group

The simplest examples of strongly interlocking representations, in particular those listed under (1.5a) through (1.5d), have the important property that they also furnish a representation of the extended Lorentz group. Indeed, these representations satisfy the criteria[17] discussed in Part II, Sec. 3 of Ref. 1. Thus, a parity operator S, a time reversal operator T, and a total reflection operator $J = TS$ can be constructed in the standard manner as discussed in Ref. 1. Furthermore, the criteria given in Ref. 1 for the existence of an invariant nondegenerate bilinear form of wave functions are also met[18] by our representations (1.5a)–(1.5d). Therefore, we may construct the usual bilinears:

the scalar	(ψ, ψ),
the pseudoscalar	$(T\psi, \psi)$,
the vector	$(L^\mu\psi, \psi)$,
the pseudovector	$(TL^\mu\psi, \psi)$,
the tensor	$(L^\mu L^\nu\psi, \psi)$,
the pseudotensor	$(TL^\mu L^\nu\psi, \psi)$,

and so on. Finally, for finding the irreducible parts of the last two quantities (or of higher order tensors) we can use our main theorem.

*On leave of absence from the Universidad de Madrid. Supported in part by a grant from G.I.F.T.

[1] I. M. Gelfand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963).

[2] We use the metric $g_{00} = -g_{kk} = 1$, $g_{\mu\nu} = 0$ for $\mu \neq \nu$. Summation over dummy indices is always understood.

[3] The κ need not be independent of p. For example, in the Nambu–Fronsdal equation $\kappa = ap^2 + b$, where a and b are real numbers. In most familiar cases (like the Dirac or Majorana equation) κ is simply a multiple of the identity operator.

[4] Here we also imply that $\kappa \neq 0$.

[5] We use the notation of Ref. 1. In terms of l_0 and l_1 the two Casimir invariants of $SL(2, C)$ are given as $(1/2)J_{\mu\nu}J^{\mu\nu} = l_0^2 + l_1^2 - 1$ and $(1/4)\epsilon_{\mu\nu\rho\sigma}J^{\mu\nu}J^{\rho\sigma} = 2il_0l_1$.

[6] There is no compelling reason to do this. For example, the set of components of ψ may carry a representation of the Poincaré group, or a representation of a semisimple group larger than $SL(2, C)$ [such as $SU(2,2)$], or even a more unusual nonsemisimple group [such as $SL(2, C) \times SO(3,2)$]. For a survey, see L. Castell, Nuovo Cimento **A 50**, 945 (1967).

[7] If the above condition on the ψ space is not required, then the J, L algebra may or may not close and if it closes, we may have commutators $[L, L]$ different from (1.3). For details, see L. Castell's paper in Ref. 6.

[8] Equation (1.3) is not given in Ref. 1 but, apart from general arguments, it can be also directly verified if one uses the explicit construction of the L^μ as given in Ref. 1. On the other hand, the remarkable book by E. M. Corson [*Introduction to Tensors, Spinors, and Relativistic Wave-Equations* (Blackie, London, 1954)] pays special attention to the closing relation (1.3) (cf. in particular, Sec. 36). This book is also highly recommended for a survey of the older literature. See also the article by L. O. Raifeartaigh, in *Battelles Rencontres, 1970* (Benjamin, New York, 1971).

[9] The customary Dirac matrices γ^μ are related to the L^μ by $L^\mu = (1/2)\gamma^\mu$. This means that the mass constant in the standard Dirac equation is one-half of the κ that occurs in the generic equation.

[10] Apparently, this relation was first derived by T. Takabayashi, cf. *Proceedings of the 1967 International Conference on Particles and Fields*, (Interscience, New York, 1967) p. 413, and by D. Tz. Stoyanov and I. T. Todorov, J. Math. Phys. 9, 2146 (1968). The corresponding relation in the Dirac case is $(L^\mu p_\mu)^2 = (1/4)p^2$.

[11] The two Majorana representations are the only possible self-interlocking representations. Indeed, $(0, -1/2)$ is equivalent to $(0, 1/2)$ and $(-1/2, 0)$ is equivalent to $(1/2, 0)$. The strongly interlocking criterion is trivially satisfied.

[12] These representations are infinite dimensional and nonunitary. We note here that the representations (1.5c) are also nonunitary, and so are those belonging to class (1.5b) unless l_1 is pure imaginary. Naturally, the finite dimensional representations (1.5a) are nonunitary.

[13] Recall that $(1/2)J_{\sigma\tau}J^{\sigma\tau}$ is a multiple of the identity in a strongly interlocking representation.

[14] This result has been first obtained with the laborious use of an explicit representation, cf. Ref. 10.

[15] For example, in the Majorana representation we obtain the familiar result $|L'_0| = l + (1/2)$.

[16] It is well known that, in the general case of an $SL(2, C)$ representation, the parity operator is proportional to $\exp(i\pi L_0)$. See, for example, p. 152 of Corson, Ref. 8.

[17] For Case (1.5b), this is true only if we take $l_1 \neq 0$.

[18] For Case (1.5b), this is true only if l_1 is either pure real or or pure imaginary.

Thin-sandwich conjecture and a fifth initial value equation for the gravitational field

C. Martin Pereira

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

(Received 5 February 1973; revised manuscript received 17 April 1973)

The thin-sandwich conjecture for Einstein's field equations of general relativity is discussed. Three of the four initial-value equations of general relativity are considered. This system of three equations is shown to have an integrability condition when the spatial metric, its time rate of change, and the lapse function N are given. The nature of this integrability condition is discussed in detail. It is also noted that this integrability condition can be expected to play a significant role in any proof of existence for a thin-sandwich problem in which the spatial metric and its time derivative are chosen arbitrarily. A local proof of existence is given for the first three thin-sandwich equations. The proof allows the spatial metric components, their time rate of change, and the lapse function N to be chosen arbitrarily. No coordinate conditions of any kind are used in the proof.

I. THE THIN-SANDWICH CONJECTURE OF GENERAL RELATIVITY

The thin-sandwich conjecture of general relativity was first postualted in 1960. Since that time, this conjecture has been examined by a number of authors,[1-3] but no proof of the conjecture in its original form has been obtained. The conjecture states that given the spatial metric and its time rate of change on a spacelike 3-surface, there always exists a solution to Einstein's field equations. [It is assumed here that in addition the energy-momentum tensor $T^{\mu\nu}(x)$ is given throughout all space–time.] A proof of the conjecture in its original form would show (a) that there does exist for the gravitational field a natural set of field coordinates and corresponding field velocities which can be chosen arbitrarily on a space-like 3-surface, and (b) that the field coordinates may be chosen to consist of the spatial metric components on a spacelike 3-surface.

The purpose of the present paper is to point out an important theoretical feature of the original thin-sandwich problem which has heretofore been overlooked. A new thin-sandwich equation has been found. This equation is an integrability condition associated with the first three thin-sandwich equations. It plays the same role with respect to the first three thin-sandwich equations as is played by the energy–momentum conservation law with respect to Einstein's equations. Furthermore, the fact that an integrability condition is associated with the original thin-sandwich problem means that a direct application of the familiar Cauchy–Kowalewski existence theorem is inadequate for a discussion of this problem. A completely satisfactory proof of existence requires the more general existence theorems of Riquier[4-7] or Cartan.

Several modified forms of the thin-sandwich conjecture have been proven recently. It will be useful to compare these results with the results described here. The original conjecture requires that the proof admit 12 specific arbitrary functions, namely the six spatial metric components γ_{ij} and their six time derivatives $\gamma_{ij,4}$. Bruhat's thin-sandwich proof required the imposition of coordinate conditions and therefore did not admit 12 arbitrary functions. The recent proof of Komar[2] avoids any coordinate conditions and therefore admits 12 arbitrary functions. It thus represents a considerable advance over earlier work. Komar uses a canonical transformation to change the thin-sandwich equations into a more convenient form. However, because of this canonical transformation, Komar's proof does not allow the spatial metric and its time rate of change to be chosen arbitrarily. Rather, Komar's arbitrary functions are the quantities $(\gamma p^{-2})\gamma_{ij}$ and their time derivatives. (Here p is the trace of the canonical momentum and γ is the determinant of the spatial metric.) Komar's arbitrary quantities become singular when $p = 0$ and this leads to difficulties. Komar has stated that his canonical transformation changes the thin-sandwich equations into an elliptic system of equations. However, this is not the only effect of the transformation. The quantity p contains first derivatives of the shift functions N_i. The consequence of this is that Komar's transformation

$$\gamma_{ij} \rightarrow (\bar{\gamma})^{-1/2}\,|\bar{p}|\,\bar{\gamma}_{ij}$$

introduces new second derivatives of the N_i into the three thin-sandwich equations

$$p^{ij}{}_{,j} = 0.$$

As a result, the structure of these equations is changed to such an extent that *they no longer possess an integrability condition.*

Bergmann[3] has noted that Komar's transformation can be greatly generalized. Bergmann shows that elliptic equations will result for a very wide choice of arbitrary functions. Bergmann's arbitrary functions are the five conformal metric components $\gamma^{-1/3}\gamma_{ij}$, their five time derivatives, and the one quantity $\bar{\gamma} = F(\gamma, p)$ and its time derivative. Here $F(\gamma, p)$ is unrestricted except for the requirement $\partial F/\partial p \neq 0$. In addition to the production of elliptic equations, Bergmann's transformations also change the equations to new equations which possess no integrability conditions. The reason for this is the same as for Komar's transformation.

York has also advocated the use of the conformal metric in the initial value problem.[8] Working independently, York[9] was also led to a treatment of the thin-sandwich problem in which the five conformal metric components and their time derivatives are chosen arbitrarily. York chooses as the last two arbitrary functions the quantity $p\gamma^{-1/2}$ and its time derivative. York's procedure also produces elliptic equations and removes the integrablity condition. (The reason for this is again the same as for Komar's procedure.) York then gives a detailed discussion of the Dirichlet problem for his elliptic system of equations.

The interested reader should examine carefully the effect of Komar's transformation by applying the transformation to Eqs. 7-3.15a and 7-3.15b of the article by

Copyright © 1973 by the American Institute of Physics

Arnowitt, Deser, and Misner in the book edited by L. Witten (1962) (see Ref. 1.) One more point should be understood when comparing the present paper with the discussion of Komar. In Komar's paper[2] the system of equations considered consists of ten equations, the six equations defining p_{ij} in terms of $\gamma_{ij,4}$ plus the four initial value equations written in terms of the p_{ij}. Komar's transformation allows one to write these ten equations as five second-order differential equations plus five equations defining the traceless part of the quantity p_{ij}. (See Komar, Ref. 2, Eq. 3.10). In the present paper, on the other hand, one is to imagine that one has already substituted for the p_{ij} to obtain four equations that are second order in the N^i. Also, instead of the p_{ij}, I use a closely related quantity Q_{ij}.

One more point should be made here. Komar[2] and Bergmann[3] prove formal ellipticity for their modified thin-sandwich problems, but they do not attempt an explicit proof of uniqueness and stability.

This ends the discussion of the work of previous authors on various modified thin-sandwich theorems. In the remainder of this paper the term "thin-sandwich problem" will be used to refer to the original thin-sandwich problem in which the γ_{ij} and their time derivatives are chosen arbitrarily.

II. DERIVATION OF A FIFTH INITIAL VALUE EQUATION FOR THE GRAVITATIONAL FIELD

The thin-sandwich problem of general relativity consists of the four equations

$$[N^{-1}(Q^{ij} - \gamma^{ij}Q^m{}_m)]_{;j} = 8\pi N^{-1}S^i, \tag{1a}$$

$$(Q^i{}_i)^2 - Q^j{}_iQ^i{}_j - (N)^2\bar{R} = 16\pi(N)^2T^4{}_4, \tag{1b}$$

where

$$2Q_{ij} = \gamma_{ij,4} - N_{j;i} - N_{i;j} \tag{2}$$

and

$$S^i = \gamma^{ki}(N)^2T^4{}_k \quad \text{with} \quad i = 1,2,3,\ \mu = 1,\ldots,4. \tag{3}$$

Here γ_{ij} is the spatial metric tensor and the $T^4{}_\mu$ are components of the energy–momentum tensor $T^\nu{}_\mu$. The quantity Q^{ij} is proportional to the extrinsic curvature of the three-surface.[10] If one writes $g_{\mu\nu}$ for the metric of space–time, then

$$\gamma_{ij} \equiv g_{ij}, \qquad \gamma^{ij}\gamma_{jk} \equiv \delta^i{}_k, \tag{4a}$$

$$\gamma^{ij} = g^{ij} - g^{i4}g^{j4}(g^{44})^{-1}, \tag{4b}$$

$$N_i \equiv g_{4i}, \qquad N \equiv (-g^{44})^{-1/2}. \tag{4c}$$

Note that covariant differentiation is with respect to γ_{ij}. Raising and lowering of indices is done using γ_{ij}. Our notation for the spatial metric is taken from the discussion of York.[8-11] We denote the spatial Ricci tensor formed from the γ_{ij} by $\bar{R}^k{}_m$ and the spatial Ricci scalar by $\bar{R}$.

Equations (1) are the initial value equations of the gravitational field.[11] A precise definition of the thin-sandwich problem is as follows.

Thin-sandwich problem: Show that for every set of given functions γ_{ij} and $\gamma_{ij,4}$ there always exists a corresponding set of functions N_i and N such that Eqs. (1) are satisfied.

The first three initial value equations, Eqs. (1a), can be written out explicitly as follows:

$$N^{[n;j]}{}_{;j} - N^{-1}N_{;j}[N^{(n;j)} - \gamma^{nj}\gamma_{im}N^{(i;m)} - e^{nj}] - \bar{R}^{nj}N_j - e^{nj}{}_{;j} = -8\pi S^n \tag{5}$$

where we have made use of identity

$$\gamma^{nm}[N_{n;k;m} - N_{n;m;k}] = -\bar{R}^j{}_kN_j. \tag{6}$$

Also note that

$$N^{[n;j]} \equiv \tfrac{1}{2}(N^{n;j} - N^{j;n}) \tag{7a}$$

and

$$N^{(n;j)} \equiv \tfrac{1}{2}(N^{n;j} + N^{j;n}). \tag{7b}$$

It is now possible to derive a new thin-sandwich equation by taking the divergence of Eq. (5) and then using Eq. (5) to eliminate the remaining second derivatives from the result. This procedure is equivalent to an application to Eqs. (5) of Riquier's general procedure for the derivation of integrability conditions. The resulting equation

$$[\bar{R}^n{}_jN^j]_{;n} + N^{-1}N_{;j;n}[N^{(n;j)} - \gamma^{jn}\gamma_{im}N^{(i;m)} - e^{nj}] + e^{nj}{}_{;j;n} = 8\pi N^{-1}(NS^j)_{;j} \tag{8}$$

is the integrability condition associated with the restricted thin-sandwich problem. Here

$$2e_{nj} = \gamma_{nj,4} - \gamma_{nj}\gamma^{im}\gamma_{im,4}.$$

(For a precise definition of the term "integrability condition", and a discussion of Riquier's procedure, see Refs. 4–6).

Note that Eqs. (5) involve second derivatives of the N^i. However, because of the identity

$$[N^{i;j} - N^{j;i}]_{;j;i} = 0.$$

Equation (8) involves only first derivatives of the N^i. Thus Eq. (8) has the following properties

(a) It is algebraically independent of Eqs. (5) and also independent of Eqs. (1b).

(b) It can be derived from Eqs. (5) by differentiation.

(c) The highest derivative of the N^i contained in Eqs. (8) are first derivatives.

(d) The highest derivatives of N contained in Eq. (8) are second derivatives.

These properties (which can be verified by inspection) are by themselves enough to imply that Eq. (8) is the integrability condition for the system (5).

Note that Eq. (8) does contain third derivatives of the form $\gamma_{ij,4mn}$. But in the thin-sandwich picture, the $\gamma_{ij,4}$ are given functions so the quantities $\gamma_{ij,4mn}$ may be thought of as given beforehand.

It appears that Eq. (8) is being published here for the first time.[12] In terms of the Q_{ij}, the identity may be written

$$[Q^{ij} - \gamma^{ij}Q^m{}_m]_{;j;i} - N^{-1}N_{;i;j}[Q^{ij} - \gamma^{ij}Q^m{}_m] = 8\pi N^{-1}[NS^i]_{;i}, \quad (9)$$

where the Q_{ij} satisfy Eq. (2).

The significance of Eq. (8) may be summarized by a comparison with Einstein's equations. Einstein's equations

$$G^{\mu\nu} = -8\pi T^{\mu\nu} \quad (10)$$

imply the energy–momentum conservation law

$$T^{\mu\nu}{}_{;\nu} = 0. \quad (11)$$

Equation (11) is the integrability condition corresponding to Eqs. (10) when the quantities $T^{\mu\nu}(x)$ are given.[5,6] Equation (8) is the integrability condition associated with the restricted thin-sandwich conjucture. Thus Eq. (8) plays the same role with respect to Eqs. (1a) as is played by the energy–momentum conservation law with respect to Einstein's equations.

Equation (8) is algebraically independent of Eqs. (1). Furthermore it is unique in the same sense that the energy–momentum conservation law is unique. It thus deserves full status as a fifth initial value equation.

Once Eq. (8) has been derived, it becomes possible to prove the restricted thin-sandwich theorem as follows. Replace the system (1a) by the equivalent system

$$[N^{-1}(Q^{Aj} - \gamma^{Aj}Q^m{}_m)]_{;j} = 8\pi N^{-1}S^A, \qquad \text{first members } N^A{}_{,33} \quad (12a)$$

$$\{[N^{-1}(Q^{3j} - \gamma^{3j}Q^m{}_m)]_{;j}\}_{x^3=0} = 8\pi[N^{-1}S^3]_{x^3=0}, \qquad N^2{}_{,32} \quad (12b)$$

$$[Q^{ij} - \gamma^{ij}Q^m{}_m]_{;j;i} - N^{-1}N_{;i;j}(Q^{ij} - \gamma^{ij}Q^m{}_m) = 8\pi N^{-1}(NS^i)_{;i}, \qquad N^3{}_{,3} \quad (12c)$$

where the Q^{ij} and the S^i are defined in Eq. (2) and Eq. (3), and $A = 1, 2$, and where it is assumed that $R^3{}_3 \neq 0$. Also note that the initial value 2-surface $x^3 = 0$ is assumed here to be noncompact.

The method used here is analogous to Lichnerowicz's proof of the existence of solutions to Einsteins' equations by the use of the contracted Bianchi identities.[13] Note that Eq. (12b) is an initial value equation on the two surface $x^3 = 0$. The remaining equations (12a) and (12c) satisfy the requirements of the Cauchy-Kowalewski existence theorem. They can be thought of as determining the derivatives shown beside each equation and referred to as first members. An application of this existence theorem to Eqs. (12) now proves the restricted thin-sandwich theorem stated above. (The given functions are assumed to be analytic and existence is proven only locally.) Note that *no coordinate conditions of any kind have been used in the proof.*

The preceding proof of existence has a number of useful consequences. From the form of the first members of Eqs. (12) one can determine all the arbitrary functions in the solution of the restricted thin-sandwich problem. In addition to the thirteen functions of three variables $\gamma_{ij}, \gamma_{ij,4}, N$, there are the four functions of two variables $(N^i)_{x^3=0}$, $(N^1{}_{,3})_{x^3=0}$, and the one function of one variable $(N^2{}_{,3})_{x^3=x^2=0}$.

Note how the existence of the integrability condition (8) has the effect of converting one of the thin-sandwich equations into a constraint on the permissible initial values for the N^i given on an initial 2-surface. This means that there will be some choices of initial conditions on the 2-surface which will be incompatible with local existence of a solution to the thin-sandwich problem. Of course, this does not prevent the existence of local solutions when a correct choice of initial conditions is made.

Furthermore, Eq. (8) is the only consequence of Eqs. (1a) which contains no derivatives of the N^i higher than first derivatives. For if any other such equation existed and was algebraically independent of Eq. (8) it would change the nature of the arbitrary functions described above and thus contradict the proof of existence just completed.

Since it has now been demonstrated that the first three thin-sandwich equations admit the 13 arbitrary[14] functions $\gamma_{ij}, \gamma_{ij,4}, N$ it is natural to ask if the addition of the fourth thin-sandwich equation, Eq. (1b), to the system has the effect of simply removing N from the list so that 12 arbitrary functions remain. A successful proof of this fact would prove the thin-sandwich conjecture locally. In an attempt to answer this question, the same techniques applied here have also been applied to the full system of four thin-sandwich equations (1). One can expect that the application of Riquier's methods will bring a new flexibility to the analysis. One can also expect that the four equations (1) will also have an integrability condition in the case where the spatial metric and its time derivative are given. Furthermore, this integrability condition should be essentially equivalent to Eq. (8). The results of this investigation into the full system of four equations (1) will be published elsewhere.[7]

Finally, it is of interest to note that in both Maxwell's and Einstein's equations, the presence of an integrability condition is associated with the existence of a gauge transformation which transforms one set of solutions of the equations into another. Since Eq. (8) is also an integrability condition, it is therefore natural to ask whether there is also some generalized gauge transformation[15] associated with the thin-sandwich equations. A full investigation of this question is currently in progress.

ACKNOWLEDGMENTS

I wish to thank J. W. York, V. Moncrief, and R. H. Gowdy for helpful remarks. I especially thank K. Kuchař for valuable discussions.

[1]D. H. Sharp, Princeton A.B. senior thesis, Princeton University, 1960 (unpublished); R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. 126, 1864 (1962); Y. Choquet-Bruhat, *Gravitation: An Introduction to Current Research.* edited by L. Witten (Wiley, New York, 1962), Chap. 4; J. A. Wheeler, *Relativity, Groups and Topology,* edited by C. De Witt and B. De Witt (Gordon and Breach, New York, 1964), pp. 345–364; P. G. Bergmann, *Relativity,* edited by M. Carmeli, S. Fickler, and L. Witten (Plenum, New York, 1970), Chap. 4; A. Komar,

J. Math. Phys. 11, 820 (1970).
[2]A. Komar, Phys. Rev. D **4**, 927 (1971).
[3]P. G. Bergmann, GRG **2**, 363 (1971).
[4]J. F. Ritt, *Differential Algebra* (Dover, New York, 1966) or Am. Math. Soc. Colloq. Publ. 33, (1950) Chap. VIII.
[5]C. M. Pereira, Ph.D. thesis, Case Western Reserve University, 1967.
[6]C. M. Pereira, J. Math. Phys. **13**, 1542 (1972).
[7]C. M. Pereira, "The Thin-Sandwich Conjecture" (to be published).
[8]J. W. York, Phys. Rev. Lett. **26**, 1656 (1971); J. W. York, Phys. Rev. Lett. **28**, 1082 (1972).
[9]J. W. York (unpublished).
[10]See J. A. Wheeler, Ref. 1, pp. 348–349.
[11]J. W. York, J. Math. Phys. **13**, 125 (1972).
[12]Working independently, V. Moncrief has derived a special case of this identity: the special case N = const. [V. Moncrief, Ph.D. thesis, Univ. of Maryland, 1972, p. 57 (unpublished).] Moncrief's stimulating discussion does not however demonstrate — even in the N = const. case — that the identity is an integrability condition.
[13]A. Lichnerowicz, *Théories relativistes de la gravitation et de l'électromagnetisme* (Masson et Cie, Paris, 1955), pp. 31–40; A. Lichnerowicz, *Problèmes globaux en mécanique relativiste* (Hermann, Paris, 1939).
[14]One should not attach too much significance to the restriction $R^3_{\ 3} \neq 0$ in the proof. If $R^3_{\ 3} = 0$, then $R^1_{\ 1} \neq 0$ or $R^2_{\ 2} \neq 0$ will suffice. If $R^i_{\ j} = 0$, the number of arbitrary functions of two variables is reduced from four to three, but the proof still goes through provided for example $N_{;3}{}^{;2} \neq 0$. This last restriction ($N_{;3}{}^{;2} \neq 0$) can easily be satisfied in the case of the Schwarzschild metric and thus should be satisfied by any solution which contains isolated sources. I wish to thank R. H. Gowdy for raising questions which the discussion in the present reference is designed to answer.
[15]For example, gauge invariance implies via Noether's theorem the existence of a Noether identity and the Noether identity implies the presence of an integrability condition. For Noether's theorem see Ref. 16. Note that the gauge transformation discussed here would if it exists be closely related to Eq. (8). It is the author's opinion that this gauge transformation would be quite different from the gauge transformation discussed in Ref. 8.
[16]A. Trautman, *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), Chap. 5.

Depolarization of electromagnetic waves excited by distributions of electric and magnetic sources in inhomogeneous multilayered structures of arbitrarily varying thickness. Generalized field transforms

E. Bahar

Electrical Engineering Department, University of Nebraska, Lincoln, Nebraska 68508
(Received 7 March 1973)

A suitable basis for the full wave expansion of electromagnetic fields in inhomogeneous multilayered structures of arbitrarily varying thickness is presented in this paper. To this end, we formulate appropriate sets of transform pairs for the transverse electric and magnetic fields. Since arbitrary distribution of electric and magnetic sources are considered, the complete expansion must be composed of both vertically and horizontally polarized waves. Each set of generalized transforms, for the vertically and horizontally polarized waves, consists of two infinite integrals (continuous spectrum) which correspond to the radiation and the lateral wave terms as well as a finite number of terms (discrete spectrum) which correspond to the surface waves. For a general three-dimensional distribution of sources in any of the structure's layers, the transverse electric and magnetic fields are in general two component vector functions. Thus, the transform pairs involve vector rather than scalar functions. Exact boundary conditions are employed in the analysis rather than approximate surface impedance boundary conditions. When the boundary media of the structure are regarded as perfect electric or magnetic walls, or are characterized by surface impedances, the fields are expressed exclusively in terms of infinite sets of waveguide modes.

1. INTRODUCTION

Full wave solutions to the problem of electromagnetic wave propagation in inhomogeneous multilayered structures of arbitrarily varying thickness were derived recently using Fourier-type transform pairs for the transverse components of the electric and magnetic fields.[1,2] These transform pairs provide suitable bases for the expansion of the electromagnetic fields in all the layers of the structure.

However, only infinite line sources, oriented parallel to the layers of the structure, were considered. Thus, horizontal magnetic line sources excite only vertically polarized waves while horizontal electric line sources excite only horizontally polarized waves. For these cases, the vertically and horizontally polarized waves are completely decoupled and the problem can be solved in terms of scalar functions that represent the transverse electric and magnetic field components.

In this paper, arbitrary three-dimensional source distributions are assumed and, for convenience, it is assumed that both electric and magnetic sources are present. In these cases, therefore, even when the layers of the structure are uniform, both vertically and horizontally polarized waves are excited. In addition, the transverse electric and magnetic fields are in general two-component vectors.

Problems of radio wave propagation in layered media are often solved in terms of electric and magnetic vector potentials or Hertz potentials. However, the manner in which one chooses to represent the electromagnetic fields in terms of potentials is not unique. Thus, for instance, the fields of a horizontal electric dipole may be expressed in terms of a two component magnetic vector (or Hertz) potential[3] or in terms of vertically oriented electric and magnetic vector potentials in conjunction with a judicious use of the reciprocity theorem for electromagnetic waves.[4]

For three-dimensional electric current sources of arbitrary orientation, the two half-space problem has been formulated rather elegantly in terms of dyadic Green's functions.[5] When magnetic sources are also considered, it is necessary to employ two such sets of dyadic Green's functions if this approach is to be used.

In this work, we find it more convenient to work directly in terms of the transverse electric and magnetic fields. To this end, we formulate two sets of transform pairs for the transverse electric and magnetic fields which account for both vertically and horizontally polarized waves. Since the transverse fields are two components vector fields, the transform pairs involve vector functions rather than scalar functions.

For the general problem considered, there is no particular axis of symmetry for the fields; thus, we use a right-hand Cartesian coordinate system. For structures with uniform layers, it is possible to expand the fields in terms of the familiar two dimensional Fourier transforms associated with the variables in planes parallel to the layers of the structure x–z plane (see Fig. 1). However, since these transform pairs are to be applied to problems in which the thickness and the electromagnetic parameters of the structure are assumed to vary along the x axis (see Fig. 2), we employ a combination of the familiar Fourier and generalized Fourier transform pairs associated with the variables in the transverse y–z planes. The field expansions in terms of the generalized Fourier transforms consist of two infinite integrals (continuous spectrum) which correspond to the radiation and the lateral wave terms as well as a finite number of terms (discrete spectrum) which correspond to the surface waves.[1] The completeness of the expansion in terms of the generalized Fourier transforms and its relationship to the familiar Fourier transforms have already been established.[1] It is shown that the solutions obtained in terms of the field transforms are suitable for integration using the steepest descent method.

When vector (or Hertz) potentials are employed to solve these problems, one generally reduces the problem to the solution of vector wave equations. However,

 Copyright © 1973 by the American Institute of Physics

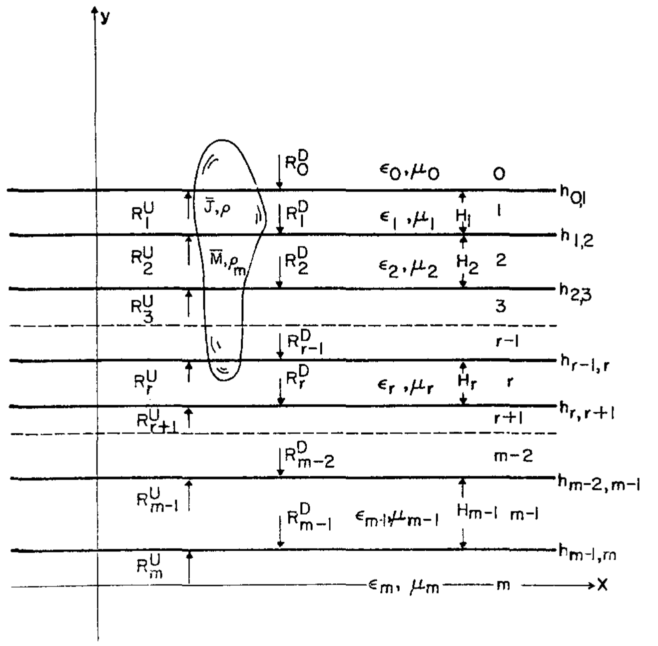

FIG. 1. Electric and magnetic sources distributed in the layers of a uniform multilayered structure.

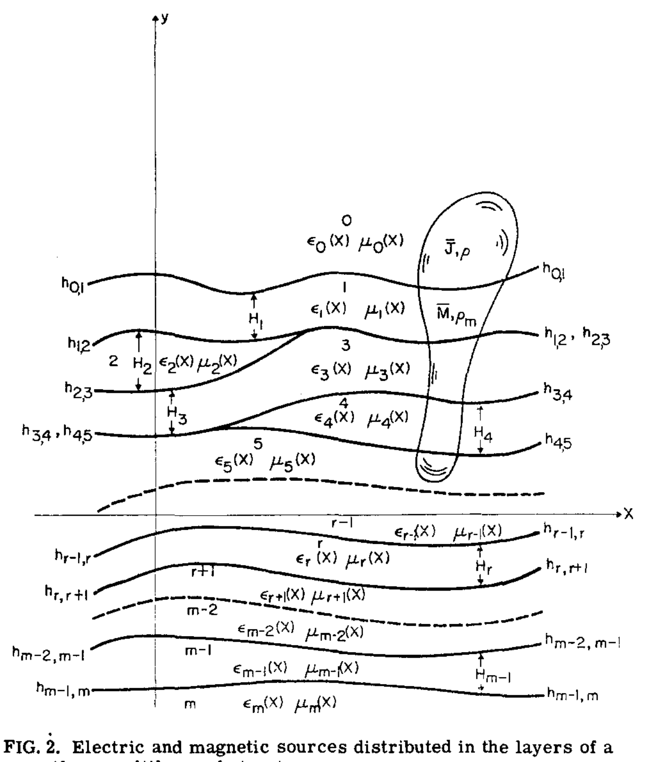

FIG. 2. Electric and magnetic sources distributed in the layers of a nonuniform multilayered structure.

using the transform pairs formulated in this paper, we reduce the problem to the solution of first order ordinary differential equations for the wave amplitudes. When the layers of the structure are uniform, these first order differential equations are completely uncoupled.

2. FORMULATION OF THE PROBLEM

We consider the excitation of electromagnetic fields in stratified media with an arbitrary number of uniform layers (see Fig. 1). The general three-dimensional sources are assumed to be distributed in any of the $(m + 1)$ media of the structure. Since it is often convenient to represent uniform, infinitesimal electric current loops by magnetic dipoles, we shall also assume that both electric currents and charges, $\bar{J}$ and ρ, respectively, as well as magnetic currents and charges, M and ρ_m respectively, are present. An $\exp(i\omega t)$ time dependence is assumed.

The ith medium of the structure is characterized by the electromagnetic parameters ϵ_i and μ_i, which, in general, may be complex. Thus, if the dielectric coefficient and the conductivity of the medium are ϵ_{ir} and σ_i respectively,

$$\epsilon_i = \epsilon_{ir} - i\sigma_i/\omega. \tag{2.1a}$$

The interface between the layers i and $i + 1$ is given by the surface $y = h_{i,i+1}$. The thickness of the ith layer is

$$H_i = h_{i-1,i} - h_{i,i+1}, \quad i = 1, 2, \ldots, m - 1. \tag{2.1b}$$

To obtain an appropriate set of basis functions, we note that for vertically polarized waves $H_y = 0$ and for horizontally polarized waves $E_y = 0$. Thus, we first obtain from Maxwell's equation a set of scalar inhomogeneous wave equations for the normal components E_y and H_y and express the horizontal components

$$\bar{E}_H = E_x \bar{a}_x + E_z \bar{a}_z \tag{2.2a}$$

and

$$\bar{H}_H = H_x \bar{a}_x + H_z \bar{a}_z \tag{2.2b}$$

in terms of the sources and the normal components of the electromagnetic fields.[6] Thus, for each of the $m + 1$ media we get

$$(\nabla^2 + k^2)E_y = i\omega\mu J_y - \frac{1}{i\omega\epsilon}\frac{\partial}{\partial y}\nabla\cdot\bar{J} + \nabla_H\cdot(\bar{M}_H \times \bar{a}_y), \tag{2.3a}$$

$$(\nabla^2 + k^2)H_y = i\omega\epsilon M_y - \frac{1}{i\omega\mu}\frac{\partial}{\partial y}\nabla\cdot\bar{M} + \nabla_H\cdot(\bar{a}_y \times \bar{J}_H), \tag{2.3b}$$

$$\left(\frac{\partial^2}{\partial y^2} + k^2\right)\bar{E}_H = \frac{\partial}{\partial y}\nabla_H E_y - i\omega\mu\nabla_H H_y \times \bar{a}_y + i\omega\mu\bar{J}_H - \frac{\partial}{\partial y}(\bar{M}_H \times \bar{a}_y), \tag{2.3c}$$

and

$$\left(\frac{\partial^2}{\partial y^2} + k^2\right)\bar{H}_H = \frac{\partial}{\partial y}\nabla_H H_y - i\omega\epsilon\bar{a}_y \times \nabla_H E_y + i\omega\epsilon\bar{M}_H - \frac{\partial}{\partial y}(\bar{a}_y \times \bar{J}_H), \tag{2.3d}$$

in which the operators ∇ and ∇_H are defined as

$$\nabla = \nabla_H + \bar{a}_y\frac{\partial}{\partial y} = \bar{a}_x\frac{\partial}{\partial x} + \bar{a}_z\frac{\partial}{\partial z} + \bar{a}_y\frac{\partial}{\partial y}. \tag{2.4a}$$

In the ith medium $\epsilon \to \epsilon_i$, $\mu \to \mu_i$ and the wavenumber is

$$k = \omega(\mu\epsilon)^{1/2} \to \omega(\mu_i\epsilon_i)^{1/2} = k_i. \tag{2.4b}$$

For source-free regions, the solutions to the scalar wave equation (2.3a), subject to the boundary conditions at each interface,

$$\bar{E}_H(x, h^+_{i,i+1}, z) = \bar{E}_H(x, h^-_{i,i+1}, z) \tag{2.5a}$$

and

$$\bar{H}_H(x, h^+_{i,i+1}, z) = \bar{H}_H(x, h^-_{i,i+1}, z), \tag{2.5b}$$

can be expressed as

$$E_{y0} = \exp(\pm iux) Z^V(v,y)\psi^V(v,y)\exp(\pm iwz). \qquad (2.6a)$$

For the ith layer

$$v_i = (k_i^2 - u^2 - w^2)^{1/2}, \quad \mathrm{Im}(v_i) \le 0, \qquad (2.6b)$$

and the basis function $\psi^V(v,y)$ satisfies the one-dimensional wave equation

$$\left(\frac{\partial^2}{\partial y^2} + v^2\right)\psi^V(v,y) = 0 \qquad (2.7a)$$

and the boundary conditions at each interface,

$$\psi^V(v, h^+_{i,i+1}) = \psi^V(v, h^-_{i,i+1}) \qquad (2.7b)$$

and

$$\frac{1}{\epsilon_i}\frac{\partial}{\partial y}\psi^V(v,h^+_{i,i+1}) = \frac{1}{\epsilon_{i+1}}\frac{\partial}{\partial y}\psi^V(v,h^-_{i,i+1}),$$

$$i = 0,1,\ldots,m-1. \qquad (2.7c)$$

The wave impedance $Z^V(v,y)$ for the ith medium is

$$Z^V(v,y) = Z_i^V = (u^2 + w^2)/u\omega\epsilon_i. \qquad (2.7d)$$

Similarly, for source-free regions, the solutions to the scalar wave equation (2.3b), subject to the boundary conditions (2.5), can be expressed as

$$H_{y0} = \exp(\pm iux) Y^H(v,y)\psi^H(v,y)\exp(\pm iwz), \qquad (2.8)$$

in which v is given by (2.6b). The basis function $\psi^H(v,y)$ satisfies the one-dimensional wave equation (in each of the $m+1$ media)

$$\left(\frac{\partial^2}{\partial y^2} + v^2\right)\psi^H(v,y) = 0 \qquad (2.9a)$$

and the boundary conditions at each interface

$$\psi^H(v, h^+_{i,i+1}) = \psi^H(v, h^-_{i,i+1}) \qquad (2.9b)$$

and

$$\frac{1}{\mu_i}\frac{\partial}{\partial y}\psi^H(v,h^+_{i,i+1}) = \frac{1}{\mu_{i+1}}\frac{\partial}{\partial y}\psi^H(v,h^-_{i,i+1}),$$

$$i = 0,1,\ldots,m-1. \qquad (2.9c)$$

The wave admittance $Y^H(v,y)$ for the ith medium is

$$Y^H(v,y) = Y_i^H = (u^2 + w^2)/u\omega\mu_i. \qquad (2.9d)$$

The horizontal components of the electric and magnetic fields associated with the vertical components E_y and H_y can be derived from (2.3c) and (2.3d).

For the purpose of our analysis, it is necessary to derive suitable expansions for the electric and magnetic fields tangent to the transverse, y–z planes. To this end, we now construct the appropriate field transform pairs for the transverse electric and magnetic fields $\bar{E}_T$ and $\bar{H}_T$ respectively, where

$$\bar{E}_T(x,y,z) = E_y(x,y,z)\bar{a}_y + E_z(x,y,z)\bar{a}_z \qquad (2.10a)$$

and

$$\bar{H}_T(x,y,z) = H_y(x,y,z)\bar{a}_y + H_z(x,y,z)\bar{a}_z. \qquad (2.10b)$$

3. THE FIELD TRANSFORM PAIRS FOR THE TRANSVERSE ELECTRIC AND MAGNETIC FIELDS

In order to construct the expressions for the transform pairs for the transverse electric and magnetic fields $\bar{E}_T$ and $\bar{H}_T$ respectively, we employ the following completeness relationships for the one-dimensional Fourier transform:

$$\delta(z - z_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \epsilon^{-iw(z-z_0)}\,dw, \text{ and}$$

$$\delta(w - w') = \frac{1}{2\pi}\int_{-\infty}^{\infty} \epsilon^{iz(w-w')}\,dz, \qquad (3.1)$$

in which $\delta(\alpha - \beta)$ is the Dirac delta function.

The following completeness and orthogonal relationships have also been established for the generalized Fourier transform[1]:

$$\begin{aligned}\delta(y - y_0) &= \sum_v Z^P N^P \psi^P(v,y)\psi^P(v,y_0)\\ &\equiv \int_{-\infty}^{\infty} Z^P N_0^P \psi_0^P(v,y)\psi_0^P(v,y_0)dv_0\\ &\quad + \int_{-\infty}^{\infty} Z^P N_m^P \psi_m^P(v,y)\psi_m^P(v,y_0)dv_m\\ &\quad + \sum_{n=1}^{N} Z^{Pn}\psi_s^{Pn}(v,y)\psi_s^{Pn}(v,y_0)\end{aligned} \qquad (3.2a)$$

and

$$\int_{-\infty}^{\infty} Z^P N_q^P \psi_q^P(v,y)\psi^P(v',y)dy = \Delta(v,v') \equiv \delta_{q,r}\begin{cases}\delta(v - v'), & v' \ne v_s,\\ \delta_{v,v_s}, & v' = v_s,\end{cases} \qquad (3.2b)$$

where the superscript P equals V or H and the subscripts q and r are equal to $0, m$, or s, and $\delta_{q,r}$ is the Kronecker delta. The scalar functions are

$$R_{P0}^{Dh}\psi_0^P(v,y) = \begin{cases}\exp(iv_0 y) + R_{P0}^{Dh}\exp(-iv_0 y), \\ \quad\text{for medium } 0,\\ \prod_{q=1}^{r}\dfrac{T_{Pq-1}^{D}}{T_{Pq}^{DH}}\exp\left(i\sum_{q=1}^{r} v_{q-1,q}h_{q-1,q}\right)\\ \quad\times[\exp(iv_r y) + R_{Pr}^{Dh}\exp(-iv_r y)],\\ \quad\text{for medium } r = 1,2,3,\ldots,m,\end{cases} \qquad (3.3a)$$

$$R_{Pm}^{Uh}\psi_m^P(v,y) = \begin{cases}\prod_{q=1}^{m-r}\dfrac{T_{Pm+1-q}^{U}}{T_{Pm-q}^{UH}}\exp\left(i\sum_{q=1}^{m-r} v_{m-q,m+1-q}h_{m-q,m+1-q}\right)\\ \quad\times[\exp(-iv_r y) + R_{Pr}^{Uh}\exp(iv_r y)], \quad \text{for medium } r = 0,1,2,\ldots,m-1,\\ \exp(-iv_m y) + R_{Pm}^{Uh}\exp(iv_m y), \quad \text{for medium } m,\end{cases} \qquad (3.3b)$$

and

$$\psi_s^{Pn}(v,y) = \psi_s^{Pn}(v,h_{0,1}) \begin{cases} \exp[-iv_0^n(y-h_{0,1})], & \text{for medium } 0, \\ \dfrac{1}{T_{P1}^{DH}} \exp(-iv_1^n h_{0,1})[\exp(iv_1^n y) + R_{P1}^{Dh} \exp(-iv_1^n y)], & \text{for medium } 1, \\ \dfrac{1}{T_{P1}^{DH}} \exp(-iv_1^n h_{0,1}) \prod_{q=2}^{r} \dfrac{T_{Pq-1}^{D}}{T_{Pq}^{DH}} \\ \times \exp\left(i\sum_{q=2}^{r} v_{q-1,q}^n h_{q-1,q}\right)[\exp(iv_r^n y) + R_{Pr}^{Dh} \exp(-iv_r^n y)], \\ \text{for medium } r = 2,3,\ldots,m, \end{cases} \tag{3.3c}$$

where

$$[\psi_s^{Pn}(v,h_{0,1})]^2 = \left[u\Big/iZ_0^P v_0 \frac{d}{du}\frac{1}{R_{P0}^D}\right]_{v=v_n}. \tag{3.3d}$$

The scalar functions for vertically and horizontally polarized waves $\psi^V(v,y)$ and $\psi^H(v,y)$, (2.7) and (2.9), are given by (3.3) on replacing the letter P in all the expressions by V and H, respectively. The reflection coefficient at the $i, i+1$ interface for waves incident from above is R_{Pi}^D and R_{Pi}^U is the reflection coefficient at the $i-1, i$ interface for waves incident from below (See Fig. 1). Thus, for $P = V$ or H,

$$R_{Pm}^D = 0, \quad R_{Pi}^D = \frac{(R_{i+1,i}^P + R_{Pi+1}^{DH})}{(1 + R_{i+1,i}^P R_{Pi+1}^{DH})}, \quad i = 0,1,\ldots,m-1,$$

and

$$R_{P0}^U = 0, \quad R_{Pi}^U = \frac{(R_{i-1,i}^P + R_{Pi-1}^{UH})}{(1 + R_{i-1,i}^P R_{Pi-1}^{UH})}, \quad i = 1,2,\ldots,m, \tag{3.4a}$$

where $R_{i,i\pm1}^V$ and $R_{i,i\pm1}^H$ are the two medium Fresnel reflection coefficients for vertically and horizontally polarized waves respectively,

$$R_{i+1,i}^V = -R_{i,i+1}^V = (v_i\epsilon_{i+1} - v_{i+1}\epsilon_i)/(v_i\epsilon_{i+1} + v_{i+1}\epsilon_i),$$
$$R_{i+1,i}^H = -R_{i,i+1}^H = (v_i\mu_{i+1} - v_{i+1}\mu_i)/(v_i\mu_{i+1} + v_{i+1}\mu_i) \tag{3.4b}$$

and

$$R_{Pi}^{DH} = R_{Pi}^D \exp(-i2v_iH_i), \quad R_{Pi}^{Dh} = R_{Pi}^D \exp(i2v_ih_{i,i+1}),$$
$$R_{Pi}^{UH} = R_{Pi}^U \exp(-i2v_iH_i), \quad R_{Pi}^{Uh} = R_{Pi}^U \exp(-i2v_ih_{i-1,i}). \tag{3.4c}$$

The transmission coefficients are

$$T_{Pi}^D = 1 + R_{Pi}^D, \quad T_{Pi}^U = 1 + R_{Pi}^U,$$
$$T_{Pi}^{DH} = 1 + R_{Pi}^{DH}, \quad T_{Pi}^{UH} = 1 + R_{Pi}^{UH} \tag{3.4d}$$

and the normalization coefficients are

$$N_q^P = \begin{cases} R_{P0}^{Dh}/2\pi Z_0^P, & q = 0, \\ R_{Pm}^{Uh}/2\pi Z_m^P, & q = m, \\ 1, & q = s. \end{cases} \tag{3.4e}$$

The generalized Fourier transform (3.2) consists of two infinite integrals (continuous part of the wavenumber spectrum) which are associated with the radiation and the lateral wave terms and a finite set of surface wave terms (discrete part of the wavenumber spectrum). The relationship between this transform and the familiar transform has been considered earlier.[1] The infinite integrals in the v plane are shown to be associated with branch cut integrals $\mathrm{Im}(v_0) = 0$ and $\mathrm{Im}(v_m) = 0$ in the complex u plane, while the surface wave terms are associated with the residues of the poles at $1/R_{P0}^D = 0$ (or $1/R_{Pm}^U = 0$). The modal equation which determines the surface wave parameters v^n is given by

$$1 - R_{Pi}^U R_{Pi}^D \exp(-i2v_iH_i) = 0, \quad \mathrm{Im}(v) \le 0, \tag{3.5}$$

for P equals V or H and $i = 1,2,3,\ldots,$ or $m-1$. On the basis of the discussion in Sec. 2, we formulate the following field transform pairs:

$$\bar{E}_T(x,y,z) = \sum_v \int_{-\infty}^{\infty} [E^V(x,v,w)\bar{e}_T^V + E^H(x,v,w)\bar{e}_T^H]\, dw, \tag{3.6a}$$

where

$$E^P(x,v,w) = \iint_{-\infty}^{\infty} \bar{E}_T(x,y,z)\cdot(\bar{h}_P^T \times \bar{a}_x)\,dy\,dz, \quad P = V \text{ or } H, \tag{3.6b}$$

and

$$\bar{H}_T(x,y,z) = \sum_v \int_{-\infty}^{\infty} [H^V(x,v,w)\bar{h}_T^V + H^H(x,v,w)\bar{h}_T^H]\,dw, \tag{3.7a}$$

where

$$H^P(x,v,w) = \iint_{-\infty}^{\infty} \bar{H}_T(x,y,z)\cdot(\bar{a}_x \times \bar{e}_P^T)\,dy\,dz, \quad P = V \text{ or } H. \tag{3.7b}$$

The symbol $\sum$ which denotes the summation (integration) over the entire v spectrum is to be interpreted as in (3.2a). The basis functions for the vertically polarized waves are

$$\bar{e}_T^V = Z^V\left(\bar{a}_y\psi^V(v,y) - \frac{\bar{a}_z iw}{u^2+w^2}\frac{\partial\psi^V(v,y)}{\partial y}\right)\phi(w,z) \tag{3.8a}$$

and

$$\bar{h}_T^V = \bar{a}_z\psi^V(v,y)\phi(w,z) \tag{3.8b}$$

and the complementary basis functions for the vertically polarized waves are

$$\bar{e}_V^T = Z^V N^V\left(\bar{a}_y\psi^V(v,y) + \frac{\bar{a}_z iw}{u^2+w^2}\frac{\partial\psi^V(v,y)}{\partial y}\right)\phi^c(w,z) \tag{3.8c}$$

and

$$\bar{h}_V^T = \bar{a}_z N^V\psi^V(v,y)\phi^c(w,z), \tag{3.8d}$$

in which

$$\phi(w,z) = \exp(-iwz) \quad \text{and} \quad \phi^c(w,z) = (1/2\pi)\exp(iwz).$$

For the horizontally polarized waves, the basis functions and the complementary basis functions are respectively

$$\bar{e}_T^H = \bar{a}_z\psi^H(v,y)\phi(w,z), \tag{3.9a}$$

$$\bar{h}_T^H = Y^H \left(-\bar{a}_y \psi^H(v,y) + \frac{\bar{a}_z iw}{u^2+w^2} \frac{\partial \psi^H(v,y)}{\partial y}\right)\phi(w,z), \tag{3.9b}$$

and

$$\bar{e}_H^T = \bar{a}_z N^H \psi^H(v,y)\phi^c(w,z), \tag{3.9c}$$

$$\bar{h}_H^T = Y^H N^H \left(-\bar{a}_y \psi^H(v,y) - \frac{\bar{a}_z iw}{u^2+w^2} \frac{\partial \psi^H(v,y)}{\partial y}\right)\phi^c(w,z). \tag{3.9d}$$

To verify the field transforms (3.6) and (3.7), we substitute (3.6b) into (3.6a) and (3.7b) into (3.7a) and make use of the orthogonal relationships

$$\left.\begin{aligned} &\iint_{-\infty}^{\infty} e_T^P \cdot (h_Q^T \times a_x)' dydz \\ &\iint_{-\infty}^{\infty} h_T^Q \cdot (\bar{a}_x \times e_P^T)' dydz \end{aligned}\right\} = \delta_{P,Q}\Delta(v-v')\delta(w-w'), \tag{3.10}$$

in which P and Q are equal to V or H and $\Delta(v-v')$ is defined in (3.2b). The primes in some of the terms in (3.10) indicate that the variables in these terms are u', v', and w'. The orthogonal relationship (3.10) for $P = Q = V$ or H is a direct consequence of (3.1) and (3.2b). For the case $P = H$ and $Q = V$ the integrand in (3.10) vanishes since $\bar{e}_T^H$ and $(\bar{a}_x \times e_H^T)$ are orthogonal to $(\bar{h}_V^T \times \bar{a}_x)$ and $\bar{h}_T^V$, respectively. Therefore it remains to be shown that either one of the following related orthogonal relationships are satisfied:

$$\iint_{-\infty}^{\infty} \bar{e}_T^V \cdot (\bar{h}_H^T \times \bar{a}_x)' dydz = 0,$$

$$\iint_{-\infty}^{\infty} \bar{h}_T^H \cdot (\bar{a}_x \times \bar{e}_V^T)' dydz = 0. \tag{3.11a}$$

Substituting, for $\bar{e}_T^V$ and $\bar{h}_H^T$, (3.8a) and (3.9d), respectively, in (3.11a), we get

$$\begin{aligned} -\frac{1}{2\pi} &\iint_{-\infty}^{\infty} Z_T^V Y_T^H{}' N^{H'} \exp[-iz(w-w')] \\ &\times \left[\left(\frac{iw}{u^2+w^2}\right)' \psi^V(v,y) \frac{\partial}{\partial y} \psi^H(v',y) \right. \\ &\left. + \frac{iw}{u^2+w^2} \frac{\partial}{\partial y} \psi^V(v,y)\psi^H(v',y)\right] dydz \\ &= -iw\delta(w-w')\int_{-\infty}^{\infty} \frac{N^{H'}}{uu'} \frac{\partial}{\partial y}\left[\psi^V(v,y)\psi^H(v',y)\right. \\ &\left. + \frac{1}{k^2}\frac{\partial}{\partial y}\psi^V(v,y)\frac{\partial}{\partial y}\psi^H(v',y)\right] dy \\ &= \frac{iw\delta(w-w')}{uu'} N^{H'} \sum_{i=1}^{m} \left[\psi^V(v,y)\psi^H(v',y)\right. \\ &\left. + \frac{1}{\omega\epsilon}\frac{\partial}{\partial y}\psi^V(v,y)\frac{1}{\omega\mu}\frac{\partial}{\partial y}\psi^H(v',y)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}. \end{aligned} \tag{3.11b}$$

On applying the boundary conditions (2.7b), (2.7c), (2.9b), and (2.9c) for ψ^V and ψ^H and their derivatives in the above expressions, (3.11c) is shown to vanish. It should be noted that since ψ^V and ψ^H are piecewise continuous functions of y, their derivatives are not continuous at the interface between two media. Thus, in all the above expressions for the basis functions and complementary basis functions, (3.8) and (3.9), respectively, we perform differentiation with respect to y for each medium separately. Subsequently, all the integrations with respect to y must also be performed for each medium separately (3.11c).

For ideal dielectric loaded rectangular waveguides, the basis functions satisfy a biorthogonal relationship.[7] In this case, it can be shown that the complementary or reciprocal basis functions (3.8a), (3.8b), (3.9a), (3.9b) are proportional to the basis functions (3.8c), (3.8d), (3.9c), (3.9d); hence, they are self-reciprocal. For half-spaces with nondissipative media, a biorthogonal relationship exists between the basis functions and their complex conjugates.[8] In this case the reciprocal basis functions can be shown to be proportional to the complex conjugate of the basis functions. However, for the general case no such simple relationship exists between the base and the reciprocal base.[9] The distinction between the basis and the reciprocal basis is essentially that between covariant and contravariant vectors.

4. TRANSFORMATION OF MAXWELL'S EQUATIONS INTO FIRST ORDER ORDINARY DIFFERENTIAL EQUATIONS FOR THE WAVE AMPLITUDES

As indicated in the introduction (Sec. 1), rather than resolve Maxwell's equations by solving the wave equations for the normal components of the electric and magnetic fields E_y and H_y, (2.3a) and (2.3b) respectively, we proceed by transforming Maxwell's equation into first order ordinary differential equations for the wave amplitudes of the transverse electric and magnetic fields (2.10). The reason for following the latter procedure is that it can be carried out whether or not the thickness and the electromagnetic parameters of each layer of the structure are uniform.[10] We shall assume here that the layers of the stratified structure are uniform (Fig. 1).

Eliminating the normal components of the electric and magnetic fields from Maxwell's curl equations, we get the following set of differential equations for the transverse electric and magnetic fields $\bar{E}_T$ and $\bar{H}_T$ in terms of the electric and magnetic sources $\bar{J}$ and $\bar{M}$, respectively:

$$-\frac{\partial \bar{E}_T}{\partial x} = i\omega\mu(\bar{H}_T \times \bar{a}_x) - \frac{1}{i\omega\epsilon}\nabla_T\nabla_T\cdot(\bar{H}_T \times \bar{a}_x) + \bar{M}_T \times \bar{a}_x + \frac{1}{i\omega\epsilon}\nabla_T J_x, \tag{4.1a}$$

$$-\frac{\partial \bar{H}_T}{\partial x} = i\omega\epsilon(\bar{a}_x \times \bar{E}_T) - \frac{1}{i\omega\mu}\nabla_T\nabla_T\cdot(\bar{a}_x \times \bar{E}_T) + \bar{a}_x \times \bar{J}_T + \frac{1}{i\omega\mu}\nabla_T M_x, \tag{4.1b}$$

in which the operator ∇_T is given by

$$\nabla_T = \bar{a}_y \frac{\partial}{\partial y} + \bar{a}_z \frac{\partial}{\partial z}. \tag{4.1c}$$

Scalar multiply (4.1a) by $(\bar{h}_V^T \times \bar{a}_x)'$ and integrate with respect to y and z over the entire y–z plane. Using the properties of the field transform pair for the transverse electric field, (3.6a) and (3.6b), it follows that

$$-\iint \frac{\partial \bar{E}_T}{\partial x}\cdot(\bar{h}_V^T \times \bar{a}_x)' dydz$$

$$= -\frac{d}{dx}\iint_{-\infty}^{\infty} \bar{E}_T\cdot(\bar{h}_V^T \times \bar{a}_x)' dydz = -\frac{dE^{V'}}{dx}. \tag{4.2}$$

To integrate the next two terms in (4.1a), we employ Green's theorem in two dimensions and note that

$$[1 + (1/k^2)\nabla_T\nabla_T]\cdot(\bar{h}_V^T \times \bar{a}_x)' \equiv L(\bar{h}_V^T \times \bar{a}_x)' = (u'/\omega\mu)\bar{e}_V^{T'}. \tag{4.3a}$$

Thus,

$$\iint_{-\infty}^{\infty} i\omega\mu\, L(\bar{H}_T \times \bar{a}_x)\cdot(\bar{h}_V^T \times \bar{a}_x)'\,dydz$$
$$= iu'H^{V'} - \sum_{i=1}^{m}\left[\frac{1}{i\omega\epsilon_i}\,[\nabla_T\cdot(\bar{H}_T \times \bar{a}_x)(\bar{h}_V^T \times \bar{a}_x)'\right.$$
$$\left. - \nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)'(\bar{H}_T \times \bar{a}_x)]\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}\cdot\bar{a}_y\,dz, \qquad (4.3b)$$

in which we have employed the properties of the field transform pair for the transverse magnetic field (3.6c) and (3.6d) and the operator L is defined in (4.3a). Also, from Maxwell's curl equations, we have

$$\nabla_T\cdot(\bar{H}_T \times \bar{a}_x) = i\omega\epsilon E_x + J_x. \qquad (4.4a)$$

In addition,

$$\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x) = N^V\frac{\partial\psi^V}{\partial y}\phi^c \quad\text{and}\quad \bar{H}_T \times \bar{a}_x\cdot\bar{a}_z = \bar{H}_T\cdot\bar{a}_y. \qquad (4.4b)$$

We now note that $E_x, H_y, (1/\epsilon)\partial\psi^V/\partial y$, and $\bar{h}_V^T\cdot\bar{a}_y$ are continuous at each interface. Thus, on applying Green's theorem in two dimensions, the line integral in (4.3b) can be written as

$$\sum_{i=1}^{m}\int_{-\infty}^{\infty}\left[\frac{1}{i\omega\epsilon}J_x(\bar{h}_x \times \bar{a}_x)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}\cdot\bar{a}_y dz$$
$$= -\iint_{-\infty}^{\infty}\frac{1}{i\omega\epsilon}\nabla_T J_x\cdot(\bar{h}_V^T \times \bar{a}_x)dydz$$
$$- \iint_{-\infty}^{\infty}\frac{1}{i\omega\epsilon}J_x\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)dydz. \qquad (4.4c)$$

Hence, (4.1a) reduces to

$$-\frac{dE^V}{dx} - iuH^V$$
$$= \iint_{-\infty}^{\infty}M_z(\bar{h}_V^T\cdot\bar{a}_z)dydz - \iint_{-\infty}^{\infty}\frac{1}{i\omega\epsilon}J_x\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)dydz$$
$$\equiv f^V(x). \qquad (4.5a)$$

We now multiply (4.1b) by $(\bar{a}_x \times \bar{e}_V^T)'$ and integrate with respect to y and z over the entire y–z plane. Following the procedure applied to (4.1a), we get

$$-\frac{dH^V}{dx} - iuE^V$$
$$= \iint_{-\infty}^{\infty}\bar{J}_T\cdot\bar{e}_V^T dydz - \iint_{-\infty}^{\infty}\frac{1}{i\omega\mu}M_x\nabla_T\cdot(\bar{a}_x \times \bar{e}_V^T)dydz$$
$$\equiv g^V(x). \qquad (4.5b)$$

Equations (4.1a) and (4.1b) are now scalar multiplied by $(\bar{h}_H^T \times \bar{a}_x)'$ and $(\bar{a}_x \times \bar{e}_H^T)'$, respectively, and integrated with respect to y and z over the y–z plane. Thus, it can be shown that

$$-\frac{dE^H}{dx} - iuH^H$$
$$= \int_{-\infty}^{\infty}\bar{M}_T\cdot\bar{h}_H^T dydz - \iint_{-\infty}^{\infty}\frac{1}{i\omega\epsilon}J_x\nabla_T\cdot(\bar{h}_H^T \times \bar{a}_x)dydz$$
$$\equiv f^H(x) \qquad (4.6a)$$

and

$$-\frac{dH^H}{dx} - iuE^H$$
$$= \int_{-\infty}^{\infty}J_z(\bar{e}_H^T\cdot\bar{a}_z)dydz - \iint_{-\infty}^{\infty}\frac{1}{i\omega\mu}M_x\nabla_T\cdot(\bar{a}_x \times \bar{e}_H^T)dydz$$
$$\equiv g^H(x). \qquad (4.6b)$$

The electric and magnetic field transforms for the vertically and horizontally polarized waves are expressed in terms of forward and backward wave amplitudes. Thus, we set

$$H^P = a^P + b^P \quad\text{and}\quad E^P = a^P - b^P, \quad P = V \text{ or } H. \qquad (4.7)$$

Furthermore, we define the following functions of x that involve the sources $\bar{J}$ and $\bar{M}$, as

$$A^P \equiv -(f^P + g^P)/2 \quad\text{and}\quad B^P \equiv -(f^P - g^P)/2,$$
$$P = V \text{ or } H, \qquad (4.8)$$

where f^V, g^V, f^H, and g^H are defined in (4.5) and (4.6). Expressing (4.5) and (4.6) in terms of the forward and backward wave amplitudes a^P and b^P (P equals V or H) respectively, (4.7), we get the following sets of uncoupled first order ordinary differential equations for the wave amplitudes:

$$\frac{da^P}{dx} + iua^P = A^P \quad\text{and}\quad \frac{db^P}{dx} - iub^P = -B^P, \quad (4.9a)$$

in which P is equal to V or H. The solution to (4.9a) subject to the boundary conditions (radiation condition),

$$a^P(x = -\infty) = 0, \quad b^P(x = \infty) = 0 \qquad (4.9b)$$

can be readily shown to be

$$a^P = \exp(-iux)\int_{-\infty}^{x}\exp(iux')A^P(x')dx' \qquad (4.10a)$$

and

$$b^P = \exp(iux)\int_{x}^{\infty}\exp(-iux')B^P(x')dx'. \qquad (4.10b)$$

Thus, employing the field transform pairs formulated in Sec. 3, we have derived rigorous expressions for the vertically and horizontally polarized waves excited by a general distribution of electric and magnetic sources in any of the uniform layers of a stratified structure. In the next section, it is shown that these expressions are particularly suitable for integration by the steepest descent method.

5. THE FIELDS FOR VERTICAL AND HORIZONTAL ELECTRIC DIPOLES

We shall use the analysis of Sec. 4 to write the expression for the electromagnetic fields for vertical and horizontal electric dipoles. Fields due to vertically and horizontally oriented uniform, infinitesimal current loops may be obtained in a similar manner by considering the dual problem of excitation by equivalent vertical and horizontal magnetic dipoles. For a vertical electric dipole located at $\bar{r} = \bar{r}_0$, we write

$$\bar{J}(x,y,z) = \mathcal{I}_y\delta_y(\bar{r} - \bar{r}_0)\bar{a}_y = \mathcal{I}_y\delta(x - x_0)\delta(y - y_0)\delta(z - z_0)\bar{a}_y, \qquad (5.1a)$$

in which $\delta(\bar{r} - \bar{r}_0)$ is the three-dimensional Dirac delta function and $\mathcal{I}_y$ is the electric current moment measured in amp meters. Thus,

$$A^H = B^H = 0$$

and

$$A^V = -B^V = -\mathcal{I}_y(e_V^T \cdot a_y)\delta(x-x_0)/2$$
$$= -\mathcal{I}_y[Z^V N^V \psi^V(v,y)\phi^c(w,z)]_0 \delta(x-x_0)/2. \quad (5.1b)$$

Obviously, no horizontally polarized waves are excited. The transverse magnetic and electric fields in any of the $m+1$ media are

$$\bar{H}_T = \bar{a}_z \,\mathrm{sgn}(x_0 - x)\frac{\mathcal{I}_y}{4\pi}\sum_v \int_{-\infty}^{\infty} \exp[-iu|x-x_0|-iw(z-z_0)] \times [Z^V N^V \psi^V]_0 \psi^V(v,y)dw \quad (5.2a)$$

and

$$\bar{E}_T = \frac{\mathcal{I}_y}{4}\sum_v \int_{\infty}^{\infty} \exp[-iu|x-x_0|-iw(z-z_0)][Z^V N^V \psi^V]_0 \times Z^V\left(-\bar{a}_y\psi^V(v,y) + \frac{\bar{a}_z iw}{u^2+w^2}\frac{\partial}{\partial y}\psi^V(v,y)\right)dw. \quad (5.2b)$$

The axial components of the magnetic and electric fields are

$$H_x = \frac{\mathcal{I}_y}{4\pi}\sum_v \int_{-\infty}^{\infty} \exp[-iu|x-x_0|-iw(z-z_0)] \times [Z^V N^V \psi^V]_0 \frac{w}{u}\psi^V(v,y)dw \quad (5.2c)$$

and

$$E_x = \mathrm{sgn}(x_0 - x)\frac{\mathcal{I}_y}{4\pi}\sum_v \int_{-\infty}^{\infty} \exp[-iu|x-x_0|-iw(z-z_0)] \times [Z^V N^V \psi^V]_0 \frac{1}{i\omega\epsilon}\frac{\partial}{\partial y}\psi^V(v,y)dw, \quad (5.2d)$$

in which

$$\mathrm{sgn}(x-x_0) = \begin{cases} 1 & \text{for } x > x_0 \\ -1 & \text{for } x < x_0. \end{cases} \quad (5.2e)$$

At the plane $x = x_0$ it can be shown from (5.2) that

$$\bar{a}_x \times \bar{H}|_{x_0^+} = -\tfrac{1}{2}\mathcal{I}_y\,\delta(y-y_0)\delta(z-z_0)\bar{a}_y = \tfrac{1}{2}\bar{J}_{TS}, \quad (5.3a)$$

where $\bar{J}_{TS}$ is the transverse surface current per unit area at $x = x_0$. Similarly,

$$\bar{a}_x \cdot \bar{E}|_{x_0^+} = -\mathcal{I}_y\delta'(z-z_0)\delta(y-y_0)/2i\omega\epsilon_0$$
$$= -(\nabla_T \cdot \bar{J}_{TS})/2i\omega\epsilon_0 = \rho_s/2\epsilon_0, \quad (5.3b)$$

in which ρ_s is the surface charge per unit area. The tangential electric field and axial magnetic field are continuous at $x = x_0$. For $x > x_0 = 0, y$ and $y_0 > h_{0,1} = 0$, and $z = z_0 = 0$, the z component of the magnetic radiation field can be written as

$$H_z = -\frac{\mathcal{I}_y}{8\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\exp(iv_0y_0) + R_{V0}^D \exp(-iv_0y_0)] \times \exp(-iv_0 y - iux)dv_0 dw. \quad (5.4a)$$

By making the substitutions

$$x = r\sin\theta, \quad y = r\cos\theta, \quad y_0 = d$$

and

$$u = k_0 \sin\theta' \cos\phi', \quad v_0 = k_0\cos\theta', \quad w = k_0 \sin\theta' \sin\phi' \quad (5.4b)$$

and using the steepest descent method, it can readily be shown that the azimuthal magnetic field is

$$H_\phi = \frac{ik_0\mathcal{I}_y \sin\theta \exp(-ik_0 r)}{4\pi r}[\exp(ik_0 d\cos\theta) + R_{V0}^D \exp(-ik_0 d\cos\theta)]. \quad (5.4c)$$

The lateral wave contribution to the magnetic field [the infinite integral with respect to v_m, (3.2a)] can be evaluated in a similar manner using the steepest descent method. The contributions from the surface waves can be obtained directly from (2.5). Any of the other components of the electromagnetic field can be evaluated in a similar manner.

For a horizontal electric dipole, we write as in (5.1a)

$$\bar{J}(x,y,z) = \mathcal{I}_x\delta(r-r_0)\bar{a}_x. \quad (5.5a)$$

In this case, both vertically and horizontally polarized waves are excited:

$$A^V = B^V = \mathcal{I}_x[\nabla_T \cdot (\bar{h}_V^T \times \bar{a}_x)/2i\omega\epsilon]_0 = \mathcal{I}_x\left[N^V \frac{\partial \psi^V}{\partial y}\phi^c\big/2i\omega\epsilon\right]_0 \quad (5.5b)$$

and

$$A^H = B^H = \mathcal{I}_x[\nabla_T\cdot(\bar{h}_H^T \times \bar{a}_x)/2i\omega\epsilon]_0 = \mathcal{I}_x[(w/u)N^H\psi^H\phi^c]_0. \quad (5.5c)$$

The transverse magnetic and electric vertically polarized fields are

$$\bar{H}_T^V = a_z\frac{\mathcal{I}_x}{4\pi}\sum_v\int_{-\infty}^{\infty} \exp[-iu|x-x_0|-iw(z-z_0)] \times \left[\frac{N^V}{i\omega\epsilon}\frac{\partial\psi^V}{\partial y}\right]_0 \psi^V(v,y)dw \quad (5.6a)$$

and

$$\bar{E}_T^V = \mathrm{sgn}(x-x_0)\frac{\mathcal{I}_x}{4\pi}\sum_v\int_{-\infty}^{\infty} \exp[-iu|x-x_0| - iw(z-z_0)]\left[\frac{N^V}{i\omega\epsilon}\frac{\partial\psi^V}{\partial y}\right]_0 \times Z^V\left(\bar{a}_y\psi^V(v,y) - \frac{\bar{a}_z iw}{u^2+w^2}\frac{\partial\psi^V(v,y)}{\partial y}\right)dw. \quad (5.6b)$$

The corresponding axial magnetic and electric fields are

$$H_x^V = \mathrm{sgn}(x_0 - x)\frac{\mathcal{I}_x}{4\pi}\sum_v\int_{-\infty}^{\infty} \exp[-iu|x-x_0|-iw(z-z_0)] \times\left[\frac{N^V}{i\omega\epsilon}\frac{\partial\psi^V}{\partial y}\right]_0 \frac{w}{u}\psi^V(v,y)dw \quad (5.6c)$$

and

$$E_x^V = \frac{\mathcal{I}_x}{4\pi}\sum_v\int_{-\infty}^{\infty}\exp[-iu|x-x_0|-iw(z-z_0)] \times\left[\frac{N^V}{i\omega\epsilon}\frac{\partial\psi^V}{\partial y}\right]_0 \frac{1}{i\omega\epsilon}\frac{\partial}{\partial y}\psi^V(v,y)dw. \quad (5.6d)$$

The transverse magnetic and electric horizontally polarized fields are

$$\bar{H}_T^H = \frac{\mathcal{I}_x}{4\pi}\sum_v\int_{-\infty}^{\infty}\exp[-iu|x-x_0|-iw(z-z_0)]\left[\frac{w}{u}N^H\psi^H\right]_0 \times Y^H\left(-\bar{a}_y\psi^H(x,y) + \bar{a}_z\frac{iw}{u^2+w^2}\frac{\partial}{\partial y}\psi^H(v,y)\right)dw \quad (5.7a)$$

$$\bar{E}_T^H = \bar{a}_z\,\mathrm{sgn}(x-x_0)\frac{\mathcal{I}_x}{4\pi}\sum_v\int_{-\infty}^{\infty}\exp[-iu|x-x_0|-iw(z-z_0)] \times\left[\frac{w}{u}N^H\psi^H\right]_0 \psi^H(v,y)dw. \quad (5.7b)$$

The corresponding axial magnetic and electric fields are

$$H_x^H = \operatorname{sgn}(x_0 - x)\,\frac{\mathcal{J}_x}{4\pi}\sum_v \int_{-\infty}^{\infty} \exp[-iu|x-x_0| - iw(z-z_0)] \times\left[\frac{w}{u}N^H\psi^H\right]_0 \frac{1}{i\omega\mu}\,\frac{\partial}{\partial y}\,\psi^H(v,y)\,dw \quad (5.7c)$$

and

$$E_x^H = -\,\frac{\mathcal{J}_x}{4\pi}\sum_v \int_{-\infty}^{\infty} \exp[-iu|x-x_0| - iw(z-z_0)] \times\left[\frac{w}{u}N^H\psi^H\right]_0 \frac{w}{u}\,\psi^H(v,y)\,dw. \quad (5.7d)$$

For the case of the horizontal electric dipole, the tangential magnetic field and the normal electric and magnetic fields are continuous; however, the tangential electric field is discontinuous. It can be shown from (5.6b) and (5.7b) that the total tangential electric field $\bar{E}_T^V + \bar{E}_T^H$ at $x = x_0^+$ is

$$\begin{aligned}\bar{E}_T|_{x_0^+} &= -\,\mathcal{J}_x[\delta'(y-y_0)\delta(z-z_0)\bar{a}_y \\ &\quad + \delta(y-y_0)\delta'(z-z_0)\bar{a}_z]/2i\omega\epsilon \\ &= -\,\nabla_T J_{sx}/2i\omega\epsilon. \end{aligned} \quad (5.8a)$$

Also from (5.6c) and (5.7c) it can be shown that

$$\bar{a}_x\cdot\bar{H}|_{x_0^+} = \nabla_T\cdot(\bar{a}_x \times \nabla_T J_{sx})/2k^2 \equiv 0. \quad (5.8b)$$

Equations (5.8) and the continuity of the tangential magnetic field and normal electric field at $x = x_0$ are consistent with the boundary conditions for double layered sources.[11]

6. CONCLUDING REMARKS

We have formulated in this paper a set of transform pairs for the transverse electric and magnetic fields in multilayered structures. These transforms provide a suitable basis for the full wave expansion of the electromagnetic fields into vertically and horizontally polarized waves. Each set of transforms consists of two infinite integrals (continuous part of the wave spectrum) which correspond to the radiation and lateral wave terms as well as a finite number of terms (discrete part of the wave spectrum) which correspond to the surface waves.

These transforms have been applied to the problem of electromagnetic wave propagation in multilayered structures excited by an arbitrary distribution of electric and magnetic sources in any of the uniform layers of the stratified medium. They enable the conversion of Maxwell's equations into a complete set of first order uncoupled ordinary differential equations for the vertically and horizontally polarized wave amplitudes (4.9) which are readily solved (4.10).

The special cases of radiation by vertical and horizontal electric dipoles are considered in some detail in Sec. 5. It is shown that the solutions satisfy the proper boundary conditions as an interface with single or double layered source distributions. It was also demonstrated that the expressions for the continuous part of the wave spectrum are conducive to integration by the steepest descent method.

In the companion paper[10] it is shown that the generalized field transforms also form a suitable basis for the expansion of the transverse electric and magnetic fields excited by an arbitrary distribution of electric and magnetic sources in nonuniform multilayered structures. In this case however, since the basis functions are also functions of the variable x, the resulting first order ordinary differential equations for the vertically and horizontally polarized wave amplitudes are coupled. This accounts not only for the forward and backward scattering of the primary fields in the nonspecular direction (with respect to the x–z plane) but also wave coupling between the vertically and horizontally polarized waves.

The generalized field transform pairs can also be applied to problems in which one or both of the bounding media of the multilayered structure are regarded as perfect electric or magnetic walls or if they are characterized by surface impedances. Thus, for instance, if the structure is a multilayered waveguide with perfectly conducting walls, the continuous part of the wave spectrum vanishes and the fields are given in terms of an infinite set of waveguide modes (discrete part of the wave spectrum).[1] On the other hand, one can derive the fields due to arbitrary electric and magnetic source distributions in free space by setting $R_{P0}^{D} \to 0$ in the first infinite integral in (3.2a) and disregarding the second infinite integral (lateral wave) and the surface wave terms.

ACKNOWLEDGMENTS

The work reported in this paper was supported by the National Science Foundation and the Engineering Research Center of the University of Nebraska. The author wishes to thank J. R. Wait, F. G. Ullman, and L. Lipsky for their comments. The manuscript was typed by Mrs. E. Everett.

[1]E. Bahar, J. Math. Phys. **14**, 1024 (1973).
[2]E. Bahar, J. Math. Phys. **14**, 1030 (1973).
[3]A. Sommerfield, *Particle Differential Equations in Physics* (Academic, New York, 1949).
[4]J. R. Wait, *Electromagnetic Waves in Stratified Media* (Pergamon, 1962).
[5]Chen-To Tai, *Dyadic Green's Functions in Electromagnetic Theory* (Intext Educational Publishers, Scranton, Pennsylvania, 1971).
[6]S. N. Samaddar, Electron. Lett. 3 (1), 3 (1967).
[7]R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960).
[8]B. F. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall, Englewood Cliffs, New Jersey, 1973).
[9]B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956).
[10]E. Bahar, J. Math. Phys. **14**, 1510 (1973).
[11]J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, **14**, 1510 (1973).

Depolarization of electromagnetic waves excited by distributions of electric and magnetic sources in inhomogeneous multilayered structures of arbitrarily varying thickness. Full wave solutions

E. Bahar

Electrical Engineering Department, University of Nebraska, Lincoln, Nebraska 68508

(Received 7 March 1973)

Full wave solutions to the problem of depolarization of electromagnetic waves excited by general three-dimensional distributions of electric and magnetic sources in inhomogeneous multilayered structures of arbitrarily varying thickness are derived. Generalized field transforms that provide an appropriate basis for the expansion of transverse electromagnetic fields are employed to convert Maxwell's equations into a set of coupled first order ordinary differential equations for the forward and backward, vertically and horizontally polarized wave amplitudes. The continuous parts of the complete wave spectrum correspond to the radiation and lateral wave terms while the discrete part of the wave spectrum corresponds to the surface wave or trapped waveguide modes. Exact boundary conditions are imposed at all the interfaces of the structure and the solution is not restricted by the surface impedance concept. When the bounding media of the structure are characterized by perfect electric or magnetic walls, the fields are expressed exclusively in terms of waveguide modes. On the other hand, if the electromagnetic parameters are functions of one coordinate variable, the solutions are expressed exclusively in terms of an infinite integral—the radiation term. The solutions are shown to satisfy the reciprocity relationships.

1. INTRODUCTION

Full wave solutions to the problem of depolarization of electromagnetic waves excited by general three-dimensional distributions of electric and magnetic sources, in inhomogeneous, multilayered structures of arbitrarily varying thickness are derived (see Fig. 1).

For the purpose of the analysis, generalized field transforms are employed to provide a suitable complete expansion for the vertically and horizontally polarized transverse electric and magnetic fields.[1] The complete expansion into vertically and horizontally polarized waves, consist of two infinite integrals (continuous wave spectrum) which constitute the radiation and the lateral waves as well as a finite set of terms (discrete wave spectrum) that constitute the surface waves. The generalized field transforms are used to convert Maxwell's equations into a set of coupled first order ordinary differential equations for the forward and backward, vertically and horizontally polarized wave amplitudes.

In our analysis, exact boundary conditions are imposed at all the interfaces of the structure; thus, the solutions are not restricted by the approximate surface impedance concept. It is shown that these solutions satisfy the reciprocity relationships in electromagnetic theory.

A very broad group of problems may be solved using the analysis derived in this paper. These include depolarization of electromagnetic waves due to variations in the height as well as due to variations in the electromagnetic parameters of the earth's surface as for the case of electromagnetic waves obliquely incident on a coast line. The analysis is also applicable to problems of electromagnetic radiation by sources embedded in the nonuniform layers of the earth's crust (hardened communication systems) and to problems of wave scattering at the earth's surface by objects of finite cross section buried in the earth's crust (remote sensing, see Fig. 2). Nonuniform artificial surface wave structures, nonuniform layered waveguides, and electromagnetic radiation in one dimensionally inhomogeneous media are other important special cases that can be considered likewise. Thus, for instance, when the bounding media of the nonuniform structure are regarded as perfect electric or magnetic walls or characterized by surface impedances,

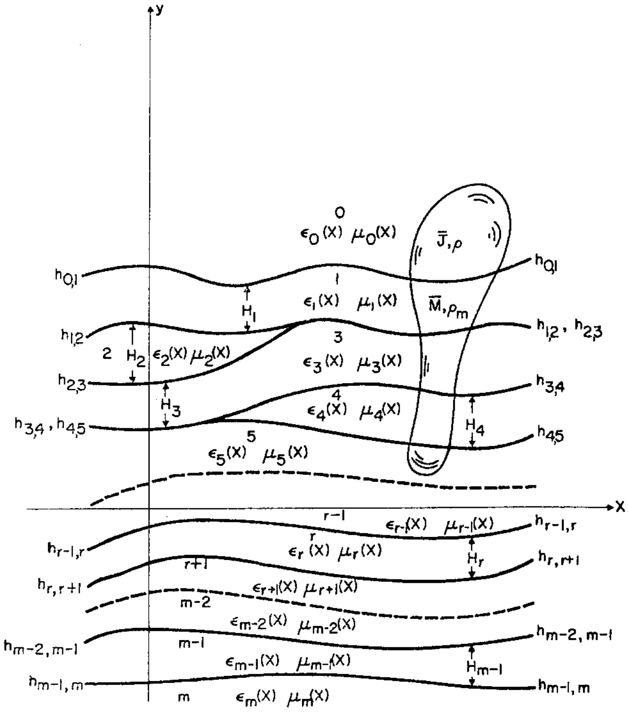

FIG. 1. General source distribution in nonuniform multilayered structures.

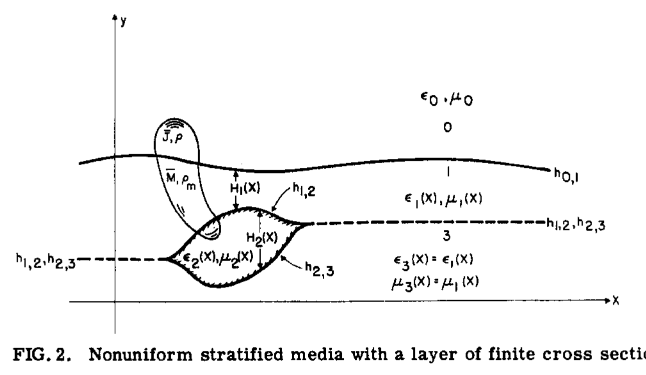

FIG. 2. Nonuniform stratified media with a layer of finite cross section.

Copyright © 1973 by the American Institute of Physics

the electromagnetic fields are expressed exclusively in terms of infinite sets of waveguide modes. However, if the electromagnetic parameters are functions of only one coordinate variable, the fields are given exclusively in terms of a single infinite integral–the radiation term.

A right-hand Cartesian coordinate system is used since for the general problem considered there is no axis of symmetry for the fields. The height of the interfaces of the layered structure as well as the electromagnetic parameters are assumed to be arbitrary functions of the variable x (see Figs. 1 and 2).

2. FORMULATION OF THE PROBLEM

The excitation of electromagnetic waves by general three-dimensional distributions of sources in nonuniform multilayered structures is considered (see Fig. 1). For convenience it is assumed that both electric and magnetic sources, J and ρ and M and ρ_m respectively, are present in any of the $(m+1)$ media of the structure. A suppressed $\exp(i\omega t)$ time dependence is assumed in this work.

The ith medium of the structure is characterized by the electromagnetic parameters $\epsilon_i(x)$ and $\mu_i(x)$ which in general may be complex to account for dissipation of electromagnetic power in the medium. Thus if ϵ_{ir} and σ_i are the dielectric coefficient and conductivity respectively of the ith medium,

$$\epsilon_i(x) = \epsilon_{ir}(x) - \sigma_i(x)/\omega. \tag{2.1a}$$

The interface between medium i and $i+1$ is given by the surface $y = h_{i,i+1}$ and the thickness of the ith layer is

$$H_i(x) = h_{i-1,i}(x) - h_{i,i+1}(x), \quad i = 1, 2, \ldots, m-1. \tag{2.1b}$$

Maxwell's equations for the transverse electric and magnetic fields, $\bar{E}_T$ and $\bar{H}_T$ respectively, are

$$-\frac{\partial \bar{E}_T}{\partial x} = i\omega\mu(\bar{H}_T \times \bar{a}_x) - \frac{1}{i\omega\epsilon}\nabla_T\nabla_T\cdot(\bar{H}_T \times \bar{a}_x) + \bar{M}_T \times \bar{a}_x + \frac{1}{i\omega\epsilon}\nabla_T J_x \tag{2.2a}$$

and

$$-\frac{\partial \bar{H}_T}{\partial x} = i\omega\epsilon(\bar{a}_x \times \bar{E}_T) - \frac{1}{i\omega\mu}\nabla_T\nabla_T\cdot(\bar{a}_x \times \bar{E}_T) + \bar{a}_x \bar{J}_T + \frac{1}{i\omega\mu}\nabla_T M_x, \tag{2.2b}$$

in which the operator ∇_T is given by

$$\nabla_T = \bar{a}_y\frac{\partial}{\partial y} + \bar{a}_z\frac{\partial}{\partial z} \tag{2.2c}$$

and the transverse vectors are

$$\bar{A}_T = \bar{a}_y A_y + \bar{a}_z A_z. \tag{2.2d}$$

The following field transform pairs provide the basis for the complete expansion of the transverse electric and magnetic fields into vertically and horizontally polarized waves:

$$\bar{E}_T(x,y,z) = \sum_v \int_{-\infty}^{\infty} [E^V(x,v,w)\bar{e}_T^V + E^H(x,v,w)\bar{e}_T^H]dw, \tag{2.3a}$$

where

$$E^P(x,v,w) = \int_{-\infty}^{\infty} \bar{E}_T(x,y,z)\cdot(\bar{h}_P^T \cdot a_x)dydz, \quad P = V \text{ or } H, \tag{2.3b}$$

$$\bar{H}_T(x,y,z) = \sum_v \int_{-\infty}^{\infty} [H^V(x,v,w)\bar{h}_T^V + H^H(x,v,w)\bar{h}_T^H]dw, \tag{2.4a}$$

where

$$H^P(x,v,w) = \int_{-\infty}^{\infty} \bar{H}_T(x,y,z)\cdot(\bar{a}_x \times \bar{e}_P^T)dydz, \quad P = V \text{ or } H. \tag{2.4b}$$

The letters V and H are used as subscripts or superscripts to denote vertical or horizontal polarization respectively. The explicit expressions for the basis functions $\bar{e}_T^P$ and $\bar{h}_T^P$ and the complementary basis functions $\bar{e}_P^T$ and $\bar{h}_P^T$ (P equals V or H) for stratified media with an arbitrary number of layers $(m+1)$, have been derived earlier.[1] The symbol $\sum_v$ in (2.3a) and (2.4a) denote the summation (integration) over the entire v spectrum. It consists of two infinite integrals (continuous part of the wave spectrum) which originate from branch cut integrals $\mathrm{Im}(v_0) = 0$ and $\mathrm{Im}(v_m) = 0$ and a set of terms (discrete part of the wave spectrum) which are the residue contributions at the poles of the reflection coefficients R_{P0}^D ($P = V$ or H) looking into the layered structure from above (see Fig. 1). The modal equation for the discrete part of the spectrum, (the surface or trapped waveguide modes) is[2]

$$1 - R_{Pi}^U \exp(-i2v_iH_i) = 0, i = 1, 2, \ldots, \text{or } m-1, \tag{2.5a}$$

in which R_{Pi}^U and R_{Pi}^D are the reflection coefficients in medium i looking upwards and downwards, respectively (see Fig. 1) and

$$v_i = (k_i^2 - u^2 - w^2)^{1/2}, \quad \mathrm{Im}(v_i) \le 0, \tag{2.5b}$$

where $k_i = \omega(\mu_i\epsilon_i)^{1/2}$ is the wavenumber for medium i. Since the electromagnetic parameters and the heights of the interfaces of the layered structure are assumed to be functions of x, in this paper the basis functions and the complementary basis functions are functions of the transverse variables y and z as well as the axial coordinate variable x. Nevertheless, the orthogonality relationships derived for uniform layered structures can be used in the problem presently under consideration. It can be readily shown[1] that, for any surface $x = $ const,

$$\left.\begin{aligned}&\int_{-\infty}^{\infty} \bar{e}_T^P\cdot(\bar{h}_Q^T \times \bar{a}_x)'dydz\\ &\int_{-\infty}^{\infty} \bar{h}_T^Q\cdot(\bar{a}_x \times \bar{e}_P^T)'dydz\end{aligned}\right\} = \delta_{PQ}\Delta(v-v')\delta(w-w'), \tag{2.6a}$$

in which P and Q are equal to V or H and for the primed terms the variables are understood to be u', v' and w':

$$\Delta(v-v') = \delta_{q,r}\begin{cases}\delta(v-v'), & v' \ne v_s,\\ \delta_{v,v_s}, & v' = v_s,\end{cases} \tag{2.6b}$$

where q and r are equal to $0, m$, or s (corresponding to the radiation, lateral wave and surface wave constituents of the wave spectrum[2] and $\delta(\alpha-\beta)$ and $\delta_{\alpha,\beta}$ are the Dirac delta function and the Kronecker delta.

In the following section, we employ the transform pairs for the transverse electric and magnetic fields, $\bar{E}_T$ and $\bar{H}_T$ [(2.3) and (2.4)], respectively, to convert Maxwell's equations for $\bar{E}^T$ and $\bar{H}^T$ [(2.2)] into a set of coupled first order ordinary differential equations for the forward and backward, vertically and horizontally polarized wave amplitudes $a^P(x,v,w)$ and $b^P(x,v,w)$. The wave amplitudes are related as follows to the scalar transforms E^P and H^P, (2.3) and (2.4):

$$H^P = a^P + b^P \quad \text{and} \quad E^P = a^P - b^P, \quad P = V \text{ or } H. \tag{2.7}$$

3. COUPLED EQUATIONS FOR THE VERTICALLY AND HORIZONTALLY POLARIZED WAVE AMPLITUDES

Beginning with (2.2a), we substitute (2.3a) and (2.4a) for the transverse electric and magnetic fields $\bar{E}_T$ and $\bar{H}_T$ respectively, and scalar multiply the equation by $(\bar{h}_V^T \times \bar{a}_x)'$, and integrate with respect to y and z over the entire y-z plane. We note that the transverse fields and the basis functions are in general piecewise continuous functions; thus, the integration must be performed in each medium of the structure separately. Furthermore, since the layered structure varies along the x axis, the field expansions (2.3a) and (2.4a) do not converge uniformly at all points of the y-z plane; hence, in general it is not permissible to interchange orders of integration and differentiation. Thus, we have

$$-\int_{-\infty}^{\infty} \frac{\partial \bar{E}_T}{\partial x}\cdot(\bar{h}_V^T \times \bar{a}_x)'dydz = \frac{d}{dx}\int_{-\infty}^{\infty}\bar{E}_T\cdot(\bar{h}_V^T \times \bar{a}_x)'dydz$$
$$+\int_{-\infty}^{\infty}\bar{E}_T\cdot\frac{\partial}{\partial x}(\bar{h}_V^T \times \bar{a}_x)'dydz$$
$$-\sum_{i=1}^{m}\int_{-\infty}^{\infty}\frac{dh_{i-1,i}}{dx}\left[\bar{E}_T\cdot(\bar{h}_V^T \times \bar{a}_x)'\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}dz \qquad (3.1)$$

and, using Green's theorem in two dimensions, we get

$$\int_{-\infty}^{\infty}\frac{1}{\omega\epsilon}\nabla_T J_x\cdot(\bar{h}_V^T \times \bar{a}_x)'dydz$$
$$= -\int_{-\infty}^{\infty}\frac{1}{\omega\epsilon}J_x\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)'dydz$$
$$-\sum_{i=1}^{m}\int_{-\infty}^{\infty}\left[\frac{1}{\omega\epsilon}J_x(\bar{h}_x \times \bar{a}_x)'\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}\cdot\bar{a}_y dz \qquad (3.2)$$

and

$$\int_{-\infty}^{\infty}\frac{1}{\omega\epsilon}\nabla_T\nabla_T\cdot(\bar{H}_T \times \bar{a}_x)\cdot(\bar{h}_V^T \times \bar{a}_x)'dydz$$
$$= \int_{-\infty}^{\infty}\frac{1}{\omega\epsilon}(\bar{H}_T \times \bar{a}_x)\cdot\nabla_T\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)'dydz$$
$$-\sum_{i=1}^{m}\int_{-\infty}^{\infty}\left[\frac{1}{\omega\epsilon}[\nabla_T\cdot(\bar{H}_T \times \bar{a}_x)(\bar{h}^T \times \bar{a}_x)'\right.$$
$$\left.-\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)'(\bar{H}_T \times \bar{a}_x)]\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}\cdot\bar{a}_y dz. \qquad (3.3)$$

From Maxwell's equations and the boundary conditions at the interfaces, we have

$$\nabla_T\cdot(\bar{H}_T \times \bar{a}_x) = i\omega\epsilon E_x + J_x, \qquad (3.4a)$$

$$\left[E_x = -\frac{dh_{i-1,i}}{dx}E_y\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}, \qquad i = 1, 2, \ldots, m \qquad (3.4b)$$

and

$$\left[\bar{H}_T \times \bar{a}_x\cdot\bar{a}_y = H_z\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0. \qquad i = 1, 2, \ldots, m \qquad (3.4c)$$

It can also be shown that

$$[(1/\omega\epsilon)\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \qquad i = 1, 2, \ldots, m \qquad (3.5a)$$

$$[1 + (1/k^2)\nabla_T\nabla_T]\cdot(\bar{h}_V^T \times \bar{a}_x) = (u/\omega\mu)\bar{e}_V^T, \qquad (3.5b)$$

and

$$\bar{E}_T\cdot(\bar{h}_V^T \times \bar{a}_x) = E_y(\bar{h}_V^T\cdot\bar{a}_z). \qquad (3.5c)$$

Thus, on employing (3.2) through (3.5) and the properties of the field transforms (2.3b) and (2.4b), (3.1) reduces to

$$-\frac{d}{dx}E^V(x, v', w') - iu'H^V(x, v', \omega')$$
$$+\sum_v\int_{-\infty}^{\infty}(E^VC_V^V + E^HC_V^H)dw = f^V(x), \qquad (3.6a)$$

in which

$$C_V^P(v', w', v, w) \equiv \int_{-\infty}^{\infty}\bar{e}_T^P\cdot\frac{\partial}{\partial x}(\bar{h}_V^T \times \bar{a}_x)'dydz,$$
$$P = V \text{ or } H, \qquad (3.6b)$$

and

$$f^V(x) \equiv \int_{-\infty}^{\infty}[M_z\bar{h}_V^{T\prime}\cdot\bar{a}_z - (1/i\omega\epsilon)J_x\nabla_T\cdot(\bar{h}_V^T \times \bar{a}_x)]dydz. \qquad (3.6c)$$

We now scalar multiply (2.2b) by $(\bar{a}_x \times \bar{e}_V^T)'$ and integrate with respect to y and z over the entire y-z plane. Following the procedure applied to (3.1) and noting that

$$\nabla_T\cdot(\bar{a}_x \times \bar{E}_T) = i\omega\mu H_x + M_x, \qquad (3.7a)$$

$$\left[H_x = -\frac{dh_{i-1,i}}{dx}H_y\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}, \qquad i = 1, 2, \ldots, m \qquad (3.7b)$$

$$\left[\bar{a}_x \times \bar{E}_T\cdot\bar{a}_y = -E_z\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \qquad (3.7c)$$

and

$$[(1/\omega\mu)\nabla_T\cdot(\bar{a}_x \times \bar{e}_V^T)]_{h_{i-1,i}^-}^{h_{i-1,i}^+} = 0, \qquad (3.8a)$$

$$[1 + (1/k^2)\nabla_T\nabla_T]\cdot(\bar{a}_x \times \bar{e}_V^T) = (u/\omega\epsilon)\bar{h}_V^T, \qquad (3.8b)$$

$$\bar{H}_T\cdot(\bar{a}_x \times \bar{e}_V^T) = -H_y(\bar{e}_V^T\cdot\bar{a}_z) + H_z(\bar{e}_V^T\cdot\bar{a}_y), \qquad (3.8c)$$

we can show that (2.2b) reduces to

$$-\frac{d}{dx}H^V(x, v', w') - iu'E^V(x, v', w')$$
$$+\sum_v\int_{-\infty}^{\infty}(H^VD_V^V + H^HD_V^H)dw = g^V(x), \qquad (3.9a)$$

in which

$$D_V^P(v', w', v, w) \equiv \int_{-\infty}^{\infty}h_T^P\cdot\frac{\partial}{\partial x}(\bar{a}_x \times \bar{e}_V^T)'dydz$$
$$-\int_{-\infty}^{\infty}\sum_{i=1}^{m}\frac{dh_{i-1,i}}{dx}[(\bar{h}_T^P\cdot\bar{a}_z)(\bar{e}_V^T\cdot\bar{a}_y)']_{h_{i-1,i}^-}^{h_{i-1,i}^+}dz,$$
$$P = V \text{ or } H, \qquad (3.9b)$$

and

$$g^V(x) = \int_{-\infty}^{\infty}\bar{J}_T\cdot\bar{e}_V^{T\prime}dydz - \int_{-\infty}^{\infty}\frac{1}{i\omega\mu}M_x\nabla_T\cdot(\bar{a}_x \times \bar{e}_V^T)'dydz. \qquad (3.9c)$$

On scalar multiplying (2.2a) and (2.2b) by $(\bar{h}_H^T \times \bar{a}_x)'$ and $(\bar{a}_x \times \bar{e}_H^T)'$ respectively, and integrating with respect to y and z over the entire y-z plane, we obtain in a similar manner

$$-\frac{d}{dx}E^H(x, v', w') - iu'H^H(x, v', w')$$
$$+\sum_v\int_{-\infty}^{\infty}(E^VC_H^V + E^HC_H^H)dw = f^H(x) \qquad (3.10a)$$

and

$$-\frac{d}{dx}H^H(x,v',w')-iu'E^H(x,v',w') + \sum_v \int_{-\infty}^{\infty}(H^V D_D^V + H^H D_H^H)\,dw = g^H(x), \quad (3.10\text{b})$$

in which

$$C_H^P(v',w',v,w) \equiv \int_{-\infty}^{\infty} \bar{e}_T^P \cdot \frac{\partial}{\partial x}(\bar{h}_H^T \times \bar{a}_x)'dydz + \int_{-\infty}^{\infty}\sum_{i-1}^{m}\frac{dh_{i-1,i}}{dx}(\bar{e}_T^P\cdot\bar{a}_z)(\bar{h}_H^T\cdot\bar{a}_y)'\Big|_{h_{i-1,i}^-}^{h_{i-1,i}^+}dz \quad (3.11\text{a})$$

$$D_H^P(v',w',v,w) \equiv \int_{-\infty}^{\infty}\bar{h}_T^P\cdot\frac{\partial}{\partial x}(\bar{a}_x\times\bar{e}_H^T)'dydz \quad (3.11\text{b})$$

for $P = V$ or H and

$$f^H(x) \equiv \int_{-\infty}^{\infty}\bar{M}_T\cdot\bar{h}_H^T\,dydz - \int_{-\infty}^{\infty}\frac{1}{i\omega\epsilon}J_x\nabla_T\cdot(\bar{h}_H^T\times\bar{a}_x)\,dydz \quad (3.11\text{c})$$

and

$$g^H(x) \equiv \int_{-\infty}^{\infty}J_z(\bar{e}_H^T\cdot\bar{a}_z)dydz - \int_{-\infty}^{\infty}\frac{1}{i\omega\epsilon}M_x\nabla_T\cdot(\bar{a}_x\times\bar{e}_H^T)\,dydz. \quad (3.11\text{d})$$

Expressing the scalar transforms E^P and H^P ($P = V$ or H) in terms of the wave amplitudes (2.7), we obtain the coupled first order ordinary differential equations for the vertically and horizontally polarized wave amplitudes. Thus,

$$-\frac{da^P}{dx} - iua^P = \sum_Q\sum_{v'}\int(S_{PQ}^{BA}a^Q + S_{PQ}^{BB}b^Q)\,dw' - A^P, \quad Q = V \text{ and } H, \quad (3.12\text{a})$$

and

$$-\frac{db^P}{dx} + iub^P \equiv \sum_Q\sum_{v'}\int(S_{PQ}^{AA}a^Q + S_{PQ}^{AB}b^Q)\,dw' + B^P, \quad Q = V \text{ and } H, \quad (3.12\text{b})$$

in which $P = V$ or H and we have interchanged the primed with the unprimed variables:

$$A^P = -(f^P + g^P)/2, \quad B^P = -(f^P - g^P)/2, \quad P = V \text{ or } H \quad (3.13)$$

The transmission scattering coefficients are defined as

$$S_{PQ}^{\alpha\beta}(v,w;v',w') = -[C_P^Q(v,w;v',w') + D_P^Q(v,w;v',w')]/2 \quad \alpha \neq \beta, \quad (3.14\text{a})$$

in which α and β are A or B and P and Q are V or H. The reflection scattering coefficients are

$$S_{PQ}^{\alpha\alpha}(v,w;v',w') = [C_P^Q(v,w;v',w') - D_P^Q(v,w;v',w')]/2$$

for $\alpha = A$ or B and P and $Q = V$ or H. (3.14b)

4. THE COUPLING COEFFICIENTS AND RECIPROCITY RELATIONSHIPS

In this section we obtain explicit expressions for the transmission and reflection scattering coefficients (3.14a) and (3.14b) respectively. Thus it is necessary to evaluate the integrals in (3.6b), (3.9b), (3.11a), and (3.11b) for C_Q^P and D_Q^P. To this end, we find it very useful to demonstrate first that our solutions for the forward and backward vertically and horizontally polarized wave amplitudes satisfy the reciprocity relationships. In view of the normalization used in our analysis, it can be shown that the reciprocity conditions to be satisfied are[2,3]

$$S_{PQ}^{BA}(v,w;v',w') = -S_{QP}^{AB}(v',w';v,w)N^P(v,w)/N^Q(v',w'), \quad (4.1\text{a})$$

$$S_{PQ}^{\alpha\alpha}(v,w;v',w') = S_{QP}^{\alpha\alpha}(v',w';w)N^P(v,w)/N^Q(v',w') \quad (4.1\text{b})$$

for P and Q equal to V or H and α equal to A or B. The normalization coefficients for the vertically and horizontally polarized waves are defined in Ref. 1. Since the orthogonality relationships (2.6) are satisfied for all y-z planes, it follows that

$$\frac{d}{dx}\int\bar{e}_T^P\cdot(\bar{h}_Q^T\times\bar{a}_x)'dydz = 0 = \int_{-\infty}^{\infty}\bar{e}_T^P\cdot\frac{\partial}{\partial x}(\bar{h}_Q^T\times\bar{a}_x)'dydz + \int_{-\infty}^{\infty}\bar{h}_Q^T\cdot\frac{\partial}{\partial x}(\bar{a}_x\times\bar{e}_T^P)dydz - \sum_{i=1}^{m}\int_{-\infty}^{\infty}\left(\frac{dh_{i-1,i}}{dx}\,\bar{e}_T^P\cdot(\bar{h}_Q^T\times\bar{a}_x)'\right)\Big|_{h_{i-1,i}^-}^{h_{i-1,i}^+}dy$$

$$= C_Q^P(v'w';v,w) + \frac{D_P^Q(v,w;v',w')N^Q(v',w')}{N^P(v,w)} \quad (4.2)$$

for P and $Q = V$ or H. Using the above relationship between the coupling coefficients C_Q^P and D_P^Q, we can show readily that the transmission and reflection scattering coefficients (3.14) satisfy the reciprocity relationships (4.1). Furthermore, in view of (4.2) it is not necessary to evaluate both C_Q^P and D_P^Q.

Using the orthogonal properties of the scalar functions $\psi^P(v,y)$ and $\phi(w,z)$ used in the definition of the basis functions $\bar{e}_T^P$ and $\bar{h}_T^P$,[1] we can show that for $v \neq v'$

$$C^{VV}(v'w';v,w) = \delta(w-w')\sum_{i=1}^{m}\left[\frac{Z^V(v,y)N^{V'}}{v'^2-v^2}\times\left(\psi^V(v,y)\frac{\partial^2\psi^V(v',y)}{\partial x\partial y} - \frac{\partial\psi^V(v,y)}{\partial y}\,\frac{\partial\psi^V(v',y)}{\partial x}\right)\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}$$

$$\equiv \delta(w-w')C^{VV}(v',v,w). \quad (4.3\text{a})$$

For the surface wave terms $v = v' = v^n$, we get

$$C^{VV}(v',w';v,w) = \tfrac{1}{2}\delta(w-w')\sum_{i=1}^{m}\left[\frac{dh_{i-1,i}}{dx}Z^V(v,y)[\psi^V]^2 + \frac{1}{2v^2}\frac{\partial Z^V}{\partial x}\left\{y\left(\frac{\partial\psi^V}{\partial y}\right)^2 + yv^2(\psi^V)^2 - \frac{\psi^V\partial\psi^V}{\partial y}\right\}\right]_{h_{i-1,i}^-}^{h_{i-1,i}^+}$$

$$\equiv \delta(w-w')C^{VV}(v',v,w), \quad (4.3\text{b})$$

in which

$$\frac{\partial Z^V}{\partial x} = Z^V\left(\frac{u^2-w^2}{u^2+w^2}\,\frac{1}{u}\frac{du}{dx} - \frac{1}{\epsilon}\frac{d\epsilon}{dx}\right) \quad (4.3\text{c})$$

and du/dx is determined by the modal equation (2.5). To obtain the explicit expression for D^{HH}, apply the following duality transformations to the expression for C^{VV} (4.3):

$$\psi^V \to \psi^H, \quad Z^V \to Y^H, \quad N^V \to N^H \quad \text{and} \quad \epsilon \to \mu \quad (4.3\text{d})$$

The coefficient $C_V^H = 0$ since $\bar{e}_T^H$ is orthogonal to $(\bar{h}_V^T \times \bar{a}_x)$. Integrating the expression for C_H^V, (3.11a), we can show that

$$C_H^V(v',w';v,w) = \delta(w-w')\sum_{i=1}^{m}\Bigg[\frac{iwN^{H'}}{uu'}\Big\{\psi^V(\partial\psi^H/\partial x)' + \frac{1}{k^2}\frac{\partial\psi^V}{\partial y}\left(\frac{\partial^2\psi^H}{\partial x\partial y}\right)' + \frac{dh_{i-1,i}}{dx}\frac{(u^2+w^2)'}{k^2}\frac{\partial\psi^V}{\partial y}\psi^{H'} - \frac{1}{u'}\frac{du'}{dx}\psi^V\psi^{H'} + \frac{1}{k^2(v^2-v'^2)}$$
$$\times\left((u^2+w^2)\frac{1}{\mu}\frac{d\mu}{dx} - \frac{v^2}{\mu'}\frac{d\mu'}{dx}\right)\times\left[\frac{\partial\psi^V}{\partial y}\left(\frac{\partial\psi^H}{\partial y}\right)' + v'^2\psi^V\psi^{H'}\right] + \frac{1}{(v^2-v'^2)}\frac{1}{\epsilon}\frac{d\epsilon}{dx}$$
$$\times\left[\frac{\partial\psi^V}{\partial y}\left(\frac{\partial\psi^H}{\partial y}\right)' + v^2\psi^V\psi^{H'}\right]\Big\}\Bigg]\begin{matrix}h^+_{i-1,i}\\h^-_{i-1,i}\end{matrix}$$
$$\equiv \delta(w-w')C_H^V(v',v,w). \tag{4.3e}$$

In the expressions derived for the coupling coefficients, (4.3), it is understood that, for $P = V$ or H,

$$\frac{\partial\psi^P}{\partial x} = \left(\sum_{i=1}^{m}\frac{dh_{i-1,i}}{dx}\frac{\partial}{\partial h_{i-1,i}} + \sum_{i=0}^{m}\frac{d\mu_i}{dx}\frac{\partial}{\partial\mu_i} + \frac{d\epsilon_i}{dx}\frac{\partial}{\partial\epsilon_i}\right)\psi^P. \tag{4.4}$$

In view of (4.3), (3.3) can be simplified by integrating with respect to w'; thus,

$$-\left(\frac{d}{dx} + iu\right)a^P(x,v,w) = \sum_Q\sum_{v'}[S_{PQ}^{BA}(v,v'w)a^Q(x,v',w) + S_{PQ}^{BB}(v,v',w)b^Q(x,v',w)] - A^P(x,v,w) \tag{4.5a}$$

and

$$-\left(\frac{d}{dx} - iu\right)b^P(x,v,w) = \sum_Q\sum_{v'}[S_{PQ}^{AA}(v,v',w)a^Q(x,v',w) + S_{PQ}^{AB}(v,v',w)b^Q(x,v',w)] + B^P(x,v,w), \tag{4.5b}$$

in which $P, Q = V$ or H, the symbol $\sum_{v'}$ is defined as in (2.3) and (2.4), and

$$S_{PQ}^{\alpha\beta}(v,v',w) = -[C_P^Q(v,v',w) + D_P^Q(v,v',w)]/2,$$
$$\alpha \neq \beta,\ \alpha,\beta = A \text{ or } B, \tag{4.5c}$$

and

$$S_{PQ}^{\alpha\alpha}(v,v'w) = [C_P^Q(v,v',w) - D_P^Q(v,v',w)]/2,$$
$$\alpha = A \text{ or } B. \tag{4.5d}$$

The above simplifications result from our assumption that $\partial h_{i-1,i}/\partial z = 0$. For the special case when the source distributions are independent of z (uniform infinite line sources), the source coefficients A^P and B^P [(3.13)] are proportional to $\delta(w)$. As a result, the parameter w can be eliminated from all the field expressions (the waves are propagating normal to the z axis, $w = 0$) and the fields are independent of z. In this case, all the cross polarization terms, which are proportional to w, (4.3e), vanish. This special case has been considered earlier in detail.[4] However, for the finite source distributions considered in this paper, none of the scattering coefficients (4.5) vanish. For instance, for the radiation field, $S_{VH}^{BA}(v,v',w)$ corresponds to the coupling between an incident forward propagating horizontally polarized wave propagating in the direction $(u',v',w)/k$ and a forward scattered vertically polarized wave propagating in the direction $(u,v,w)/k$. Similarly, $S_{HV}^{BB}(v,v',w)$ corresponds to the coupling between an incident backward propagating vertically polarized wave propagating in the direction $(u',v',w)/k$ and a forward scattered horizontally polarized wave in the direction $(u,v,w)/k$. Since the complete full wave expansion of the fields (2.3) and (2.4) consist of radiation, lateral wave, and surface wave terms, this analysis also accounts for the coupling between these constituents of the full wave expansions.

5. ITERATIVE SOLUTIONS

The coupled first order ordinary differential equations for the vertically and horizontally polarized wave amplitudes [(4.5)] are similar in form to those derived for source free waveguides with nonuniform cross sections.[5,6] In the present analysis, however, the wave spectrum is not only discrete (surface waves or trapped waveguide modes) but also continuous (the radiation and lateral wave terms). In addition, the excitation from local sources is accounted for directly in the coupled differential equations through the terms A^P and B^P [(3.13)]. The scattering coefficients $S_{PQ}^{\alpha\beta}$, (4.5), are in general arbitrary functions of the independent variable x. Thus, these sets of differential equations are often solved using sophisticated numerical methods. In this section, however, we consider iterative solutions that are suited to physical interpretation.

To obtain the first order iterative solutions, we ignore the wave coupling in (4.5) and solve the resulting uncoupled equations subject to the boundary condition

$$a^P(x\to-\infty) = 0 \quad\text{and}\quad b^P(x\to\infty) = 0. \tag{5.1a}$$

Thus, it follows that

$$a^P(x,v,w) = \int_{-\infty}^{x}\exp\left(-i\int_{x'}^{x}u\,dx''\right)A^P(x',v,w)\,dx' \tag{5.1b}$$

and

$$b^P(x,v,w) = \int_{x}^{\infty}\exp\left(i\int_{x'}^{x}u\,dx''\right)B^P(x',v,w)\,dx',$$
$$P = V \text{ or } H. \tag{5.1c}$$

In the above expressions the parameter u may in general be a function of x. The first order iterative solutions (5.1) correspond to the primary (unscattered) fields which can be determined using (2.3a) and (2.4a). When the sources are far from the observation point, the radiation and lateral wave terms may be integrated readily using the steepest descent method.[1] These first order iterative solutions are now substituted on the right-hand side of (4.5) and the source terms set equal to zero. The resulting differential equations are integrated to give the following second order iterative solutions for the scattered wave amplitudes[4]:

$$a^P(x,v,w) = -\int_{-\infty}^{x}\exp\left(-i\int_{x'}^{x}u\,dx''\right)\times\left(\sum_Q\sum_{v'}S_{PQ}^{BA}(v,v',w)a^Q(x',v',w) + S_{PQ}^{BB}(v,v'w)b^Q(x'v',w\right)dx' \tag{5.2a}$$

and

$$b^P(x,v,w) = \int_{x}^{\infty}\exp\left(i\int_{x'}^{x}u\,dx''\right)\times\left(\sum_Q\sum_{v'}S_{PQ}^{AA}(v,v',w)a^Q(x',v',w) + S_{PQ}^{AB}(v,v',w)b^Q(x',v',w)\right)dx'. \tag{5.2b}$$

This iterative procedure has been carried out to compute the vertically polarized fields excited by infinite line sources over nonuniform stratified earth.[7]

6. CONCLUDING REMARKS

Through the use of generalized field transforms,[1] Maxwell's equations for inhomogeneous media with arbitrary source distributions are converted into a set of first order ordinary differential equations for the forward and backward vertically and horizontally polarized wave amplitudes. Since the stratified medium is nonuniform, the wave amplitudes are coupled. Explicit expressions for the coupling coefficients are derived and the results are shown to satisfy the reciprocity relationships in electromagnetic theory. An iterative method to solve the coupled equations for the wave amplitudes is outlined.

The solutions derived in this paper can be applied to a very broad class of problems such as propagation in the vicinity of a coast line where both the height and electromagnetic parameters of the earth's surface vary. The analysis can also be applied to problems of remote sensing such as mapping the water table or detection of buried objects of finite cross section (Fig. 2). Exact boundary conditions are imposed in the derivation of these solutions and, since the field expansions do not in general converge uniformly at all points, orders of integration (summation) and differentiation are not interchanged.

It is interesting to note that when surface impedances are used to characterize the bounding media, certain additional terms arise in the expression for the scattering coefficients, (4.5). These terms which are proportional to the surface impedances or admittances[6,8] have no counterpart in our present analysis in which exact boundary conditions are imposed. Hence, these spurious terms, which tend to vanish as the conductivity of the bounding media increases, should be disregarded when the surface impedance concept is employed. In addition, when surface impedances are used in the analysis, the lateral wave contributions to the full wave expansions, (2.3) and (2.4), vanish. Thus boundary value problems should be carefully examined before employing impedance boundary conditions.

ACKNOWLEDGMENTS

The work reported in this paper was supported by the National Science Foundation and the Engineering Research Center of the University of Nebraska. The author wishes to thank J. R. Wait, F. G. Ullman, and L. Lipsky for their comments. The manuscript was typed by Mrs. E. Everett.

[1]E. Bahar, J. Math. Phys. **14**, 1502 (1973).
[2]E. Bahar, J. Math. Phys. **14**, 1024 (1973).
[3]D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis* (Pergamon, New York, 1967).
[4]E. Bahar, J. Math. Phys. **14**, 1030 (1973).
[5]S. A. Schelkunoff, Bell Syst. Tech. J. **34**, 995 (1955).
[6]R. L. Gallawa, J. Res. Natl. Bur. Stand. (U.S.) D: Radio Propagat. **68**, 1201 (1964).
[7]E. Bahar and G. Govindarajan, "Radiation by a Line Source Over Nonuniform Stratified Earth, J. Geophys. Res. 78 (2), 393 (1973).
[8]E. Bahar, IEEE Trans. Antennas Propag. **AP20** (3), 354 (1972).

Generalized semidirect product in group unifications*

M. Vujičić and F. Herbut

Institute "Boris Kidrich"-Vinča, Belgrade, and Institute of Physics, Belgrade, Yugoslavia

Dj. Šijački†

Institute "Boris Kidrich"-Vinča, Belgrade, Yugoslavia

(Received 3 January 1973; revised manuscript received 27 March 1973)

The case of a unification E of two given groups K and H when K is invariant in E (i-unification) is investigated. Necessary and sufficient conditions for the existence of E are found. It is shown that any i-unification can be constructed as a factor group of the simplest i-unification, i.e., the semidirect product $K \circledS_\tau H$ with respect to a given homomorphism $\tau: H \to \mathrm{Aut}(K)$. This construction is called the generalized semidirect product (GSP). The i-unifications are put in correspondence with the group extensions via the GSP. The significance of the GSP is illustrated in elementary particle and solid state physics.

1. INTRODUCTION

The problem of coupling two partial symmetries of a physical system is encountered in different fields of physics: combination of the Lorentz group with the translational group into the Poincaré group; coupling of the point groups and the subgroups of the translational group into the crystallographic space groups; attempts to unify internal and space–time symmetries of elementary particles, etc. This task has sometimes a trivial solution in terms of a semidirect product, like in the first mentioned example, as well as in all the 73 symmorphic space groups.[1] In the remaining space groups (there are 157 of them) the coupling is nontrivial: The point groups are nontrivially extended by translational subgroups.[2]

The idea of coupling an internal, e.g., the $SU(3)$ symmetry group, with the Poincaré group comes from the need to explain the fact that $SU(3)$-multiplets of elementary particles have common spin and parity and different masses. Attempts to find a satisfactory coupling proceeded along the lines of group unifications and group extensions[3] (precise definitions of these concepts in Sec. 2). In the unification approach one treats the Poincaré group as a subgroup of a higher symmetry group E, and in the extensions it is a factor group. In the trivial case of a semidirect product these two concepts coincide.

In our previous paper[4] we have discussed extensions of a group G by a group K when a homomorphism $\sigma: G \to \mathrm{Aut}(K)$ exists [$\mathrm{Aut}(K)$ being the group of all automorphisms in K]. All such extensions were shown to possess a generalized semidirect product (GSP) form, which may considerably simplify their construction and the evaluation of their irreducible representations (on the basis of well-known methods for finding representations of semidirect products[5]).

Recent research[6] has revealed that 101 of the 157 nonsymmorphic space groups can be put into a GSP form. Though for them the point groups are not exact symmetries (i.e., they are not subgroups of the space groups), there appear new exact symmetry groups H as the second factors in the GSP expression [cf. Eq. (8) below]. These 101 space groups become in this way unifications of their translational and H-subgroups. The latter have a clear meaning in terms of properties of crystals.

It is a general feature of all extensions E of G by K expressible as GSP's that they are also unifications of K and H. The purpose of this work is to find the widest class of extensions obtainable as GSP's. It is substantially wider than the class treated in Ref. 4. The latter contains, in the terminology of this paper, only central GSP's [not to be confused with central extensions—Eq. (12) below].

In Sec. 2 we establish an equivalence between all extensions and the relevant unifications (which we call i-unifications). Lacking in the general case the mentioned homomorphism $\sigma: G \to \mathrm{Aut}(K)$, we found that unification theory, in which instead of G, the group H is the starting point, gave a more suitable mathematical framework in which we derived the general theory of GSP. We have reasons to believe that the groups H are endowed with important physical meaning, and the study of these groups is far more natural in terms of unifications.

2. i-UNIFICATIONS AND EXTENSIONS

Let K and H be two given groups. A unification of K and H is a third group E which is the product of two subgroups $i(K)$ and $u(H)$ that are isomorphic to K and H respectively.

Definition 1: A group E is an *i-unification* of two groups K and H if $i(K)$ is invariant in E, i.e.,

$$E = i(K)u(H), \qquad i(K) \lhd E, \qquad u(H) < E, \tag{1}$$

where $<$ and $\lhd$ denote the subgroup and the invariant subgroup relations respectively.

On the other hand, a group E is an extension of a group G by a group K if there is an invariant subgroup $i(K)$ in E, such that $i(K) \cong K$, and $E/i(K) \cong G$, i.e., if one has an exact sequence

$$1 \to K \xrightarrow{i} E \to G \to 1. \tag{2}$$

The extension E can be defined as the set $\{(\alpha, a) \mid \alpha \in K, a \in G\}$ with the composition law

$$(\alpha, a)(\beta, b) = (\alpha\Psi[a](\beta)\omega(a, b), ab), \tag{3}$$

where $\Psi: G \to \mathrm{Aut}(K)$, $\omega: G \times G \to K$ give the system of automorphisms and the system of factors respectively satisfying certain necessary and sufficient conditions.[7,8]

Lemma 1: (A) An arbitrary i-unification E of K and H is also an extension of G by K, where G is any group isomorphic to $E/i(K) \cong u(H)/i(K_0)$, and $K_0 < K$ is defined by $i(K_0) = i(K) \cap u(H)$.

(B) An arbitrary extension E of G by K can be

Copyright © 1973 by the American Institute of Physics

viewed as an i-unification of K and H, where H is the corresponding extension of G by K_0, the latter being any subgroup of K invariant under all $\Psi[a]$ and satisfying $\omega(a,b) \in K_0, \forall a,b \in G$.

Proof: (A) Obviously $i(K_0)$ is an invariant subgroup of $u(H)$, so that $u(H)/i(K_0)$ exists. The isomorphism $E/i(K) \cong u(H)/i(K_0)$ is due to the fact that each coset in E with respect to $i(K)$ contains precisely one coset of $u(H)$ with respect to $i(K_0)$, as easily established.

Let $\{h(a) \mid a \in G\}$ be any set of representatives one from each coset of $u(H)$ with respect to $i(K_0)$. Then one defines the maps Ψ and ω by

$$\Psi[a](\alpha) = i^{-1}(h(a)i(\alpha)h(a)^{-1}), \quad \forall \alpha \in K, \forall a \in G, \tag{4a}$$

$$h(a)h(b) = i(\omega(a,b))h(ab), \quad \forall a,b \in G, \tag{4b}$$

which satisfy the mentioned necessary and sufficient conditions.[8] The groups E and H can now be equivalently written as $\{(\alpha,a) \mid \alpha \in K, a \in G\}$ and $\{(\gamma,a) \mid \gamma \in K_0, a \in G\}$ respectively, with (3) as their composition law. It is possible to use the same Ψ and ω, defined by (4), both for E and H because $i(K_0)$ is invariant under all automorphisms $\Psi[a]$, and all $\omega(a,b) \in i(K_0)$.

(B) Let E be an extension of G by K with given mappings Ψ and ω. There always exists a subgoup K_0 with the required properties (at least $K_0 = K$), which in its turn implies the existence of a group $H < E$, which is an extension of G by K_0 with the same ω and each $\Psi[a]$ restricted to K_0. It is easily seen that $E = i(K)H$, which makes it an i-unification of K and H. QED

3. i-UNIFICATION AS GENERALIZED SEMIDIRECT PRODUCT

In this section we discuss necessary and sufficient conditions for two groups K and H to have an i-unification E.

To obtain necessary conditions, let us assume that E, K, H, i, and u satisfying Eq. (1) are given. The intersection of $i(K)$ and $u(H)$ is, in general, nontrivial, and it defines two subgroups $K_0 < K$ and $H_0 \lhd H$ by

$$i(K_0) = u(H_0) = i(K) \cap u(H). \tag{5}$$

This equation implies an isomorphism $l: K_0 \to H_0$, which is $l \equiv u^{-1} \circ i$.

Owing to the fact that $i(K) \lhd E$, the conjugations in E give rise to a homomorphism $\tau: H \to \mathrm{Aut}(K)$, i.e.,

$$\tau[x](\alpha) = i^{-1}(u(x)i(\alpha)u(x)^{-1}), \quad \forall \alpha \in K, \forall x \in H. \tag{6}$$

Restricting α to K_0 or x to H_0, Eq. (6) reduces to the respective equations

$$\tau[x](\gamma) = l^{-1}(xl(\gamma)x^{-1}), \quad \forall \gamma \in K_0, \forall x \in H, \tag{7a}$$

$$\tau[l(\gamma)](\alpha) = \gamma\alpha\gamma^{-1}, \quad \forall \alpha \in K, \forall \gamma \in K_0. \tag{7b}$$

Equations (7a) and (7b) are necessary conditions which we show now to be also sufficient.

Theorem: Let the groups K and H have the subgroups $K_0 < K$, $H_0 \lhd H$, isomorphic via $l: K_0 \to H_0$, and let a homomorphism $\tau: H \to \mathrm{Aut}(K)$ be given so that Eqs. (7a) and (7b) are satisfied. Then the *generalized semidirect product* (GSP)

$$E \equiv (K \circledT H)/K_0', \tag{8}$$

where

$$K_0' = \{(\gamma, l(\gamma^{-1})) \mid \gamma \in K_0\}, \tag{9}$$

is an i-unification of K and H, with $i(\alpha) = (\alpha,1)K_0'$, $\forall \alpha \in K$ and $u(x) = (\epsilon,x)K_0'$, $\forall x \in H$ [cf. (1)], 1 and ϵ being the unit elements in H and K respectively.

Proof: The semidirect product $K \circledT H = \{(\alpha,x) \mid \alpha \in K, x \in H\}$ has the composition law defined by $(\alpha,x)(\beta,y) = (\alpha\tau[x](\beta), xy)$. By making use of Eqs. (7a) and (7b) it is straightforward to show that its subset K_0' is closed under multiplication, inversion, and conjugation by any element of $K \circledT H$, so that it is an invariant subgroup.

It is a known property of the semidirect product that the isomorphisms $i'(\alpha) \equiv (\alpha,1)$, $\forall \alpha \in K$, and $u'(x) \equiv (\epsilon,x)$, $\forall x \in H$ are such that $K \circledT H$ can be viewed as an i-unification of K and H. Since $i = w \circ i'$ and $u = w \circ u'$, where $w: K \circledT H \to E$ is the natural homomorphism $[\mathrm{Ker}(w) = K_0']$ preserving all relations in (1), the group E is also an i-unification of K and H. QED

Remark 1: It is noteworthy that, for given K, H and τ, the simplest possible i-unification $K \circledT H$ contains among its factor groups all the other i-unifications.

Corollary: In the GSP given by (8) relations (5) are satisfied.

Proof: The set $i(K) \cap u(H)$ consists of those cosets in $K \circledT H$ which can be simultaneously written as $(\alpha,1)K_0'$ and as $(\epsilon,x)K_0'$. The former can be expressed as $\{(\alpha\gamma, l(\gamma^{-1})) \mid \gamma \in K_0\}$, and it contains (ϵ,x) if and only if $\alpha \in K_0$ and $x \in H_0$, which implies (5). QED

Remark 2: If K, H, K_0, and H_0 are generated each by a subset of elements (in particular by a finite number of generators), then for the possibility to construct the GSP [cf. (8)] it is sufficient to check Eqs. (7a) and (7b) only for the generating elements.

4. SPECIAL CASES OF *GSP* AND EXAMPLES

Different choices of the homomorphism τ and the subgroup K_0 provide us with important special cases of GSP.

Definition 2: If K_0 is the center C of K or its subgroup C_0, then the corresponding GSP we call *central* GSP; otherwise, a *noncentral* one.

In Ref. 4 we have shown that any extension of G by K can be put in the central GSP form if Ψ is a homomorphism (denoted by σ).[9] The following lemma establishes a natural connection between the possible homomorphic property of Ψ and the corresponding central character of the GSP.

Lemma 2: If E is an extension of G by K with $\Psi: G \to \mathrm{Aut}(K)$ being a homomorphism, then it has a central GSP form, and vice versa, a central GSP always implies Ψ to be a homomorphism.

Proof: The center C is invariant under every automorphism in K. Conjugating $i(\alpha)$, $\alpha \in K$ with $h(a)h(b)$, $a,b \in G$, and making use of (4a) and (4b), one arrives at

$$\Psi[a](\Psi[b](\alpha)) = \omega(a,b)\Psi[ab](\alpha)\omega(a,b)^{-1}.$$

Oviously Ψ is a homomorphism if and only if $\omega(a,b) \in C$, $\forall a,b \in G$. QED

Remark 3: In the general case replacing $u(x)$ in (6) by $i(\gamma)h(a)$, $\gamma \in K_0, a \in G$, and using (4a), one gets the connection between Ψ and τ:

$$\tau[x](\alpha) = \gamma\Psi[a](\alpha)\gamma^{-1}. \tag{10}$$

For central GSP's and only for them (10) simplifies to

$$\tau = \sigma \circ n, \tag{11}$$

with $\sigma = \Psi$ and $n: H \to G$.

Within the central GSP's there are important special cases due to the possibility of having τ and/or K_0 trivial. When τ is trivial, we talk about the generalized direct product (GDP) known in the literature as central extensions[11]:

$$E = (\underset{\cdot}{K} \otimes H)/C_0' \tag{12}$$

[cf. Eqs. (8) and (9), $C_0 = K_0$ is a central subgoup of K].

Remark 4: If one has two groups K and H and one wants to check if a GDP can be built out of them, then beside τ being trivial, K_0 must be central in K [condition (7b)], and H_0 central in H [condition 7(a)].

The best known example of GDP was proposed by Michel[12]:

$$E = (S \otimes \overline{P})/Z_2', \tag{13}$$

where S is a group of internal symmetries and $\overline{P}$ is the covering of the Poincaré group.

If K_0 is trivial, the GSP reduces to the semidirect product (SP). Among the well-known examples for SP are the symmorphic space groups in solid state physics.

The simplest GSP is the direct product (DP), which occurs when both τ and K_0 are trivial.

Remark 5: Let K, H, K_0 a central subgroup of K, $H_0 \lhd H$, and an isomorphism $l: K_0 \to H_0$ be given (making H_0 Abelian). In order to construct a central GSP it is less practical to look for a τ which satisfies (7a) and (7b) than for a homomorphism $\sigma: H/H_0 \to \mathrm{Aut}(K)$ restricted only by

$$\sigma[a](\gamma) = l^{-1}(h(a)l(\gamma)h(a)^{-1}), \quad \forall \gamma \in K_0, \forall a \in H/H_0, \tag{14}$$

where $h(a)$ are arbitrarily chosen representatives one from each coset of H with respect to H_0. The connection between σ and τ is given by (11) and its inverse

$$\sigma = \tau \circ h. \tag{15}$$

An instructive illustration of the central GSP which does not reduce either to the GDP or to the SP is

$$E = \{[U(1) \otimes SU(3)/Z_3] \textcircled{\tau} Z_4(X)\}/Z_2'. \tag{16}$$

If X is the time reversal,[13] then the nontrivial intersection of $i(K)$ and $u(H)$ implies the following relation between the baryon quantum number B and the spin J:

$$(-1)^B = (-1)^{2J}.$$

[For hadrons the same relation follows,[12] due to Z_2', also from (13).]

When X is the charge conjugation,[4] the nontrivial τ reflects the incompatibility of the charge conjugation quantum number and the additive quantum numbers of the internal symmetry group.

Whenever one combines a symmetry group with an involutive discrete symmetry, the minimal extensions [i.e., those for which G in (2) is Z_2] are essential. It is shown[8] that every minimal extension is an easily obtained GSP and that central as well as non-central GSP's occur naturally among them.

*This work is supported by the Republican Community for Scientific Research of SR Serbia.

†Present address: Department of Physics, Duke University, Durham, N.C. 27706.

[1]L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967).
[2]E. Ascher and A. Janner, Helv. Phys. Acta **38**, 551 (1965).
[3]G. C. Hegerfeldt and J. Hennig, Fortschr. Phys. **16**, 491 (1968).
[4]Dj. Šijački, M. Vujičić, and F. Herbut, J. Math. Phys. **13**, 1755 (1972).
[5]G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[6]I. Božović, M. Sc. Thesis (Belgrade University, 1972).
[7]A. G. Kurosh, *The Theory of Groups* (Chelsea, New York, 1955), Chap. XII.
[8]F. Herbut, M. Vujičić, and Dj. Šijački, J. Math. Phys. 14, 1121 (1973).
[9]Actually, in Ref. 4 we have proved this only for $\Psi = \sigma$ having at most one image in every coset of Aut(*K*) with respect to the subgroup of all inner automorphisms. This additional restriction, which Lemma 2 in the present paper proves to be unnecessary, was due to our desire to build our approach entirely on that of Michel (Ref. 10).
[10]L. Michel, "Invariance in Quantum Mechanics and Group Extensions", in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).
[11]L. Michel, "Relativistic Invariance", in *Particle Symmetries and Axiomatic Field Theory*, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966).
[12]L. Michel, Phys. Rev. 137, B405 (1965).
[13]F. von Kamber and N. Straumann, Helv. Phys. Acta **37**, 563 (1964).

An analytic approximation method for the one-dimensional Schrödinger equation.II

W. Hecht

Fakultät für Physik der Universität Freiburg, D-78 Freiburg, Hermann Herder-Str. 3, Germany
(Recieved 20 November 1972)

The problem with three classical turning points which must be solved in order to discuss "orbiting" collisions in molecular physics or "quasimolecular states" occurring in α-nucleus and heavy ion scattering is treated, using the approximation method already discussed in Paper I. The general statement of Paper I concerning complex conjugate turning points is thereby confirmed and the corresponding formulas known from Langer's method can be obtained as special cases.

I. FORMULAS RELEVANT FOR THE CONNECTION OF WHITTAKER FUNCTION

The approximations to the solutions of the one-dimentional Schrödinger equation with two turning points at ξ_1 and ξ_2 are, as shown in Paper I (Ref. 1),

$$(x')^{-1/2} W_{\pm\kappa,1/4}(2ixe^{-i\pi(1\mp 1/2)}), \tag{1}$$

with

$$\int_0^x \left(1 + \frac{2i\kappa}{u}\right)^{1/2} du = \int_{\bar{\xi}}^{\xi} k(s)ds, \quad \bar{\xi} - \tfrac{1}{2}(\xi_1 + \xi_2) = \pm 0, \tag{2}$$

$$|\kappa| = \frac{1}{2\pi}\left|\int_{\xi_1}^{\xi_2} |k(s)|ds\right| . \tag{3}$$

It should also be remembered that the argument of κ in a wavefunction at points $\xi < \bar{\xi}$ differs by an amount of 2π from that at points $\xi > \bar{\xi}$ for complex conjugate turning points. In order to avoid confusion, it is therefore useful to write

$$\bar{\kappa} = \kappa_{|_{\xi<\bar{\xi}}} \quad \text{with} \quad \arg\bar{\kappa} = 2\pi + \arg\kappa_{|_{\xi>\bar{\xi}}} \tag{4}$$

in all expressions where this difference may become relevant.

For $|x| \to 0$, the Whittaker function can be approximated (see, e.g., Ref. 2) by

$$W_{\kappa,1/4}(2ix) \approx \sqrt{\pi}\left(\frac{(2ix)^{1/4}}{\Gamma(\frac{3}{4} - \kappa)} - 2\,\frac{(2ix)^{3/4}}{\Gamma(\frac{1}{4} - \kappa)}\right), \qquad |x| \to 0, \quad |\kappa| \neq 0. \tag{5}$$

Inserting the analogous expansions which can be derived from Eq. (2),

$$(x')^{-1/2}(2ix)^{1/4} \approx (4e^{i\pi}\kappa)^{1/4}\frac{1}{\sqrt{K}}, \quad (x')^{-1/2}(2ix)^{3/4} \approx (4e^{-i\pi}\kappa)^{-1/4}\,\frac{1}{\sqrt{K}}\int_{\bar{\xi}}^{\xi} K(s)ds, \quad \xi \to \bar{\xi}$$

one obtains an expression that can be considered as the expansion of

$$\begin{aligned}(x')^{-1/2} W_{\kappa,1/4}(2ix) &= \sqrt{\frac{2\pi}{k}}\left[\frac{(\kappa e^{i\pi})^{1/4}}{\Gamma(\frac{3}{4} - \kappa)} \cos\left(\int_{\bar{\xi}}^{\xi} kds\right) - \frac{(\kappa e^{-i\pi})^{-1/4}}{\Gamma(\frac{1}{4} - \kappa)} \sin\left(\int_{\bar{\xi}}^{\xi} kds\right)\right] \\ &= \sqrt{\frac{\pi}{2k}}\left[\Omega(\kappa e^{i\pi}) \exp\left(-i\int_{\bar{\xi}}^{\xi} kds\right) + i\Omega(\kappa e^{-i\pi}) \exp\left(i\int_{\bar{\xi}}^{\xi} kds\right)\right]\end{aligned} \tag{6}$$

with

$$\Omega(\kappa) = \frac{1}{(\kappa)^{-1/4}\Gamma(\frac{3}{4} + \kappa)} + \frac{1}{(\kappa)^{1/4}\Gamma(\frac{1}{4} + \kappa)} \tag{7}$$

The analogous expression

$$(x')^{-1/2} W_{-\kappa,1/4}(2ixe^{-i\pi}) = \sqrt{\frac{\pi}{2k}}\left[\Omega(\kappa) \exp\left(i\int_{\bar{\xi}}^{\xi} kds\right) + i\Omega(\kappa e^{-2i\pi}) \exp\left(-i\int_{\bar{\xi}}^{\xi} kds\right)\right] \tag{8}$$

can be obtained from Eq. (6) by the substitution $k(\xi) \to e^{-i\pi}k(\xi)$.

For later use the circuit relations for the Whittaker functions

$$W_{\kappa,1/4}(2ixe^{-2i\pi}) = -\frac{2\pi i}{\Gamma(\frac{3}{4} - \kappa)\Gamma(\frac{1}{4} - \kappa)}\, e^{-i\pi\kappa}W_{-\kappa,1/4}(2ixe^{-i\pi}) + e^{-2i\pi\kappa}W_{\kappa,1/4}(2ix), \tag{9}$$

$$W_{-\kappa,1/4}(2ixe^{-3i\pi}) = -\frac{2\pi i}{\Gamma(\frac{3}{4} + \kappa)\Gamma(\frac{1}{4} + \kappa)}\, e^{-i\pi\kappa}W_{-\kappa,1/4}(2ix) - e^{-2i\pi\kappa}W_{-\kappa,1/4}(2ixe^{-i\pi}), \tag{10}$$

and their asymptotic formulas,

$$(x')^{-1/2} W_{-\kappa,1/4}(2ixe^{-i\pi-2i\pi n}) \simeq \frac{e^{2i\pi n\kappa}}{\hat{\Omega}(\kappa)}\frac{1}{\sqrt{k}} \exp\left(i\int_{\bar{\xi}}^{\xi} kds\right), \tag{11}$$

$$(x')^{-1/2} W_{\kappa,1/4}(2ixe^{-2i\pi n}) \simeq e^{i\pi\kappa-2i\pi n\kappa}\hat{\Omega}(\kappa)\,\frac{1}{\sqrt{k}} \times \exp\, -i\int_{\bar{\xi}}^{\xi} kds\,, \quad \hat{\Omega}(\kappa) = e^{-\kappa}(\kappa)^{\kappa}, \tag{12}$$

valid for $|\arg(e^{-2i\pi n}\int_{\bar{\xi}}^{\xi} kds)| \angle \pi$, should also be remembered.

II. PROBLEMS WITH THREE CLASSICAL TURNING POINTS

The shape and the size of the maximum and the minimum of the effective potential, as illustrated in Fig. 1, can be assumed to be arbitrary (indicated symbolically by the dashed line), except that no more than two turning points should come close together. It is therefore assumed that there is an energy value E_0 for which the turning points are all real and well separated from each other.

In the following all relevant variables, such as $\kappa, x, \bar{\xi}$, which lie in the region of the first (last) two turning points shall be labeled by the index 1 (2). For instance, choosing

$$\arg K = 0 \quad \text{for } \xi \to \infty \tag{13}$$

Copyright © 1973 by the American Institute of Physics

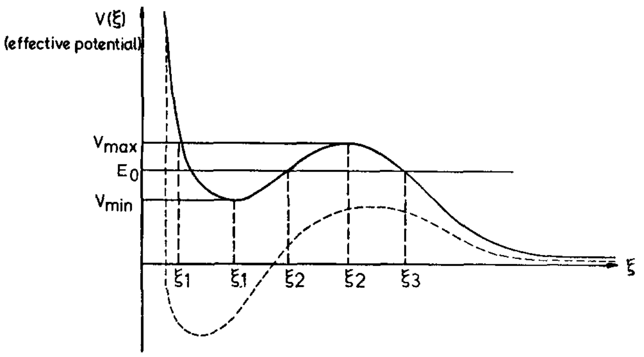

FIG. 1.

and

$$\arg(\xi - \bar{\xi}) = \pi \quad \text{for } \xi < \bar{\xi},$$

which leads to

$$\arg k = \tfrac{3}{2}\pi \quad \text{for } \xi \to 0, \tag{14}$$

one would be led, by considerations quite analogous to those of Paper I [Ref. 1; compare with Eq. (16) there] to the assignments listed in Eq. (20) for $\arg\kappa_1$ and $\arg\kappa_2$.

It is now desirable to connect the Whittaker function which approximates the physical wave function in the region of ξ_1 and ξ_2 with the corresponding Whittaker function in the region of ξ_2 and ξ_3, which results in a formula like

$$(1/\sqrt{|k|})\exp\Big(-\int_{\xi}|k|\,ds\Big) \to [\mathcal{L}(k)/\sqrt{|k|}]\exp\Big[-i\Big(\int^{\xi}|k|\,ds - \pi/4\Big)\Big] + \cdots. \tag{15}$$

From here, all relevant physical information could be obtained. In (15) it is understood, that the left (right) side should be used as approximation on the left (right) side of all turning points and that the missing integration limit should be taken to be the smallest (greatest) real turning point.

First we shall construct a connection formula for the corresponding Whittaker functions. By using Eq. (13), (14) and the asymptotics equation (11), this is seen to be

$$(x_1')^{-1/2}W_{-\kappa,1/4}(2ix_1e^{-3i\pi}) \to (-i)\,\frac{f(\kappa_1,\kappa_2)}{\hat{\Omega}(\kappa_1)\hat{\Omega}(\kappa_2)}(x_2')^{-1/2}W_{\kappa,1/4}(2ix_2) + \cdots. \tag{16}$$

The factors multiplying the yet unknown function $f(\kappa_1, \kappa_2)$ have been added for convenience.

There are possibilities of arriving at Eq. (16):

(A) Eqs. (8) and (11) can be used to relate the Whittaker functions to the corresponding WKB solutions at $\bar{\xi}_1$ which should then be identified with one another. Thus we are first led to

$$\begin{aligned}(x_1')^{-1/2}W_{-\kappa,1/4}(2ix_1e^{-3i\pi}) &\to \sqrt{\frac{\pi}{2}}(x_2')^{-1/2}\Big[e^{i\pi(\kappa_1-\kappa_2)}\Omega(\kappa_1e^{-2i\pi})\hat{\Omega}(\bar{\kappa}_2)\\ &\times W_{-\kappa_2,1/4}(2ix_2e^{-3i\pi}) - ie^{-i\pi\kappa_1}\frac{\Omega(\kappa_1)}{\hat{\Omega}(\bar{\kappa}_2)}\\ &\times W_{\kappa_2,1/4}(2ixe^{-2i\pi})\Big]\end{aligned} \tag{17}$$

By using the circuit relations (9), (10) it is now easy to obtain Eq. (16) with

$$\begin{aligned}f(\kappa_1,\kappa_2) = \sqrt{\frac{\pi}{2}}\Big(&e^{-2i\pi\kappa_2}\frac{2\pi\hat{\Omega}(\kappa_2)\hat{\Omega}(\bar{\kappa}_2)}{\Gamma(\frac{1}{4}+\kappa_2)\Gamma(\frac{3}{4}+\kappa_2)}\\ &\times e^{i\pi\kappa}\hat{\Omega}(\kappa_1)\Omega(\kappa_1e^{-2i\pi})\\ &+ e^{-i\pi\kappa_1}\hat{\Omega}(\kappa_1)\Omega(\kappa_1)e^{-2i\pi\kappa_2}\frac{\hat{\Omega}(\kappa_2)}{\hat{\Omega}(\bar{\kappa}_2)}\Big).\end{aligned} \tag{18}$$

(B) Similarly, connecting the two Whittaker functions at $\bar{\xi}_2$, we obtain Eq. (16) with

$$\begin{aligned}f(\kappa_1,\kappa_2) = \sqrt{\frac{\pi}{2}}\Big(&\frac{2\pi e^{-i\pi\kappa_1}\hat{\Omega}(\kappa_1)}{\Gamma(\frac{1}{4}+\kappa_1)\Gamma(\frac{3}{4}+\kappa_1)}\\ &e^{-2i\pi\kappa_2}\hat{\Omega}(\kappa_2)\Omega(\kappa_2) - e^{-i\pi\kappa_1}\hat{\Omega}(\kappa_2)\Omega(\kappa_2e^{-2i\pi})\Big).\end{aligned} \tag{19}$$

It can be shown [see Eqs. (20b), (20c) below] that these two formulas (18), (19) are identical for $E \approx E_0$ and can therefore be looked upon as continuations of each other. The final connection formula (15) can now be obtained by inserting the asymptotics of the Whittaker functions into Eq. (16) and using the following relations:

$$\left.\begin{aligned}e^{i\pi\kappa_1} &= \exp\Big(i\int_{\xi_1}^{\bar{\xi}_1}k\,ds\Big) = \exp\Big(i\int_{\bar{\xi}_1}^{\xi_1}k\,ds\Big)\\ e^{i\pi\kappa_2} &= \exp\Big(i\int_{\xi_2}^{\bar{\xi}_2}k\,ds\Big) = \exp\Big(i\int_{\bar{\xi}_2}^{\xi_3}k\,ds\Big)\end{aligned}\right\},\quad V_{\min} \le E \le V_{\max},$$

$$\begin{aligned}e^{2i\pi\kappa_1+i\pi\kappa_2} &= \exp\Big(i\int_{\xi_1}^{\xi_2}k\,ds\Big)\\ &= \exp\Big(-i\int_{\xi_1}^{\bar{\xi}_2}|k|\,ds\Big),\quad E > V_{\max},\end{aligned}$$

$$\begin{aligned}e^{i\pi\kappa_1+2i\pi\kappa_2} &= \exp\Big(i\int_{\xi_1}^{\xi_3}k\,ds\Big)\\ &= \exp -\Big(\int_{\xi_1}^{\xi_3}|k|\,ds\Big),\quad E < V_{\min},\end{aligned}$$

which again can be read off from Paper I and the assignments (13), (14).

Remembering

$$|\kappa_1| = \frac{1}{2\pi}\Big|\int_{\xi_1}^{\xi_2}|k|\,ds\Big|,\qquad |\kappa_2| = \frac{1}{2\pi}\Big|\int_{\xi_2}^{\xi_3}|k|\,ds\Big|$$

and using the asymptotics for the Ω functions derived in the Appendix, the final result can be formulated as follows:

$$V_{\max} < E \quad (\kappa_2 = e^{-i\pi/2}|\kappa_2|),$$

$$\begin{aligned}\mathcal{L}(k) = &\frac{2\pi e^{-2\kappa_2}(\kappa_2)^{2\kappa_2}}{\Gamma(\frac{1}{4}+\kappa_2)\Gamma(\frac{3}{4}+\kappa_2)}\\ &+ e^{-4i\pi(\kappa_1+\kappa_2)} \simeq 1 \quad \text{for } |\kappa_2| \to \infty,\end{aligned} \tag{20a}$$

$$E_0 \le E \le V_{\max} \quad (\kappa_2 = e^{i\pi/2}|\kappa_2|;\ \kappa_1 = e^{i\pi}|\kappa_1|),$$

$$\begin{aligned}\mathcal{L}(k) = e^{2i\pi\kappa_2}\Big[&e^{2i\pi\kappa_1}\frac{2\pi e^{-2\kappa_2}(\kappa_2)^{2\kappa_2}}{\Gamma(\frac{1}{4}+\kappa_2)\Gamma(\frac{3}{4}+\kappa_2)}\\ &+ e^{-2i\pi\kappa_1}\Big] \simeq 2e^{-2i\pi\kappa_2}\cos(2\pi\kappa_1)\end{aligned}$$

$$\text{for } |\kappa_2| \to \infty \tag{20b}$$

$$V_{\min} \le E \le E_0 \quad (\kappa_2 = e^{i\pi/2}|\kappa_2|;\ \kappa_1 = e^{i\pi}|\kappa_1|),$$

$$\mathcal{L}(k) = e^{-2i\pi(\kappa_1+\kappa_2)} \frac{2\pi e^{-2\kappa_1}(\kappa_1)^{2\kappa_1}}{\Gamma(\frac{1}{4}+\kappa_1)\Gamma(\frac{3}{4}+\kappa_1)} \simeq 2e^{-2i\pi\kappa_2} \times \cos(2\pi\kappa_1) \quad \text{for } |\kappa_1| \to \infty \qquad (20c)$$

$E \le V_{\min} \quad (\kappa_1 = |\kappa_1|)$

$$\mathcal{L}(k) = \frac{2\pi e^{-2\kappa_1}(\kappa_1)^{2\kappa_1}}{\Gamma(\frac{1}{4}+\kappa_1)\Gamma(\frac{3}{4}+\kappa_1)} \simeq 1 \quad \text{for } |\kappa_1| \to \infty. \qquad (20d)$$

III. DISCUSSION AND CONCLUSION

The connection formula which can be obtained by Langer's method for three real, well-separated turning points is shown to be (see, e.g., Ref. 3)

$$\frac{1}{\sqrt{|k|}} \exp\Big(-\int_{\xi}^{\xi_1} |k|\,ds\Big)_{|\xi<\xi_1} \to \Big[2e^{-2i\pi\kappa_2} \times \cos(2\pi\kappa_1) - \tfrac{1}{2} i e^{2i\pi\kappa_2} \sin(2\pi\kappa_1)\Big] \times \frac{1}{\sqrt{|k|}} \exp\Big[-i\Big(\int_{\xi_3}^{\xi} |k|\,ds - \frac{\Pi}{4}\Big)\Big]_{|\xi>\xi_3} + \ldots .$$

This does agree with Eq. (20b) or Eq. (20c) (both valid for three real turning points) only by taking $|\kappa_2| \to \infty$, which means well-separated turning points. Similarly Eq. (20a) or Eq. (20d) (both valid for one real and two complex conjugate turning points) does agree with the corresponding formular for one real turning point, which can be obtained by Langer's method, only by taking $|\kappa_2| \to \infty$ or $|\kappa_1| \to \infty$, respectively. In both cases this means, that the distances of the complex turning points from the physical region of the position variable, as measured in units of a local wavelength, should be large. Inserting Eqs. (20a) into Eq. (15), it can be verified that the resulting connection formulas are continuations of one another for the different energy regions summed up in Eqs. (20a)-(20d).

Real potential functions with at most three classical points are occurring in both molecular and nuclear physics. In order to discuss physical effects which are related to the problem of three classical points such as "orbiting" collisions in molecular physics (see, e.g., Ref. 4), or "quasimolecular states" occurring in α-nucleus and heavy ion scattering (e.g., Ref. 5), the above mathematical formulas could therefore be readily used.

It would be desirable to extend the above considerations to become a general approximation method for the one-dimensional Schrödinger equation. For this purpose, a quantitative test for the approximations as a function of the turning points which are not taken into account would be necessary.

APPENDIX

From the asymptotic expansion for the Γ function,

$$\Gamma(b+\kappa) \simeq \sqrt{2\pi}\, e^{-\kappa}(\kappa)^{\kappa+b-1/2}, \quad |\kappa| \to \infty, \quad |\arg\kappa| < \pi$$

together with the relation

$$1/\Gamma(b-\kappa) = [2\sin\pi(b-\kappa)/2\pi]\Gamma(1-b+\kappa),$$

the following relations can be derived, which contain all relevant asymptotic formulas:

$$\left.\begin{aligned} &\Omega(\kappa) \simeq \sqrt{2/\pi}/\hat{\Omega}(\kappa) \\ &\Omega(\kappa e^{\pm i\pi}) \simeq \sqrt{2/\pi}\, e^{\pm i\pi\kappa}\hat{\Omega}(\kappa) \\ &\Omega(\kappa e^{\pm 2i\pi}) \simeq 0 \end{aligned}\right\}, \quad |\kappa| \to \infty, \quad |\arg\kappa| < \pi,$$

$$\Omega(\kappa e^{\pm 4i\pi}) = -\Omega(\kappa).$$

[1]W. Hecht, J. Math. Phys. **13**, 1291 (1972).
[2]Buchholz, *The Confluent Hypergeometric Function, Springer Tracts in Natural Philosophy* (Springer-Verlag, Berlin, 1969), Vol. 15.
[3]Berry, Proc. Phys. Soc. Lond. **88**, 285 (1966).
[4]R. Herm, J. Chem. Phys. **47**, 4290 (1967).
[5]K. W. McVoy, Phys. Rev. C **3**, 1104 (1971).

Radial Jost functions in scattering theory*

R. K. Nesbet

IBM Research Laboratory, San Jose, California 95193
(Received 2 April 1973)

Several methods have recently been proposed for representing oscillatory wavefunctions by relatively slowly varying modulations of known oscillatory functions. Direct computation of the modulating function leads to efficient numerical or variational procedures and to accurate interpolation over energy or other parameters of a scattering problem. A new method, based on the phase integral (WKB) formalism, is proposed here in a form applicable to multichannel scattering. The method generalizes the phase integral method to make use of arbitrary oscillatory comparison functions, rather than just plane waves as in the usual formalism. It makes use of the modulating factor of the radial Jost function. Several examples of the proposed method are given, including an application to a two-channel model problem.

I. INTRODUCTION

In many applications of quantum mechanics to collision theory, it is necessary to integrate systems of differential or integrodifferential equations in the range of positive energies. This produces oscillatory wavefunctions. Recently, several computational methods have been introduced with the common feature of representing the computed wavefunctions by relatively slowly varying modulation of known oscillatory functions. The modulation is computed directly, and the accuracy of numerical integration is governed by the smoothness of the *potential* function, rather than of the oscillatory wavefunction.

The method of Gordon[1] makes use of a piecewise approximation to the potential function, matching exact solutions for constant or linear potential segments at the segment boundaries. The method of Light[2] uses an exponential matrix formalism whose accuracy is governed by the variation of the potential function.

The present paper is especially concerned with several related methods, to be referred to here as *variable phase* methods,[3,4,5] whose common basis was discussed in a recent note.[6] That discussion is extended here to point out the close relationship between these methods and the well-known *phase integral* method, which is the basis of the WKB approximation in a single-channel problem.[7] The necessary definitions and derivations relevant to the present discussion are given in Sec. II.

Section III proceeds to a synthesis and generalization of these ideas, leading to a formal and computational procedure that represents an extension of the phase integral method to multichannel problems. A modulating function is defined that is closely related to the radial function considered originally by Jost,[8] and which gives the Jost function,[8] known to have very regular analytic properties, as its end value when computed over the full range of the radial variable.

The present work differs in approach from recent proposals by Thorson and collaborators[9] for computational methods based on the phase-integral formalism. In particular, the present method applies directly to multichannel scattering.

Some simple examples of the proposed formalism are given in Sec. IV, including a two-channel model problem to illustrate use of the multichannel equations proposed here. Section V concludes with a discussion of the possible advantages of this formalism, in comparison with variable phase methods.

II. VARIABLE PHASE AND PHASE INTEGRAL METHODS

Consider the radial Schrödinger equation, in Hartree atomic units,

$$\left(\frac{d^2}{dr^2} + k^2 - 2V_0(r)\right) u(r) = 2V_1(r)u(r), \tag{1}$$

where k^2 is nonnegative. For multichannel scattering, V_0, V_1, and u become matrices, and k^2 is a diagonal matrix, with positive values for open scattering channels. In the methods to be considered here, it is assumed that solutions are known of a comparison equation

$$\left(\frac{d^2}{dr^2} + k^2 - 2V_0(r)\right) w(r) = 0, \tag{2}$$

which can also be considered as a matrix of equations for multichannel scattering. For simplicity, $V_0(r)$ will be assumed here to be a diagonal matrix.

For a single open channel, the regular solution of Eq. (1) can be expressed in the form

$$u_0(r) = w_0(r)c(r) + w_1(r)s(r), \tag{3}$$

where w_0 and w_1 are real-valued solutions of Eq. (2) that are, respectively, regular and irregular at the coordinate origin, $r = 0$. These functions are assumed to be normalized so that their Wronskian is

$$w_0' w_1 - w_0 w_1' = 1, \tag{4}$$

and to have equal amplitude as $r \to \infty$, but to differ in phase by $\pi/2$. The asymptotic forms are

$$w_0 \sim k^{-1/2} \sin\theta(r), \quad w_1 \sim k^{-1/2} \cos\theta(r), \tag{5}$$

where $\theta(r)$ depends on the choice of comparison potential V_0. If this is the centrifugal potential $l(l+1)/2r^2$, then

$$\theta(r) = kr - \tfrac{1}{2}l\pi, \tag{6}$$

and w_0, w_1 are spherical Bessel functions multiplied by r. For a Coulomb or dipole potential, $\theta(r)$ must be suitably modified, but the formulas derived here retain their validity. For multichannel scattering, the functions c and s of Eq. (3) become matrices, and $u_0(r)$ becomes a matrix.

The auxiliary functions $c(r)$ and $s(r)$ are not uniquely defined by Eq. (3). In the *variable phase* method[3] (and the related methods of Johnson and Secrest[4] and of

Copyright © 1973 by the American Institute of Physics

Sams and Kouri[5]), these functions are constrained by the auxiliary condition

$$w_0 c' + w_1 s' \equiv 0, \tag{7}$$

so that

$$u_0'(r) = w_0'(r)c(r) + w_1'(r)s(r). \tag{8}$$

The effect of this constraint condition is to give coupled first-order equations for the auxiliary functions when Eq. (3) is substituted into Eq. (1):

$$\begin{aligned} c' &= 2w_1 V_1(w_0 c + w_1 s), \\ s' &= -2w_0 V_1(w_0 c + w_1 s). \end{aligned} \tag{9}$$

The regular solution u_0 of Eq. (1) is obtained from these equations if, at $r = 0$,

$$c(0) = 1, \quad s(0) = 0. \tag{10}$$

With this boundary condition, Eqs. (9) are equivalent to coupled integral equations of Volterra form,

$$\begin{aligned} c(r) &= 1 + 2\int_0^r w_1(r')V_1(r')u_0(r')dr', \\ s(r) &= -2\int_0^r w_0(r')V_1(r')u_0(r')dr'. \end{aligned} \tag{11}$$

These equations as given are in proper matrix form for multichannel scattering.

Scattering information is obtained from the function or matrix

$$t(r) = s(r)c^{-1}(r), \tag{12}$$

whose asymptotic value as $r \to \infty$ is the reactance matrix (K matrix) if V_0 is the centrifugal or Coulomb potential.

In the *phase integral* method,[7] the auxiliary functions $c(r)$ and $s(r)$ are constrained not by Eq. (7) but by the condition that the function

$$u_1(r) = w_1(r)c(r) - w_0(r)s(r) \tag{13}$$

should be a solution of Eq. (1), independent of the regular solution u_0. The formalism applicable to a single-channel problem cannot be applied directly to multichannel scattering. For a single scattering channel, the specific condition imposed on $c(r)$ and $s(r)$ is that the Wronskian

$$u_0'u_1 - u_0u_1' = 1 \tag{14}$$

should be constant. Equivalently,

$$(c^2 + s^2) + (w_0^2 + w_1^2)(cs' - sc') = 1. \tag{15}$$

This condition can be expressed most conveniently in terms of two functions

$$\begin{aligned} q_w &= (w_0^2 + w_1^2)^{-1}, \\ q &= (u_0^2 + u_1^2)^{-1} = [(c^2 + s^2)(w_0^2 + w_1^2)]^{-1}, \end{aligned} \tag{16}$$

which are real and positive throughout the interval $0 < r \le \infty$. These functions are finite in this interval if the potentials V_0 and V_1 are nonsingular except at $r = 0$. Equation (15) implies

$$\frac{d}{dr}\tan^{-1}(sc^{-1}) = \frac{cs' - sc'}{c^2 + s^2} = q - q_w. \tag{17}$$

The function $q(r)$ satisfies a nonlinear differential equation[7]

$$q^{1/2}\frac{d^2}{dr^2}q^{-1/2} = q^2 - Q^2, \tag{18}$$

where

$$Q^2 = k^2 - 2V_0 - 2V_1. \tag{19}$$

The function q_w satisfies an analogous equation. It is convenient to define q by the boundary condition

$$q - q_w \to 0 \tag{20}$$

as $r \to \infty$. Then the scattering phase shift relative to the comparison wave function is given by Eq. (17) as the phase integral,

$$\delta = \int_0^\infty [q(r') - q_w(r')]\,dr'. \tag{21}$$

It should be noted that Eq. (20) requires $c^2 + s^2 = 1$ as $r \to \infty$. This is not in general compatible with the normalization of $u_0(r)$ implied by Eq. (10). The value of $c(0)$ must be determined by inward integration of $q(r)$.

The WKB method[7] makes use of the approximation

$$q \cong Q = (k^2 - 2V)^{1/2}, \tag{22}$$

with $q_w = k$ for comparison potential $V_0 = 0$. The lower limit of the phase integral is taken to be the outermost classical turning point r_0, where $k^2 - 2V$ vanishes.

III. THE RADIAL JOST FUNCTION

The regular and irregular real wave functions defined above can be expressed in terms of complex functions[7]

$$w = w_1 + iw_0 = q_w^{-1/2}\exp\left(i\int_0^r q_w(r')dr'\right), \tag{23}$$

$$u = u_1 + iu_0 = q^{-1/2}\exp\left(i\int_0^r q(r')dr'\right). \tag{24}$$

It is convenient to consider the wavefunction

$$w\gamma = ue^{-i\delta} \tag{25}$$

such that, using Eq. (21) for the phase integral,

$$\gamma(r) = (q_w/q)^{1/2}\exp\left(-i\int_r^\infty (q - q_w)dr'\right). \tag{26}$$

From Eqs. (3) and (13),

$$\gamma = (c + is)e^{-i\delta}, \tag{27}$$

where $s(0) = 0$ as in Eq. (10), but $c(0)$ cannot in general be set equal to unity. From Eq. (20), as $r \to \infty$,

$$\gamma(\infty) = 1, \quad c(\infty) = \cos\delta, \quad s(\infty) = \sin\delta. \tag{28}$$

As $r \to 0$,

$$\gamma(0) = (q_w(0)/q(0))^{1/2}e^{-i\delta}, \tag{29}$$

so Eq. (27) requires that

$$c(0) = (q_w(0)/q(0))^{1/2}, \quad s(0) = 0. \tag{30}$$

If w is $\exp(ikr)$, then $w\gamma$ as defined here is the radial wavefunction $f(-k, r)$ considered by Jost.[8]

The function $\gamma(r)$ is expected to have very regular analytic properties, because of the simple analytic behavior of q_w and q. To take practical advantage of this, the functional form $w\gamma$ can be substituted into the radial Schrödinger equation (1), and then Eq. (2) can be used to obtain the differential equation for $\gamma(r)$,

$$\left(\frac{d^2}{dr^2} + 2\frac{w'}{w}\frac{d}{dr} - 2V_1(r)\right)\gamma(r) = 0. \tag{31}$$

In order to avoid the asymptotic solution w^*/w, this equation must be integrated inwards, starting from the boundary conditions

$$\gamma(\infty) \sim 1, \quad \gamma'(\infty) \sim 0. \tag{32}$$

From Eq. (23), the logarithmic derivative of the comparison function w is

$$\frac{w'}{w} = -\frac{1}{2}\frac{q_w'}{q_w} + iq_w. \tag{33}$$

Hence Eq. (31) involves only the slowly varying function q_w and the perturbing potential $V_1(r)$. When V_1 vanishes, the solution $\gamma(r)$ is a constant. These properties should make Eq. (31) especially useful for numerical computations.

The multichannel generalization of Eq. (31) is

$$\left(\frac{d^2}{dr^2} + 2w_p^{-1}w_p\frac{d}{dr}\right)\gamma_{pn}(r) = 2\sum_q w_p^{-1}V_{pq}w_q\gamma_{qn}(r), \tag{34}$$

where the indices range over the number of open channels coupled by these equations. For simplicity, closed channels are not considered here, the comparison potential is assumed to be diagonal, and the potentials are considered to be local operators. A matrix solution is to be obtained subject to the boundary conditions

$$\gamma_{pn}(\infty) \sim \delta_{pn}, \quad \gamma'_{pn}(\infty) \sim 0. \tag{35}$$

Then the reactance matrix is given by

$$K_{pq} = \sum_n \mathrm{Im}[\gamma^{-1}(0)]_{pn}\{\mathrm{Re}[\gamma^{-1}(0)]\}^{-1}_{nq}. \tag{36}$$

This formula is obtained by considering the matrix wavefunction $w\gamma\gamma^{-1}(0)$, which is purely real at $r = 0$, so that its imaginary part is the regular solution matrix u_0. Then by comparison with Eq. (12),

$$t(r) = \mathrm{Im}[\gamma(r)\gamma^{-1}(0)]\{\mathrm{Re}[\gamma(r)\gamma^{-1}(0)]\}^{-1}. \tag{37}$$

Since $\gamma(\infty) = 1$, the asymptotic value of this expression gives the K matrix as indicated in Eq. (36).

The matrix function $\gamma(r)\gamma^{-1}(0)$ can be shown to approach $J^*(k)$ as $r \to \infty$, where $J(k)$ is the Jost function or matrix.[8] Since $\gamma(\infty) = 1$,

$$J^*(k) = \gamma^{-1}(0). \tag{38}$$

IV. THREE SIMPLE EXAMPLES

A. Spherical Bessel functions

The regular and irregular spherical Bessel functions (multiplied by r) are solutions of Eq. (1) if

$$V_0 = 0, \quad V_1 = l(l+1)/2r^2. \tag{39}$$

Since the differential equation is homogeneous in r, it is convenient to use the dimensionless variable

$$z = kr,$$

and to replace k by unity in Eqs. (1) and (2). Then the comparison function is

$$w(z) = e^{iz} = w_0 + iw_1. \tag{40}$$

Equation (31), for $\gamma_l(z)$, is

$$\left(\frac{d^2}{dz^2} + 2i\frac{d}{dz} - \frac{l(l+1)}{z^2}\right)\gamma(z) = 0, \tag{41}$$

to be integrated inwards from boundary values

$$\gamma(\infty) = 1, \quad \gamma'(\infty) = 0. \tag{42}$$

Equation (41) has a finite series solution[10]

$$\gamma_l(z) = z^{-l}\sum_{a=0}^{l} g_a z^a, \tag{43}$$

where the coefficients are integers multiplied by powers of i,

$$g_a = \frac{(2l-a)!\, i^{l-a}}{(2l-2a)!!\, a!} = \binom{2l-a}{a}(2l-2a-1)!!\, i^{l-a}, \tag{44}$$

with the convention that $(-1)!! = 1$. Because the comparison potential is singular in this example, $\gamma_l(z)$ is singular at the origin. To insure that the function u_0 is the regular solution, γ_l must be multiplied by a phase factor to make the leading term purely real at $z = 0$. From Eq. (44), this factor is i^{-l}. The resulting function can be expressed in terms of real polynomials A_l and B_l,

$$i^{-l}\gamma_l(z) = z^{-l}[(A_l(z) - iB_l(z)]. \tag{45}$$

The function q defined by Eq. (16) is

$$q_l(z) = z^{2l}/(A_l^2 + B_l^2). \tag{46}$$

For example, for $l = 3$

$$A_3(z) = 15 - 6z^2, \quad B_3(z) = 15z - z^3,$$
$$q_3(z) = z^6(225 + 45z^2 + 6z^4 + z^6)^{-1}. \tag{47}$$

Because $q_l(z)$ is finite and positive for positive z, $A_l^2 + B_l^2$ is an even polynomial with positive coefficients. From Eq. (44), these coefficients are integers.

Equation (24) gives compact formulas for the regular and irregular spherical Bessel functions

$$j_l(z) = z^{-1}u_{l0}(z) = z^{-l-1}(A_l^2 + B_l^2)^{1/2} \times \sin\int_0^z \zeta^{2l}(A_l^2 + B_l^2)^{-1}d\zeta$$
$$n_l(z) = -z^{-1}u_{l1}(z) = -z^{-l-1}(A_l^2 + B_l^2)^{1/2} \times \cos\int_0^z \zeta^{2l}(A_l^2 + B_l^2)^{-1}d\zeta. \tag{48}$$

Equation (17) in the present case becomes

$$\frac{d}{dz}\tan^{-1}\left(\frac{-B_l}{A_l}\right) = q_l(z) - 1, \tag{49}$$

so that

$$\int_0^z \frac{\zeta^{2l}d\zeta}{A_l^2 + B_l^2} = \int_0^z q_l(\zeta)d\zeta = z - \tan^{-1}\left(\frac{B_l(z)}{A_l(z)}\right). \tag{50}$$

This agrees with the assumed functional forms

$$j_l(z) = z^{-1}\,\mathrm{Im}[w(z)i^{-l}\gamma_l(z)]$$
$$= z^{-l-1}(A_l^2 + B_l^2)^{1/2}\sin[z - \tan^{-1}(B_l/A_l)],$$

$$n_l(z) = -z^{-1}\,\mathrm{Re}[w(z)i^{-l}\gamma_l(z)]$$
$$= -z^{-l-1}(A_l^2 + B_l^2)^{1/2}\cos[z - \tan^{-1}(B_l/A_l)]. \quad (51)$$

These formulas are equivalent to the well-known finite expansion of the spherical Bessel functions,[10] but they appear here in an especially simple form for accurate numerical computation except in the limit $z \to 0$. For small values of z, the first of Eqs. (51) must lead to exact cancellation of the first $2l$ powers of z, since $j_l(z)$ varies as z^l. In fact, it can easily be verified that B_l/A_l is the lth convergent of the continued fraction

$$\tan z = \frac{z}{1-}\,\frac{z^2}{3-}\,\frac{z^2}{5-}\cdots. \quad (52)$$

This can be used to obtain simple formulas for computing A_l and B_l by recurrence on the index l.[10] From Eqs. (42) and (45), the asymptotic value of $\tan^{-1}(-B_l/A_l)$ is the phase of i^{-l}, or $-l\pi/2$, in agreement with Eq. (6).

In scattering theory, calculations with the centrifugal potential as comparison potential give the usual definition of non-Coulombic partial wave phase shifts. The

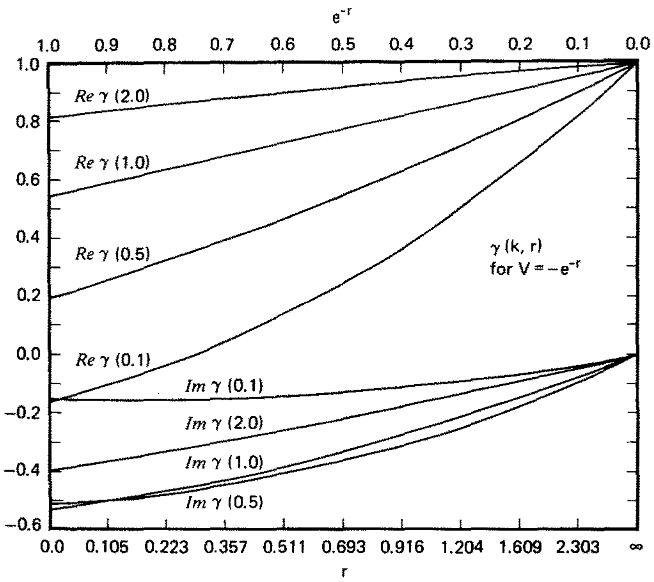

FIG. 1. $\gamma(k;r)$ for $V = -e^{-r}$.

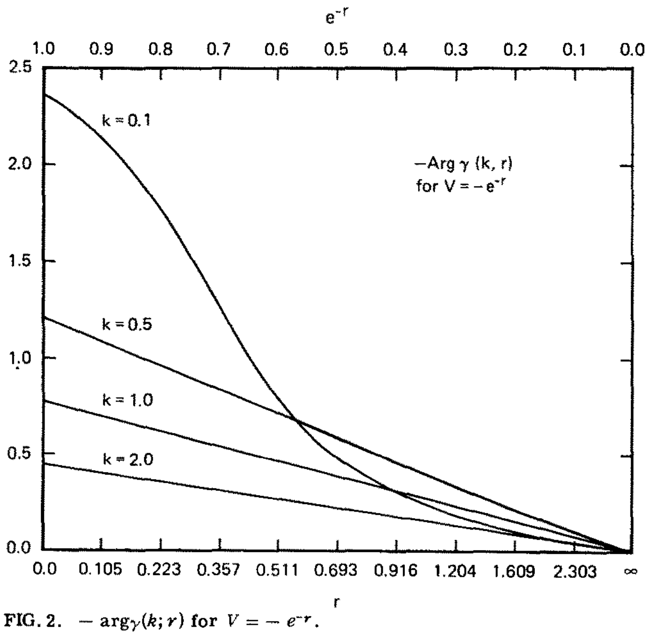

FIG. 2. $-\arg\gamma(k;r)$ for $V = -e^{-r}$.

comparison wavefunctions required are

$$w_l(z) = z^{-l}(A_l - iB_l)e^{iz}. \quad (53)$$

The logarithmic derivatives required in Eqs. (34) can be expressed entirely in terms of the polynomials A and B, using Eq. (33) and the recurrence formulas mentioned above. The result is

$$\frac{w_l'}{w_l} = \frac{-l}{z} + z\,\frac{A_lA_{l-1} + B_lB_{l-1}}{A_l^2 + B_l^2} + i\,\frac{z^{2l}}{A_l^2 + B_l^2}. \quad (54)$$

B. Exponential potential well

Exact phase shifts are known for s-wave scattering by an attractive exponential potential,[11]

$$V_0 = 0, \quad V_1 = -e^{-r}. \quad (55)$$

Equation (31) for this problem is

$$\left(\frac{d^2}{dr^2} + 2ik\,\frac{d}{dr} + 2e^{-r}\right)\gamma(r) = 0. \quad (56)$$

If $t = e^{-r}$ is used as the independent variable, this becomes

$$\frac{d^2\gamma}{dt^2} = -t^{-1}\left((1 - 2ik)\,\frac{d\gamma}{dt} + 2\gamma\right), \quad (57)$$

with boundary conditions at $t = 0$, for $\gamma(t)$,

$$\gamma(0) = 1, \quad \frac{d\gamma}{dt}(0) = \frac{-2}{1 - 2ik}. \quad (58)$$

This equation was integrated from $t = 0$ to $t = 1$, using the Runge–Kutta–Gill method,[12] for values of k ranging from 0.1 to 4.0. Real and imaginary parts of $\gamma(k;r)$ are shown in Fig. 1 and $\arg\gamma$ is shown in Fig. 2. With less than thirty integration points for any k value, the computed phase shifts agree to four significant decimals with the exact results.[11] The smooth nature of $\gamma(r)$ is evident from the figures. This property of γ makes it especially suitable for efficient numerical integration.

C. Two-channel Huck model

A simple two-channel model problem, with exact solutions computable in terms of elementary functions, has been used by Huck and others to test multichannel variational methods.[13] The comparison potential is zero, and the perturbing potential matrix is purely nondiagonal, with

$$V_{12} = V_{21} = \tfrac{1}{2}C, \quad r \le a,$$
$$= 0, \quad r > a. \quad (59)$$

The two k values are related by

$$k_1^2 - k_2^2 = 2\Delta E = 0.75. \quad (60)$$

Exact results for the elastic and inelastic cross sections Q_{pq}, computed for $k_1 = 1.0$, $k_2 = 0.5$, $a = 1.0$, and $C^2 = 2.0(2.0)12.0$ have been given elsewhere.[13]

As a test of the present method, these results have been duplicated by numerical integration of Eqs. (34) for the multichannel matrix function $\gamma_{pn}(r)$. Since the potential function vanishes for $r > 1.0$, the integration is carried inwards from 1.0 to 0.0. Then the K matrix is computed from Eq. (36), and cross sections are obtained by the usual formulas.

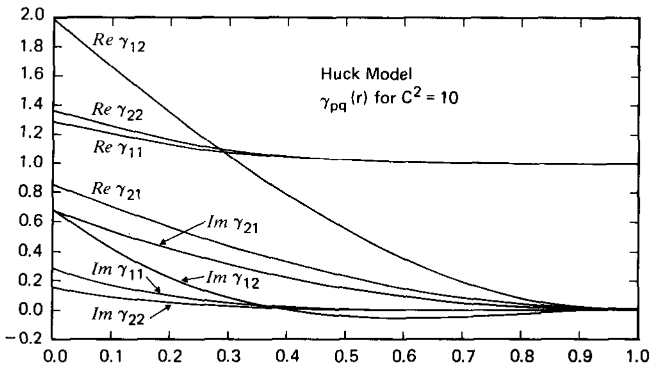

FIG. 3. $\gamma_{pq}(r)$ for Huck model, $C^2 = 10.0$.

Results for the real and imaginary parts of the functions $\gamma_{pn}(r)$, computed for $c^2 = 10.0$, are shown in Fig. 3. The smooth nature of these functions, and hence their suitability for numerical integration, is evident from the figure. Results obtained with 21 integration points in the interval $0 \leq r \leq 1$ give cross sections that agree with exact results to within two units in the sixth significant decimal place.

V. DISCUSSION

The examples given here illustrate the expected smooth analytic behavior of the function or matrix $\gamma(r)$. This formalism should lend itself to efficient numerical integration or to rapid convergence of variational expansions. Such variational expansions will be considered in a separate paper.[14]

In comparison with the method of Sams and Kouri,[5] which uses Eqs. (11) directly, and to the closely related method of Johnson and Secrest,[4] the present method loses the advantage of working with first-order differential equations. The *number* of equations for a given differential system is the same, since solving coupled equations for $c(r)$ and $s(r)$ is equivalent to solving a single equation for the complex function $\gamma(r)$. However, a second-order equation with first derivative terms, as in Eq. (31), is equivalent to two first-order equations if a method such as that of Runge and Kutta is used.

The comparative advantage of the method proposed here is expected to lie in its complete elimination of oscillatory terms in the diagonal channel part of Eqs. (34). In contrast, oscillatory functions occur explicitly in the diagonal part of the integrands of variable phase methods, as in Eqs. (11). Integration over oscillatory functions by standard methods requires closely spaced integration points. The present method may eliminate this requirement expect for the unavoidable nondiagonal terms in Eqs. (34).

The present method obtains both regular and irregular solutions of a scattering problem. While in general this may be no advantage, it opens up the possibility of defining much more general classes of comparison functions than the usual spherical Bessel or Coulomb functions. Comparison functions obtained by numerical integration could be used in the equations of this method. This would allow simple computation of the scattering effects of variations or perturbations of the comparison potential function.

Since the Jost function is obtained directly in the present method, it could be used to interpolate or extrapolate scattering data as a function of k or the energy E. It can be seen from Fig. 1 that the functions $\gamma(k;r)$ vary quite smoothly with k for all values of r. The limit as $k \to \infty$ is especially simple, since $\gamma(\infty;r) \to 1$. The data in this figure suffice to determine wavefunction and phase shift to quite high accuracy over the range $0.1 \leq k \leq \infty$.

*Supported in part by the Office of Naval Research under Contract N00014-72-C-0051.

[1]R. G. Gordon, J. Chem. Phys. **51**, 14 (1969).
[2]J. C. Light, Methods Comput. Phys. **10**, 111 (1971).
[3]B. R. Levy and J. B. Keller, J. Math. Phys. **4**, 54 (1963); C. Zemach, Nuovo Cimento **33**, 939 (1964); F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967).
[4]B. R. Johnson and D. Secrest, J. Math. Phys. **7**, 2187 (1966); B. R. Johnson and D. Secrest, J. Chem. Phys. **48**, 4682 (1968); D. Secrest, Methods Comput. Phys. **10**, 243 (1971); W. Eastes and D. Secrest, J. Chem. Phys. **56**, 640 (1972).
[5]W. N. Sams and D. J. Kouri, J. Chem. Phys. **51**, 4809 (1969); J. Chem. Phys. **51**, 4815 (1969).
[6]R. K. Nesbet, Comments At. Mol. Phys. **3**, 143 (1972).
[7]J. Heading, *An Introduction to Phase-Integral Methods* (Wiley, New York, 1962); N. Fröman and P. O. Fröman, *JWKB Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965); R. K. Nesbet, J. Math. Phys. **9**, 205 (1968).
[8]R. Jost, Physica (Utr.) **12**, 509 (1946); Helv. Phys. Acta **20**, 256 (1947); R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), Chap. 12.
[9]W. I. Newman and W. R. Thorson, Phys. Rev. Lett. **29**, 1350 (1972); Can. J. Phys. **50**, 2997 (1972); W. R. Thorson and J. H. Choi, Bull. Am. Phys. Soc. **17**, 1126 (1972).
[10]M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series, No. 55 (U. S. Government Printing Office, Washington, D. C., 1964), pp. 437–56, also p. 19.
[11]H. A. Bethe and R. Bacher, Rev. Mod. Phys. **8**, 111 (1936); H. Morawitz, J. Math. Phys. **11**, 649 (1970).
[12]Subroutine DRKGS, IBM Scientific Subroutine Package, IBM Application Program Manual H20-0205 (IBM, White Plains, New York, 1968).
[13]R. J. Huck, Proc. Phys. Soc. Lond. A **70**, 369 (1957). R. K. Nesbet, Phys. Rev. **179**, 60 (1969).
[14]R. S. Oberoi and R. K. Nesbet (to be published).

Invariants of the equations of wave mechanics*: Rigid rotator and symmetric top

Robert L. Anderson, Sukeyuki Kumei, and Carl E. Wulfman

Department of Physics, University of the Pacific, Stockton, California 95204

(Received 18 April 1972; revised manuscript received 7 December 1972)

Applying the systematic method discussed in previous papers, we derive the invariants and the groups of the time-dependent Schrödinger equations for the rigid rotator and the symmetric top. The groups for these systems are found to be $SO(3,2)$ (rigid rotator) and $SU(2,2)$ (symmetric top). For the case of the symmetric top, it is found that under the symmetry breaking $I_1 = I_2 = I_3 \to I_1 = I_2 \neq I_3$, where I_1, I_2, and I_3 are the moments of inertia of the top, two of the time-independent constants of the motion become time-dependent constants of the motion.

1. INTRODUCTION

The symmetry properties of differential equations play an important role in understanding the fundamental properties of physics. Systematic methods for finding the symmetry of differential equations were introduced by Sophus Lie.[1] However, Lie's method is not general enough to obtain many physically important groups. For example, the $O(4)$ degeneracy group of a hydrogen atom cannot be derived by his method from Schrödinger's equation.

In a series of papers (I—III),[2,3] we have demonstrated the validity and necessity of generalizing Lie's original point transformations when one is considering the symmetry properties of partial differential equations. The essential point in the generalization is that the most general transformations which leave a system of partial differential equations invariant in form[4] must allow not only for the algebraic independence of the usual independent variables, and the unknown functions, but also, in general, for the algebraic independence of a subset of derivatives of the unknown functions.

This leads to a new program for the discovery and analysis of the invariants of partial differential equations. As applied here, the program systematically determines a set of invariants $\{Q\}$ such that for a given differential operator K, if $K\psi = 0$, then $KQ\psi = 0$. The set $\{Q\}$ maps the solution space $\{\psi\}$ of K into itself. Its elements form a Lie algebra and act irreducibly on the solution space.

In order to realize this program, we have in this paper utilized a space-time dilation operator D and determined the Lie algebra corresponding to the transformed time-dependent Schrödinger operator $D(H - i\partial_t)D^{-1}$ with solution space $\{D\psi\}$. The space-time dilation operator linearizes the spectrum of $D(i\partial_t)D^{-1}$ and the dynamical algebra of $D(H - i\partial_t)D^{-1}$ contains finite order derivatives only. *The invariants Q of the original Schrödinger equation, obtained by inverse dilation, comprise an isomorphic algebra even though some of them are functions of derivatives of arbitrarily high order.*

Here we use the approach just outlined to determine dynamical algebras and dynamical groups for the rigid rotator and symmetric top. The groups will be shown to be $SO(3,2)$ (rigid rotator) and $SU(2,2)$ (symmetric top).

2. THE RIGID ROTATOR

The Schrödinger equation for the rigid rotator is

$$(L^2 - i\partial_t)\sum_{lm} e^{-il(l+1)t}Y_{lm}(\theta,\phi)C_{lm} = 0, \tag{2.1}$$

where C_{lm} are arbitrary constants and

$$L^2 = -y^2\partial_x\partial_x + 2x\partial_x - y^{-2}\partial_\phi\partial_\phi$$

with

$$x = \cos\theta, \quad y = \sin\theta.$$

The transformation operator leading to a linear spectrum[5] is seen to be

$$D = \exp\{t\partial_t \log(\bar{l} + 1)^{-1}\} \tag{2.2}$$

with

$$\bar{l} = \tfrac{1}{2}[-1 + (1 + 4L^2)^{1/2}].$$

Applying it to (2.1) yields the transformed equation

$$[L^2 - i(i\partial_t + 1)\partial_t]f(\theta,\phi,t) = 0, \tag{2.3}$$

where

$$f(\theta,\phi,t) = \sum_{lm} C_{lm}e^{-ilt}Y_{lm}(\theta,\phi).$$

To find the spectrum generating algebra of Eq. (2.3), and hence of Eq. (2.1), we choose as independent functions

$$f, f_t, f_\phi, f_x, f_{\phi\phi}, f_{x\phi}, f_{\phi t}, f_{tt}$$

$$f_{xt}, f_{\phi\phi\phi}, f_{x\phi\phi}, f_{\phi tt}, f_{x\phi t}, f_{\phi\phi t}$$

and let Q operator be of the form

$$Q = Q^{\phi\phi}\partial_\phi\partial_\phi + Q^{x\phi}\partial_x\partial_\phi + Q^x\partial_x + Q^\phi\partial_\phi + Q^t\partial_t + Q^0. \tag{2.4}$$

Then the determining equations derived by expanding the equation

$$[L^2 - i(i\partial_t + 1)\partial_t]Qf = 0,$$

and by making use of the linear independence of the above functions, are

$$Q_t^{\phi\phi} = 0, \quad Q_t^{x\phi} = 0, \quad Q_x^{x\phi} + xy^{-2}Q^{x\phi} = 0,$$

$$Q_x^{\phi\phi} - y^{-4}Q_\phi^{x\phi} = 0, \quad Q_\phi^{\phi\phi} - xy^{-2}Q^{x\phi} = 0,$$

$$Q_x^t - y^{-2}Q_t^x = 0, \quad Q_t^t - Q_x^x - xy^{-2}Q^x = 0,$$

$$Q_t^\phi - y^{-2}Q_\phi^t = 0, \quad AQ^{x\phi} - 2Q^{x\phi} - 2y^{-2}Q_\phi^x - 2y^2Q_x^\phi = 0,$$

$$AQ^{\phi\phi} + 2y^{-2}Q_x^x - 4xy^{-4}Q^x - 2y^{-2}Q_\phi^\phi = 0,$$

$$AQ^x - 2Q^x - 4xQ_x^x - 4x^2y^{-2}Q^x - 2y^2Q_x^0 = 0,$$

$$AQ^\phi - 2y^{-2}Q_\phi^0 = 0,$$

$$AQ^t + 2Q_t^0 + 2iQ_x^x + 2ixy^{-2}Q^x = 0, \quad AQ^0 = 0,$$

Copyright © 1973 by the American Institute of Physics

with

$$A = -y^2\partial_x\partial_x + 2x\partial_x - y^{-2}\partial_\phi\partial_\phi + \partial_t\partial_t - i\partial_t.$$

The operator obtained by solving these equations contains 14 parameters;

$$Q = (2.4) = \sum_{i=1}^{14} a^i Q_i,$$

where the a^i are integration constants and the Q_i are given by

$$Q_{\frac{1}{2}} = e^{\pm i\phi}(y\partial_x \mp ixy^{-1}\partial_\phi), \quad Q_3 = \partial_\phi,$$

$$Q_4 = e^{-it}(iy^2\partial_x + x\partial_t - ix), \quad Q_5 = e^{it}(-iy^2\partial_x + x\partial_t),$$

$$Q_6 = \partial_t, \quad Q_{\frac{7}{8}} = e^{it}e^{\pm i\phi}(ixy\partial_x \pm y^{-1}\partial_\phi + y\partial_t),$$

$$Q_{\frac{9}{10}} = e^{-it}e^{\pm i\phi}(-ixy\partial_x \mp y^{-1}\partial_\phi + y\partial_t - iy),$$

$$Q_{\frac{11}{12}} = e^{\pm i\phi}(\mp ixy^{-1}\partial_\phi\partial_\phi + y\partial_x\partial_\phi),$$

$$Q_{13} = \partial_\phi\partial_\phi, \quad Q_{14} = 1.$$

As these operators are derived by using the transformed equation (2.3), the corresponding operators $\tilde{Q}_i$ for the original equation (2.1) have the form

$$\tilde{Q}_i = D^{-1}Q_iD.$$

It is clear that the set $\{\tilde{Q}_i\}$ still satisfy the same commutation relations as the set $\{Q_i\}$.[6]

We investigate the properties of these operators: First we note that Q_{11}, Q_{12}, Q_{13} are expressed in terms of the angular momentum operators Q_1, Q_2, Q_3 as

$$Q_{11} = Q_1Q_3, \quad Q_{12} = Q_2Q_3, \quad Q_{13} = Q_3Q_3.$$

The remaining operators satisfy the following commutation relations:

$$[\tilde{Q}_1, \tilde{Q}_2] = 2i\tilde{Q}_3, \quad [\tilde{Q}_3, \tilde{Q}_1] = i\tilde{Q}_1, \quad [\tilde{Q}_3, \tilde{Q}_2] = -i\tilde{Q}_2,$$

$$[\tilde{Q}_4, \tilde{Q}_5] = -2i\tilde{Q}_0, \quad [\tilde{Q}_0, \tilde{Q}_4] = i\tilde{Q}_4, \quad [\tilde{Q}_0, \tilde{Q}_5] = -i\tilde{Q}_5$$

$$[\tilde{Q}_1, \tilde{Q}_5] = \tilde{Q}_7, \quad [\tilde{Q}_2, \tilde{Q}_5] = \tilde{Q}_8, \quad [\tilde{Q}_1, \tilde{Q}_4] = \tilde{Q}_9,$$

$$[\tilde{Q}_2, \tilde{Q}_4] = \tilde{Q}_{10}$$

$$[\tilde{Q}_3, \tilde{Q}_i] = 0 \quad (i = 0, 4, 5), \quad [\tilde{Q}_0, \tilde{Q}_i] = 0 \quad (i = 1, 2, 3),$$

where $\tilde{Q}_0 = -\tilde{Q}_6 + \frac{1}{2}i\tilde{Q}_{14}$.

From these it is clear that $\tilde{Q}_4$ and $\tilde{Q}_5$ shift the eigenvalue $i(l+\frac{1}{2})$ of $\tilde{Q}_0$ by unit amount. They are found to satisfy the relations

$$\tilde{Q}_4 e^{-il(l+1)t}Y_{lm} = -i[(2l+1)(2l+3)^{-1}(l+1-m)(l+1+m)]^{1/2} e^{-i(l+1)(l+2)t}Y_{l+1\,m},$$

$$\tilde{Q}_5 e^{-il(l+1)t}Y_{lm} = -i[(2l+1)(2l-1)^{-1}(l+m)(l-m)]^{1/2} e^{-i(l-1)lt}Y_{l-1\,m}.$$

Because of the presence of the factors $(2l+1)(2l+3)^{-1}$ and $(2l+1)(2l-1)^{-1}$ in these coefficients, no linear combinations of the operators $\tilde{Q}_4$ and $\tilde{Q}_5$ are skew-adjoint under the ordinary scalar product

$$(f,g) = \int_0^{2\pi}\int_0^{\pi} f^*g \sin\theta \, d\theta d\phi.$$

To construct operators with the proper adjointness, we define the new operators

$$\overline{Q}_i = (\tilde{Q}_0)^{1/2}\tilde{Q}_i(\tilde{Q}_0)^{-1/2}.^{7}$$

Then the above equations become

$$\overline{Q}_4 e^{-il(l+1)t}Y_{lm} = -i[(l+1-m)(l+1+m)]^{1/2} e^{-i(l+1)(l+2)t}Y_{l+1\,m},$$

$$\overline{Q}_5 e^{-il(l+1)t}Y_{lm} = -i[(l+m)(l-m)]^{1/2} e^{-i(l-1)lt}Y_{l-1\,m},$$

and the following operators are skew-adjoint under the scalar product defined above:

$$J_{23} = -\tfrac{1}{2}i(\overline{Q}_1 - \overline{Q}_2),$$

$$J_{31} = -\tfrac{1}{2}(\overline{Q}_1 + \overline{Q}_2), \quad J_{12} = \overline{Q}_3$$

$$J_{34} = -\tfrac{1}{2}i(\overline{Q}_4 - \overline{Q}_5), \quad J_{45} = \overline{Q}_0,$$

$$J_{53} = -\tfrac{1}{2}(\overline{Q}_4 + \overline{Q}_5),$$

$$J_{24} = \tfrac{1}{4}(\overline{Q}_7 - \overline{Q}_8 - \overline{Q}_9 - \overline{Q}_{10}),$$

$$J_{25} = -\tfrac{1}{4}i(\overline{Q}_7 - \overline{Q}_8 + \overline{Q}_9 - \overline{Q}_{10}),$$

$$J_{14} = \tfrac{1}{4}i(\overline{Q}_7 + \overline{Q}_8 - \overline{Q}_9 - \overline{Q}_{10}),$$

$$J_{15} = \tfrac{1}{4}(\overline{Q}_7 + \overline{Q}_8 + \overline{Q}_9 + \overline{Q}_{10}).$$

These operators satisfy the $O(3,2)$ algebra

$$[J_{ab}, J_{cd}] = -g_{bc}J_{ad} + g_{ac}J_{bd} - g_{ad}J_{bc} + g_{bd}J_{ac},$$

where $g_{11} = g_{22} = g_{33} = -g_{44} = -g_{55} = -1$ and $(J_{ab})^\dagger = -J_{ab}$.

Therefore the set $\{e^{-il(l+1)t}Y_{lm}(\theta,\phi)\}$ provides a basis for a unitary irreducible representation of $O(3,2)$. Although the dynamical group $O(3,2)$ of the rigid rotator is known,[8] the time-dependent Schrödinger equation was not used to derive it. It was shown in Paper II that all of our generators are the time-independent or time-dependent constants of the motion and, in this respect also, the results here differ from the results of previous workers.

3. SYMMETRIC TOP

The time-dependent Schrödinger equation for the symmetric top is given by[9]

$$\left\{-\frac{1}{2I_1}\left[(1-x^2)\partial_x^2 - 2x\partial_x + \left(\frac{I_1}{I_3} + \frac{x^2}{1-x^2}\right)\partial_\gamma^2 + (1-x^2)^{-1}\partial_\alpha^2 - \frac{2x}{1-x^2}\partial_\alpha\partial_\gamma\right] - i\partial_t\right\}\Psi = 0, \tag{3.1}$$

where $x = \cos\beta$, and α, β, γ are the Euler angles which determine the direction of the principal axes with respect to the space fixed coordinate axes. The solutions of this equation are expressed in terms of Wigner's D functions as

$$\Psi(\alpha,\beta,\gamma,t)=\sum_{jmn} C_{jmn}\frac{\sqrt{2j+1}}{4\pi} D^{j}_{mn}(\alpha,\beta,\gamma)e^{-E_{jn}t}$$
$$=\sum_{jmn} C_{jmn}\Psi^{j}_{mn}(\alpha,\beta,\gamma,t), \qquad (3.2)$$
$$E_{jn}=\frac{1}{2I_1}\cdot j(j+1)+\frac{I_1-I_3}{2I_1I_3}n^2,$$

where $j=0,\frac12,1,\frac32,\cdots,\ |m|<j,\ |n|<j$.[10]

To obtain a linear spectrum, we perform a time dilation using the operator defined by

$$D=\exp\{t\partial_t \ln[(J+\tfrac12/H]\},$$

where H is the Hamiltonian, and the operator $J+\frac12$ is defined by

$$J+\tfrac12=\tfrac12(1+8I_1\{H+\tfrac12[(I_1-I_3)/I_1I_3]\partial^2_\gamma\})^{1/2}.$$

Under this transformation Eq. (3.1) is transformed into

$$[(1-x^2)\partial^2_x-2x\partial_x+(1-x^2)^{-1}(\partial^2_\alpha+\partial^2_\gamma)$$
$$-2x(1-x^2)^{-1}\partial_\alpha\partial_\gamma+(i\partial_t)^2-\tfrac14]f=A\cdot f=0, \qquad (3.3)$$

where

$$f=\sum_{jmn} C_{jmn}\frac{\sqrt{2j+1}}{4}D^{j}_{mn}(\alpha,\beta,\gamma)e^{-i(j+1/2)t}.$$

It is clear that the operator $i\partial_t$ (not the operator $D\, i\partial_t D^{-1}$) has the linear spectrum $j+\frac12$ for the transformed eigenstates $D\Psi^{j}_{mn}$.

Now we determine the operator Q of the form

$$Q=Q^x\partial_x+Q^\alpha\partial_\alpha+Q^\gamma\partial_\gamma+Q^t\partial_t+Q^0, \qquad (3.4)$$

which satisfied the equation

$$AQf=0, \qquad (3.5)$$

where the operator A is defined in Eq. (3.3).

Expanding (3.5), and choosing the functions

$$f_{tt}, f_{xt}, f_{\alpha t}, f_{\gamma t}, f_{x\alpha}, f_{x\gamma}, f_{\alpha\alpha},$$
$$f_{\gamma\gamma}, f_{\alpha\gamma}, f_x, f_\alpha, f_\gamma, f_t, f,$$

for independent functions, we obtain fourteen determining equations:

$$-Q^t_t+Q^x_x+x(1-x^2)^{-1}Q^x=0,$$
$$-Q^x_t+(1-x^2)Q^t_x=0,$$
$$(1-x^2)Q^\alpha_t-Q^t_\alpha+xQ^t_\gamma=0,$$
$$(1-x^2)Q^\gamma_t-Q^t_\gamma+xQ^t_\alpha=0,$$
$$Q^x_\alpha-xQ^x_\gamma+(1-x^2)^2Q^\alpha_x=0,$$
$$Q^x_\gamma-xQ^x_\alpha+(1-x^2)^2Q^\gamma_x=0,$$
$$Q^\alpha_\alpha-xQ^\alpha_\gamma-Q^x_x-2x(1-x^2)^{-1}Q^x=0,$$
$$Q^\gamma_\gamma-xQ^\gamma_\alpha-Q^x_x-2x(1-x^2)^{-1}Q^x=0$$
$$Q^\alpha_\gamma-xQ^\alpha_\alpha+Q^\gamma_\alpha-xQ^\gamma_\gamma+2xQ^x_x$$
$$+(1+3x^2)(1-x^2)^{-1}Q^x=0,$$
$$(A+\tfrac14)Q^x+4xQ^x_x+2(1-x^2)^{-1}(1+x^2)Q^x$$
$$+2(1-x^2)Q^0_x=0,$$
$$(A+\tfrac14)Q^\alpha+2(1-x^2)^{-1}Q^0_\alpha-2x(1-x^2)^{-1}Q^0_\gamma=0,$$
$$(A+\tfrac14)Q^\gamma+2(1-x^2)^{-1}Q^0_\gamma-2x(1-x^2)^{-1}Q^0_\alpha=0,$$
$$(A+\tfrac14)Q^t-2Q^0_t=0,$$
$$(A+\tfrac14)Q^0+\tfrac12Q^x_x+\tfrac12x(1-x^2)^{-1}Q^x=0.$$

Solving these equations, we obtain the solution for Q in the form

$$Q\sum_{1}^{16}a^iQ_i=(3.4),$$

where the a^i are the integration constants and Q_i are defined by

$$Q_{\substack{1\\2\\3\\4}}=\{Q_{\substack{5\\6\\7\\8}}\}^*=e^{\frac12 i(t\substack{+\\-}\alpha\substack{+\\-\\+}\gamma)}\Big[\substack{-\\+\\+}\,i\Big(1\substack{+\\+\\-}x\Big)^{1/2}\Big(1\substack{-\\+\\+}x\Big)$$
$$\times\partial_x\substack{+\\+\\-}\Big(1\substack{+\\+\\-}x\Big)^{-1/2}\partial_\alpha\substack{+\\-\\+}\Big(1\substack{+\\+\\-}x\Big)^{-1/2}\partial_\gamma\substack{+\\-}\Big(1\substack{+\\+\\-}x\Big)^{1/2}\partial_t$$
$$+\tfrac12 i\Big(1\substack{+\\+\\-}x\Big)^{1/2}\Big],$$

$$Q_9=\{Q_{10}\}^*=e^{i\alpha}[(1-x^2)^{1/2}\partial_x-ix(1-x^2)^{-1/2}\partial_\alpha$$
$$+i(1-x^2)^{-1/2}\partial_\gamma],$$

$$Q_{11}=\{Q_{12}\}^*=e^{i\gamma}[(1-x^2)^{1/2}\partial_x-ix(1-x^2)^{-1/2}\partial_\gamma$$
$$+i(1-x^2)^{-1/2}\partial_\alpha],$$

$$Q_{13}=\partial_\alpha,\quad Q_{14}=\partial_\gamma,\quad Q_{15}=\partial_t,\quad Q_{16}=1.$$

As the a^i are arbitrary constants, each Q_i satisfies the equation (3.5) independently. To obtain the corresponding operators of the equation (3.1), we just perform the inverse transformation D^{-1} on each Q_i. Then the operators $\tilde{Q}_i=D^{-1}Q_iD$ satisfy[12] the equation

$$(H-i\partial_t)\tilde{Q}_i\Psi(\alpha,\beta,\gamma,t)=0 \qquad (3.6)$$

where $H-i\partial_t$ and Ψ are given in Eq. (3.1).

Equation (3.6) implies the relation

$$[H-i\partial_t,\tilde{Q}_i]=0 \quad (i=1,2,3,\ldots,16).$$

Therefore the $\tilde{Q}_i$ are the invariants which we are looking for. The action of these operators on the normalized eigenfunctions Ψ^{j}_{mn} defined in (3.2) are given by

$$\tilde{Q}_{\substack{1\\2\\3\\4}}\Psi^{j}_{mn}=\substack{-\\+\\+\\-}\sqrt2\, i\Big(\frac{2j+1}{2j}\Big)^{1/2}\Big[\Big(j\substack{-\\+\\+}m\Big)\Big(j\substack{-\\+\\+\\-}m\Big)\Big]^{1/2}$$
$$\Psi^{j-1/2}_{m\substack{+\\-\\+\\-}1/2,\ n\substack{+\\-\\-\\+}1/2} \qquad (3.7)$$

$$\tilde{Q}_{\substack{5\\6\\7\\8}}\Psi^j_{mn} = \substack{-\\ \mp\\ \pm\\ -}\sqrt{2}i\left(\frac{2j+1}{2j+2}\right)^{1/2}\left[\left(j \substack{-\\ \mp\\ +} m+1\right)\left(j \substack{-\\ \mp\\ \pm\\ -} n+1\right)\right]^{1/2} \Psi^{j+1/2}_{m\substack{\mp\\ \mp\\ +}1/2,\,n\substack{\mp\\ \pm}1/2}, \quad (3.8)$$

$$\tilde{Q}_{\substack{9\\10}}\Psi^j_{mn} = \pm[(j\mp n)(j\pm n+1)]^{1/2}\Psi^j_{m\pm1,n}$$

$$\tilde{Q}_{\substack{11\\12}}\Psi^j_{mn} = \mp[(j\mp n)(j\pm n+1)]^{1/2}\Psi^j_{m,n\pm1},$$

$$\tilde{Q}_{13}\Psi^j_{mn} = im\Psi^j_{mn}, \quad \tilde{Q}_{14}\Psi^j_{mn} = in\Psi^j_{mn},$$

$$\tilde{Q}_{15}\Psi^j_{mn} = -i(j+\tfrac{1}{2})\Psi^j_{mn}.$$

From these results it is clear that the following operators shift the eigenvalue j alone by one unit:

$$\begin{aligned}
Q_- &= \tilde{Q}_1\tilde{Q}_2 = -\tilde{Q}_3\tilde{Q}_4 \\
&= D^{-1}\{-2e^{it}[\partial_\alpha\partial_\gamma - x\partial_t^2 + i(1-x^2)\partial_x\partial_t \\
&\quad -\tfrac{1}{2}(1-x^2)\partial_x - ix\partial_t + \tfrac{1}{4}x]\}D, \\
Q_+ &= \tilde{Q}_5\tilde{Q}_6 = -\tilde{Q}_7\tilde{Q}_8 \\
&= D^{-1}\{-2e^{-it}[\partial_\alpha\partial_\gamma - x\partial_t^2 - i(1-x^2)\partial_x\partial_t \\
&\quad -\tfrac{1}{2}(1-x^2)\partial_x + ix\partial_t + \tfrac{1}{4}x]\}D. \qquad (3.9)
\end{aligned}$$

As these give rise to two term recursion relations among the functions D^j_{mn}, they will be useful for practical purposes.

To elucidate the group theoretical properties of the differential equation (3.1), we introduce the new operators $\overline{Q}_i$ defined by

$$\overline{Q}_i = (\tilde{Q}_{15})^{1/2}\tilde{Q}_i(\tilde{Q}_{15})^{-1/2}.$$

This is necessary to obtain skew-adjoint operators.[13]

Then the operators C^i_j defined by

$$C^1_2 = -\overline{Q}_9, \quad C^2_1 = \overline{Q}_{10}, \quad C^1_3 = (1/\sqrt{2})\overline{Q}_8,$$

$$C^3_1 = (1/\sqrt{2})\overline{Q}_4, \quad C^1_4 = (1/\sqrt{2})\overline{Q}_6, \quad C^4_1 = (1/\sqrt{2})\overline{Q}_2,$$

$$C^2_3 = (1/\sqrt{2})\overline{Q}_5, \quad C^3_2 = (1/\sqrt{2})\overline{Q}_1, \quad C^2_4 = -(1/\sqrt{2})\overline{Q}_7,$$

$$C^4_2 = -(1/\sqrt{2})\overline{Q}_3, \quad C^3_4 = -\overline{Q}_{11}, \quad C^4_3 = \overline{Q}_{12},$$

$$C^1_1 = i(\overline{Q}_{13} - \overline{Q}_{15}), \quad C^2_2 = i(-\overline{Q}_{13} - \overline{Q}_{15}),$$

$$C^3_3 = i(\overline{Q}_{14} + \overline{Q}_{15}), \quad C^4_4 = i(-\overline{Q}_{14} + \overline{Q}_{15})$$

satisfy the $SL(4,R)$ algebra,

$$[C^i_j, C^k_l] = \delta^i_l C^k_j - \delta^k_j C^i_l,$$

and following linear combinations of the C^i_j,

$$X^k_k = iC^k_k, \quad k = 1,2,3,4,$$

$$X^k_l = i(C^k_l + C^l_k), \quad X^l_k = C^l_k - C^k_l, \quad k < l$$
with $k, l = 1, 2$

$$X^k_l = i(C^k_l + C^l_k), \quad X^l_k = C^l_k - C^k_l, \quad k < l$$
with $k, l = 3, 4$

$$X^k_l = i(C^k_l - C^l_k), \quad X^l_k = -C^l_k - C^k_l, \quad k = 1, 2,$$
$l = 3, 4,$

satisfy the commutation relations of $SU(2,2)$[14] and are skew-adjoint under the $SU(2)$ scalar product

$$(f,g) = \int_{2\pi}^{2\pi}\int_0^{2\pi}\int_0^{\pi} f^*g \sin\beta d\beta d\alpha d\gamma. \qquad (3.10)$$

Thus it is clear that the set $\{\Psi^j_{mn}(\alpha,\beta,\gamma,t)\,|\, j = 0, \frac{1}{2}, 1, \cdots, -j < m, n < j\}$, which are the matrix elements of the regular representation of $SU(2)$, provide a basis for a unitary irreducible representation of $SU(2,2)$.

The $SU(2,2)$ group contains a variety of subgroups. Here we investigate the $SU(2)\times SU(2)$ subgroup generated by the operators $\overline{Q}_i$ $(i = 9, 10, \ldots, 14)$. For the case of the spherical top $(I_1 = I_2 = I_3)$, we have the identities $\overline{Q}_i = Q_i$ $(i = 9, 10, \ldots, 14)$ because the Q_i commute with both the dilation operators D and the operator Q_{15}. One can easily check that they also commute with the Hamiltonian. Therefore, they comprise the well-known $SU(2)\times SU(2)$ degeneracy group of the spherical top.[15] On the other hand, for the case of a symmetric top $(I_1 = I_2 \neq I_3)$, we have the identities $\overline{Q}_i = Q_i$ only for $i = 9, 10, 13, 14$, and therefore the Q_i commute with the Hamiltonian, so that degeneracy group will be $SU(2)\times U(1)$. The $\overline{Q}_{11}$ and $\overline{Q}_{12}$ are no longer time-independent constants of the motion. Here we have obtained an important result: Under the symmetry breaking $I_1 = I_2 = I_3 \to I_1 = I_2 \neq I_3$ two of the time-independent constants of the motion, Q_{11} and Q_{12}, of the spherical top turn into the time-dependent ones. This is an example of a more general phenomenon which will be discussed in detail in a future communication.

So far, we have found that the eigenstates $\{\Psi^j_{mn}(\alpha,\beta,\gamma,t)\}$, $j = 0, \frac{1}{2}, 1, \ldots, -j \leqslant m, n \leqslant j$, which comprise the regular representation of $SU(2)$, also form the basis for a unitary irreducible representation of $SU(2,2)$. However, it is clear that physically the integer and half-odd integer states cannot be mixed. If the mixing is allowed, the probability amplitude $|\Psi|^2$ is no longer invariant under a rotation of 360°.[16] Therefore, if there exists some observable which causes the mixing of integer and half odd integer states, one type of state has to be eliminated. If we restrict ourselves to the integer or half odd-integer states, then $SU(2,2)$ is no longer the dynamical group in the ordinary sense. In this case the dynamical group may be generated by the operators $\{\overline{Q}_+, \overline{Q}_-, \overline{Q}_9, \overline{Q}_{10}, \overline{Q}_{11}, \overline{Q}_{12}, \overline{Q}_{13}, \overline{Q}_{14}, \overline{Q}_{15}\}$, where the operators $\overline{Q}_+$ and $\overline{Q}_-$ are defined by

$$\overline{Q}_\pm = (\overline{Q}_{15})^{1/2}Q_\pm(\overline{Q}_{15})^{-1/2},$$

where the $Q_\pm$ are given by (3.9). However, these operators do not close under a finite number of commutation operations, that is, they generate an infinite dimensional Lie algebra.

4. CONCLUSION

Employing the method discussed in our previous papers I–III,[2,3] we have derived spectrum generating groups of the rigid rotator and the symmetric top with minimal ingenuity. All the elements of the groups are constants of the motion. Two of the time-independent constants of the motion of the spherical top are shown to be continuously connected with two of the time-independent constants of the motion of the symmetrical top.

We note once again that one cannot in general assume the form (2.4) or (3.4) for the generator of a Lie group leaving invariant a second-order partial differential equation. In fact, if one supposes the invariants of Eqs. (2.1), (3.1) are of this form, one will fail to obtain the energy shift operators. The reason becomes quite clear if we look at the form of these operators, for example, the operator $\tilde{Q}_4$ (not Q_4) of the rigid rotator. It contains infinitely many derivatives, and therefore the simple forms (2.4) or (3.4) for Q never give the energy shift operators. Therefore, if we want to get the energy shift operators by applying our method directly to the original equation (2.1) or (3.1), we have to allow infinitely many derivatives in Q. This appears to create a most cumbersome problem. But, as is shown by these and other examples,[2,3] the dilation technique can be effective in transforming to a problem involving at most a finite number of derivatives. Although the method presented here suffers from the severe restriction that one needs a knowledge of the spectrum to construct the dilation operator, it is still very useful in obtaining spectrum generating algebras for Schrödinger equations with known spectrum.

ACKNOWLEDGMENT

We are pleased to acknowledge the support of this research by a grant from the Research Corporation.

*From the M. S. thesis of S. Kumei.

[1]S. Lie, *Gesammelte Abhandlungen*, edited by F. Engel and P. Heegaard (Tuebner, Leipzig, 1922), Bd. III and suppl.; *Vorlesungen über Differentialgleichungen*, edited by Georg Scheffers (Teubner, Leipzig, 1891).

[2]R. L. Anderson, S. Kumei, and C. E. Wulfman, Phys. Rev. Lett. **28**, 988 (1972).

[3]R. L. Anderson, S. Kumei, and C. E. Wulfman, (a) Rev. Mex. Fis. **21**, 1 (1972); (b) Rev. Mex. Fis. **21**, 35 (1972).

[4]Or equivalently map the set of solutions of the given system into itself.

[5]The detailed discussion of the linearization of the spectrum will be found in Paper II.

[6]As Q_1, Q_2, Q_3 commute with the operator D, we have the identities $\tilde{Q}_i = Q_i$ for $i = 1,2,3$.

[7]As $\tilde{Q}_0$, Q_1, $\tilde{Q}_2$, $\tilde{Q}_3$ commute with $\tilde{Q}_0$, we have the identities $\tilde{Q}_i = Q_i$ for $i = 1,2,3,0.3,0$.

[8]E. Kyriakpoulos, Phys. Rev. **174**, 1846 (1968).

[9]For example, A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1957), p. 66.

[10]We use the same D function as defined in Edmonds, Ref. 9, p. 58, but the eigenstates Ψ^j_{mn} are orthonormal under the $SU(2)$ scalar product given in Eq. (3,10). Extensive group theoretical discussion of D functions will be found in Vilenkin's book.[11]

[11]N. J. Vilenkin, *Special Functions and the Theory of Representations* (Am. Math. Soc., Providence, R. I., 1968).

[12]It is important to notice that the commutation relations among the Q_i are not changed under this transformation.

[13]The factors $(Q_{15})^{(1/2)}$ and $(Q_{15})^{-(1/2)}$ in this expression remove the factors $\{(2j+1)/[2(j \pm 1)+1]\}^{(1/2)}$ of the coefficients in Eqs. (3.7) and (3.8). However, it is clear that the set $\tilde{Q}_i$ still satisfy the same commutations as the sets $\{Q_i\}$ and $\{Q_i\}$.

[14]A general discussion of the group $U(p, q)$ in relation to the group $GL(p+q, R)$ will be found in R. L. Anderson, J. Fisher, and R. Rączka, Proc. R. Soc. A 302, 491 (1968).

[15]A. W. Sáenz, Ph.D. thesis (University of Michigan, 1948).

[16]D. Bohm, *Quantum Mechanics* (Prentice-Hall, Englewood Cliffs, N. J., 1961), p. 389.

Positivity for some generalized Yukawa models in one space dimension*

Paul Federbush

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48104
(Received 22 February 1973)

For the model $(g\bar{\psi}\psi\phi^N + \phi^{2M})_{1+1}$ in a box the energy is bounded below if $M > N$.

1. INTRODUCTION

The model $(g\bar{\psi}\psi\phi^N + \phi^{2M})_{1+1}$ is considered in a box. For $M > N$ the energy is shown to be bounded below. It is not known whether the energy is bounded below for $M = N$, except of course for $M = N = 1$, and it is annoying that the method of the present paper must be modified to accommodate even this case.

Boundedness below is shown by finding an operator bound for e^{-H}. Writing

$$H = \bar{H} + N_{\tau B}, \tag{1.1}$$

we note the operator norm inequality

$$|e^{-H}| \le |e^{-N_\tau B/2} e^{-\bar{H}} e^{-N_\tau B/2}| \tag{1.2}$$

whose proof is given in Sec. 3. $e^{-\bar{H}}$ is next expanded in a Duhamel expansion, where in successive terms of the expansion more and more of the Fermion pair creation and annihilation terms (terms relevant to renormalization) are included in the exponent. This is adapted from the method of Ref. 1. Lower bounds for the portion of the interaction kept in the exponent, substituting for Wick-ordering bounds of the boson case, are obtained by use of an approximate dressing transformation in Sec. 4.

Next considered, in Sec. 5, are places in the Duhamel expansion where a fermion pair annihilation term immediately follows a fermion pair creation term in the nonexponentiated interaction. In this case the fermion operators are "pulled through" to obtain a normal ordered expression in the fermion operators in the neighboring pair. Terms where the fermions contract to a loop require renormalization cancellations. Be it noted, no fermion operator is "pulled across" more than one exponent.

To exhibit renormalization cancellations it is necessary to "pull-across" boson operators from one side of the closed fermion loop to the other. This leaves an expression involving boson annihilation operators as well as ϕ's. The annihilation operators are "pulled" to the right till they either contract away or hit the right exponent $e^{-N_\tau B/2}$ from (1.1). (In many other calculations, Ref. 2 for example, there is the vacuum available to kill annihilation operators.)

After the renormalization cancellations are exhibited and the annihilation operators are "pulled across," all the interactions not in the exponents involve only boson ϕ's (and not π's). It is nonetheless necessary to decompose some of the ϕ's into a's and a^*'s and "pull through" to contraction or the $e^{-N_\tau B/2}$ are hit. This is to leave expressions that can be dominated by the exponent ϕ^{2M} terms. These additional "pull-throughs" are defined in Sec. 6.

After the succession of fermion "pull-throughs," renormalization cancellations, and boson "pull-throughs" there remains a sum of the type

$$\sum e^{-N_\tau B/2} a^* \cdots a^* O a \cdots a e^{-N_\tau B/2} \tag{1.3}$$

The O are products of exponentials and interactions involving ϕ's, boson kinetic energy terms, and fermion operators; these products are integrated over with respect to the t_i parameters of the Duhamel expansion. O is realized as a product of unit blocks. The existence of a path space integral with positive measure enables us to estimate the product of unit blocks first treating the ϕ as numerical objects and using N_τ estimates for the fermion operators. There remains an expression in ϕ's only, integrated over path space; the switch to the path space formalism involves removal of boson kinetic energy terms. The integral over path space is then replaced by an operator expression in the ϕ, with the boson kinetic energy terms reappearing. Interaction terms in the ϕ are dominated by $\tilde{H}_{0B} + \int :\phi^{2M}:$ terms in the exponents, creation and annihilation operators by $e^{-N_\tau B/2}$.

One uses boundedness below for Hamiltonians of the form

$$H = N_B + \int :\phi^{2M}: + c\left|\int :\phi^s:\right|^r, \tag{1.4}$$

where $rs < 2M$ and r is not necessarily integral, and similar Hamiltonians, employing Nelson's original method to estimate

$$\langle 0 | e^{-V} | 0 \rangle \tag{1.5}$$

for V an expression in the ϕ (Ref. 3).

The overall estimates in Sec. 8 involve critically the "phase space," or estimates for the t_i integrals, and conditions on a number of parameters, such as the parameter determining how much of the fermion pair creation term is kept in the exponent.

The Appendix pursues the passage to boson path space with respect to the resultant "time ordering" of the fermion variables. It is there noted that in fact the operator $e^{-N_\tau B/2} e^{-\bar{H}} e^{-N_\tau B/2}$ is an analytic function of g.

The sequence of operations in the present calculation is far from unique, permitting many possible modifications. Application to further problems will naturally select the best lines of development. An obvious next step is the incorporation of the present methods with localization methods to treat boson–fermion models. This is presumably necessary to handle the infinite volume limit.

2. THE MODEL

The $(g\bar{\psi}\psi\phi^{2M})$ model is assembled in a box of length 2π with periodic boundary conditions:

$$\phi(x) = \sum \frac{1}{\sqrt{2\omega_k}} \frac{1}{\sqrt{2\pi}} (a_k e^{ikx} + a_k^* e^{-ikx}) = \sum \phi_k e^{ikx}. \tag{2.1}$$

Copyright © 1973 by the American Institute of Physics

We also define a Fourier decomposition for a Wick-ordered power of ϕ:

$$:\phi^n(x): = \sum_k :\phi^n:_k e^{ikx}. \tag{2.2}$$

The Hamiltonian is

$$H = H_{0B} + H_{0F} + V + \Delta + \int :\phi^{2M}:. \tag{2.3}$$

These quantities are defined along with others that will be useful:

$$H_{0B} = \sum \omega_n a_n^* a_n = \sum (\omega_n - \omega_n^\tau) a_n^* a_n + \sum \omega_n^\tau a_n^* a_n$$
$$= \sum \tilde{\omega}_n a_n^* a_n + \sum \omega_n^\tau a_n^* a_n = \tilde{H}_{0B} + N_{\tau B}, \tag{2.4}$$

$$H_{0F} = \sum \mu_n (b_n^* b_n + b_n'^* b_n'), \tag{2.5}$$

$$H_{0F}^{(-R)} = \sum_{|n|>R} \mu_n (b_n^* b_n + b_n'^* b_n'), \tag{2.6}$$

$$\tilde{H}_{0F}^{(R)} = \sum_{|n|\le R} \tilde{\mu}_n (b_n^* b_n + b_n'^* b_n'), \qquad \tilde{\mu}_n = \mu_n - \tfrac{1}{2}\mu_n^{\tau\prime}, \tag{2.7}$$

$$N_{\tau F}^{(R)} = \sum_{|n|\le R} \mu_n^{\tau\prime} (b_n^* b_n + b_n'^* b_n'), \tag{2.8}$$

$$\bar{H} = H - N_{\tau B} = \tilde{H}_{0B} + H_{0F} + V + \Delta + \int :\phi^{2M}:. \tag{2.9}$$

V is split into scattering and pair creation-annihilation terms:

$$V = g\int :\bar{\psi}\psi\phi^N: = V_s + V_p,$$
$$V_s = \sum_{i+j+k=0} :\phi^N:_i (v b_{-j}^* b_k + v b_{-j}'^* b_k'), \tag{2.10}$$
$$V_p = \sum_{i+j+k=0} :\phi^N:_i (v b_{-j}'^* b_{-k}^* + v b_k b_j').$$

Here the v dependence on i,j,k is suppressed. We have

$$\Delta = \sum_{r=0}^{N} \binom{N}{r}^2 (r!) \sum_k :\phi_k^{N-r}\phi_{-k}^{N-r}: h(r), \tag{2.11}$$

$$h(r) = \sum_{p_1+p_2+k_1+\dots+k_r=0} vv \prod_{m=1}^{r}\left(\frac{1}{2\pi}\frac{1}{2\omega_{k_m}}\right) \times\left(\sum_{i=1}^{r}\omega_{k_i} + \mu_1 + \mu_2\right)^{-1}. \tag{2.12}$$

United in Δ is the collection of renormalization terms. Properly there is an upper momentum cut-off; with only estimates used that are independent of the cut-off. This ideal upper cut-off is ignored. By $\Delta^{(R)}$ is meant the portion of the renormalization obtained from fermion loop contributions containing fermion momenta less than or equal R in absolute value. Similarly,

$$V_p^{(R)} = \sum_{\substack{i+j+k=0\\ |j|,|k|\le R}} :\phi^N:_i (v b_{-j}'^* b_{-k}^* + v b_k b_j'). \tag{2.13}$$

R will assume values

$$R_i = i^\alpha, \tag{2.14}$$

where α is an enormous constant to be later specified. We abbreviate $\Delta^{(R_i)}$ by $\Delta^{(i)}$.

3. THE DUHAMEL EXPANSION

We begin by proving estimate (1.2). We desire to show

$$|e^{-(A+B)}| \le |e^{-A/2}e^{-B}e^{-A/2}|. \tag{3.1}$$

We indicate two proofs. The first follows:

$$|e^{-A/2}e^{-B}e^{-A/2}| \le c \iff e^{-B} \le ce^{+A}. \tag{3.2}$$

There follows (from $x^2 \ge y^2 \implies x \ge y$ for positive operators x and y)

$$e^{-B/2^k} \le c^{1/2^k} e^{+A/2^k} \tag{3.3}$$

or

$$e^{-A/2^{k+1}} e^{-B/2^k} e^{-A/2^{k+1}} \le c^{1/2^k}.$$

Trotter's product formula then implies

$$e^{-(A+B)} \le c \iff |e^{-(A+B)}| \le c.$$

Alternatively one may use the result that if x and y are positive operators, $x \ge y \implies \ln x \ge \ln y$. From (3.2) follows

$$e^{-B} \le ce^{+A} \implies -B \le \ln c + A$$

or

$$-(A+B) \le \ln c \implies e^{-(A+B)} \le c.$$

(All estimates needed may be applied to the case where only a finite number of boson and fermion modes are kept, making domain questions trivial.) We also need later the result for positive operators x and y

$$x \ge y \implies x^c \ge y^c, \qquad 0 \le c \le 1 \tag{3.4}$$

Since (1.2) is to be used we find a Duhamel expansion for $e^{-\bar{H}}$. We define partitions of H:

$$\bar{H} = A_i + B_i, \tag{3.5}$$

$$A_i = H_{0F} + \tilde{H}_{0B} + V_p^{(i)} - D^{(i)} + \int :\phi^{2M}: \tag{3.6}$$

($D^{(i)}$ is defined in the next section). We have

$$B_i = V_s + (V_p - V_p^{(i)}) + (\Delta - \Delta^{(i)}) + (\Delta^{(i)} + D^{(i)}). \tag{3.7}$$

B_i is further split:

$$B_i = B_{i,1} + B_{i,2}, \tag{3.8}$$

$$B_{i,1} = V_s + (V_p - V_p^{(i)}),$$
$$B_{i,2} = (\Delta - \Delta^{(i)}) + (\Delta^{(i)} - D^{(i)}). \tag{3.9}$$

We use the expansion

$$e^{-\bar{H}} = e^{-A_1} - \int_0^1\int_0^1 dt_1 dt_2\, \delta(t_1+t_2-1) \times (e^{-A_2 t_2}B_{1,1}e^{-A_1 t_1} + e^{-A_3 t_2}B_{1,2}e^{-A_1 t_1}) + \cdots$$
$$+ (-1)^{n+1}\int_0^1 \int_0^1 dt_1\cdots dt_n\, \delta\left(\sum t_i - 1\right) \times \sum_{\substack{\alpha_1,\dots,\alpha_n\\ \alpha_i\in(1,2)}} e^{-A_{F_n}t_n}\cdots B_{F_2,\alpha_2}e^{-A_{F_2}t_2}B_{F_1,\alpha_1}e^{-A_1 t_1} + \cdots, \tag{3.10}$$

where

$$F_k = 1 + \sum_{i=1}^{k-1}\alpha_i. \tag{3.11}$$

F_k may also be defined inductively:

$$F_1 = 1, \qquad F_k = \alpha_{k-1} + F_{k-1}. \tag{3.12}$$

B_i has been split to facilitate renormalization cancellations.

4. DRESSING TRANSFORMATION

An approximate dressing transformation, as first introduced in Ref. 4, is now used to enable the A_i of (3.6) to be bounded below. Differing from Ref. 4, here only the fermion operators are dressed. Though "less accurate," this leads to simpler expressions sufficient for our purposes, and—all important—involving only ϕ's.

We define the dressed operators $\tilde{b}_k, \tilde{b}_j'$ (which depend on i)

$$\tilde{b}_k = b_k - \sum_{\substack{l+j+k=0\\ |j|\le R_i}} :\phi^N:_{-l} \frac{1}{\tilde{\mu}_j + \tilde{\mu}_k} v b_j'^{*}, \tag{4.1}$$

$$\tilde{b}_j' = b_j' + \sum_{\substack{l+j+k=0\\ |k|\le R_i}} :\phi^N:_{-l} \frac{1}{\tilde{\mu}_j + \tilde{\mu}_k} v b_k^{*}, \tag{4.2}$$

$$\tilde{b}_k^{*} = b_k^{*} - \sum_{\substack{l+j+k=0\\ |j|\le R_i}} :\phi^N:_{+l} \frac{1}{\tilde{\mu}_j + \tilde{\mu}_k} v b_j', \tag{4.3}$$

$$\tilde{b}_j'^{*} = b_j'^{*} + \sum_{\substack{l+j+k=0\\ |k|\le R_i}} :\phi^N:_{+l} \frac{1}{\tilde{\mu}_j + \tilde{\mu}_k} v b_k. \tag{4.4}$$

Expressing A_i in terms of these operators,

$$A_i = \tilde{H}_{0B} + H_{0F}^{(-i)} + \tfrac{1}{2} N_{\tau' F}^{(i)} - D^{(i)} + \int :\phi^{2M}: + P^{(i)} + w^{(i)} \tag{4.5}$$

with

$$P^{(i)} = \sum_{|k|\le R_i} \tilde{\mu}_k (\tilde{b}_k^{*}\tilde{b}_k + \tilde{b}_k'^{*}\tilde{b}_k') \tag{4.6}$$

and $w^{(i)}$ contains terms quadratic in g in terms of the undressed operators. $w^{(i)}$ as naturally expressed is initially antinormal ordered in the fermion operators. When $w^{(i)}$ is normal ordered it splits into two terms:

$$w^{(i)} = Q^{(i)} + D^{(i)}, \tag{4.7}$$

where $Q^{(i)}$ is quadratic in (undressed) fermion operators and $D^{(i)}$ contains no fermion operators—a contracted fermion loop. Equation (4.7) is illustrated in Figure 1. The great thing here (more manifest than a similar observation in Ref. 4) is that

$$Q^{(i)} \ge 0 \tag{4.8}$$

being a sum of the form $\sum A^{*}A$. Of course,

$$P^{(i)} \ge 0 \tag{4.9}$$

also. A useful expression for A_i follows:

$$A_i = (P^{(i)} + Q^{(i)}) + (\tilde{H}_{0B} + H_{0F}^{(-i)} + \tfrac{1}{2} N_{\tau' F}^{(i)}) + \int :\phi^{2M}:. \tag{4.10}$$

It is important for our purposes that $P^{(i)}$, $Q^{(i)}$, and $D^{(i)}$ depend on fermion operators and boson ϕ's only.

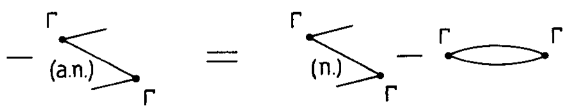

FIG. 1. Normal ordering after partial dressing, where n. and a.n. stand for normal ordered and antinormal ordered, respectively.

5. RENORMALIZATION CANCELLATION

The renormalization cancellation is the heart of the calculation. As mentioned before, after the cancellation is exhibited some boson operators must be "pulled-through" until they contract, or reach the $e^{-N_\tau B/2}$ at the edges of Eq. (1.2).

We decompose V from (2.12) into its creation and annihilation terms

$$V_p = V_{pc} + V_{pa}. \tag{5.1}$$

In the nth terms in (3.10) involving $t_1, \cdots, t_n$ we define time variables $s_0, \cdots, s_n$

$$s_0 = 0, \quad s_n = 1, \quad s_k = \sum_{i=1}^{k} t_i. \tag{5.2}$$

Each $B_{l,m}$ in (3.10) is associated to some s_i. The nth term in (3.10) is a sum of 2^n terms by virtue of the sum over α_i. We further subdivide each such term. The term $B_{i,1}$ is split into three terms

$$B_{i,1} = B_{i,s} + B_{i,a} + B_{i,c}, \tag{5.3}$$

$B_{i,s}, B_{i,a}$ and $B_{i,c}$, respectively, containing terms from V_s, $V_{p,a}$, and $V_{p,c}$. The nth term now is decomposed into 4^n terms. The renormalization cancellation takes place for a unit of the following type from the nth term of (3.10):

$$\begin{array}{ccc} s_{k+1} & s_k & s_{k-1} \\ \cdot\ \ \cdot & \cdot & \cdot \end{array}$$

$$B_{F_{k+1},a}\, e^{-A_{F_{k+1}} t_{k+1}}\, B_{F_k,c}\, e^{-A_{F_k} t_k} \tag{5.4}$$

against a unit of the following type

$$\begin{array}{ccc} s_{k+1} & s_k & s_{k-1} \\ \cdot & \cdot & \cdot \end{array}$$

$$B_{F_k,2}\, e^{-A_{F_k}(t_k + t_{k+1})} \tag{5.5}$$

from the $(n-1)$th term of (3.10). Here the sequence $s_n, \cdots, s_0$ from the nth term is associated to $s_n, \cdots, s_{\hat{k}}, \cdots, s_0$ from the $(n-1)$st term. See Fig. 2. After extracting a certain portion of (5.4), s_k when integrated from s_{k-1} to s_{k+1} allows a renormalization cancellation between (5.5) and the integrated over portion of (5.4).

We first use an identity on the first exponential in (5.4). We have

$$e^{-A_{F_{k+1}} t_{k+1}} = e^{-A_{F_k} t_{k+1}} - \int_0^{t_{k+1}} e^{-A_{F_{k+1}}(t_{k+1}-t)} \times (A_{F_{k+1}} - A_{F_k}) \cdot e^{-A_{F_k} t}\, dt. \tag{5.6}$$

Substituting (5.6) into (5.4) the second term makes a contribution irrelevant to renormalization we call a J_1 term. The first term replaces (5.4) by

$$B_{F_{k+1},a}\, e^{-A_{F_k} t_{k+1}}\, B_{F_k,c}\, e^{-A_{F_k} t_k}. \tag{5.7}$$

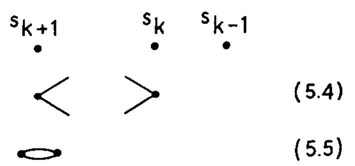

FIG. 2. The renormalization cancellation.

We now split (5.7) into two terms, the first,

$$B^{+}_{F_{k+1},a}e^{-A_{F_k}t_{k+1}}B^{+}_{F_k,c}e^{-A_{F_k}t_k}, \tag{5.8}$$

contains the terms in the B where all four fermion momenta are greater in absolute value than $R_{F_{k+1}}$; the other terms we denote by J_2. In (5.8) we pull all the fermion operators across the exponent $e^{-A_{F_k}t_{k+1}}$ until we have a normal ordered expression in these four fermion operators as indicated in Fig. 3.

See Ref. 5 for a definition of the "push-pull" operations. The first term on the right side of Fig. 3 is the portion of (5.4) involved in the renormalization cancellations:

$$\sum_{\substack{p=p_1+p_2\\ |p_2|,|p_1|>R_{F_{k+1}}}} vv:\phi^N:_p e^{-A_{F_k}t_{k+1}}:\phi^N:_{-p}e^{-A_{F_k}t_k} \times e^{-(\mu_1+\mu_2)t_{k+1}}. \tag{5.9}$$

To exhibit the renormalization it now appears necessary to push the boson operators from the $:\phi^N:_{-p}$ term in (5.9) across the exponent. This is done one boson operator at a time. We leave some annihilation operators to the right side of the exponent, pushing across the right mixture of a^*s and a's so only ϕ's arrive across the exponent. The term where all ϕ's cross the exponent is relevant to renormalizations, terms in the unpushed annihilation operators or involving contractions with the exponent are not. As soon as a contraction occurs or an annihilation term appears we stop pulling boson operators across the exponent. We now go into details.

We define

$$\tilde{\phi}_p(s)=\frac{1}{\sqrt{2\omega_p}}\frac{1}{\sqrt{2\pi}}\left(a_p e^{-2\tilde{\omega}_p(t_{k+1}-s/2)}+a^*_{-p}e^{-\tilde{\omega}_p s}\right), \tag{5.10}$$

noting

$$\tilde{\phi}_p(t_{k+1})=e^{-\tilde{\omega}_p t_{k+1}}\phi_p,\qquad \tilde{\phi}_p(0)=\phi_p-(1-e^{-2\tilde{\omega}_p t_{k+1}})a_p. \tag{5.11}$$

We also define $\hat{A}_{F_k}$:

$$A_{F_k}=\hat{A}_{F_k}+\tilde{H}_{0B}. \tag{5.12}$$

We consider the heart of (5.9):

$$:\phi^N:_p e^{-A_{F_k}t_{k+1}}:\phi^N:_{-p} = \sum_{k_1+\cdots+k_N=-p}:\phi^N:_p e^{-A_{F_k}t_{k+1}}:\phi_{k_1}\cdots\phi_{k_N}:. \tag{5.13}$$

Pulling across the $\tilde{\phi}_{k_i}(s)$, the term where all $\tilde{\phi}_{k_i}(s)$ come across the exponent yields

$$\sum_{k_1+\cdots+k_N=-p}:\phi^N:_p:\phi_{k_1}\cdots\phi_{k_N}: \times e^{-(\sum_{i=1}^N\tilde{\omega}_{k_i})t_{k+1}}e^{-A_{F_k}t_{k+1}}. \tag{5.14}$$

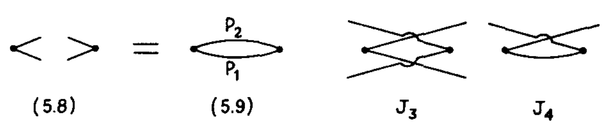

FIG. 3. Graphical illustration of fermion pull-throughs.

The terms in unpushed annihilation operators are as follows:

$$\sum_{s=0}^{N-1}\sum_{k_1+\cdots+k_N=-p}:\phi^N:_p:\phi_{k_1}\cdots\phi_{k_s}:e^{-(\sum_{i=1}^s\tilde{\omega}_{k_i})t_{k+1}} \times e^{-A_{F_k}t_{k+1}}:\phi_{k_{s+1}}\cdots\phi_{k_{N-1}}:(1-e^{-2\tilde{\omega}_{k_N}t_{k+1}})a_{k_N} \tag{5.15}$$

which we refer to as J_5, and terms subtracting from (5.15) contractions between the pulled across and unpulled across terms, explicitly

$$-\sum_{s=1}^{N-2}\sum_{k_1+\cdots+k_N=-p}\sum_{l=1}^{\min(s,N-s-1)}:\phi^N:_p\, l!\binom{s}{l}\binom{N-s-1}{l} \times :\phi_{k_{l+1}}\cdots\phi_{k_s}:e^{-(\sum_{i=1}^s\tilde{\omega}_{k_i})t_{k+1}} \times e^{-A_{F_k}t_{k+1}}:\phi_{k_{l+s+1}}\cdots\phi_{k_{N-1}}:(1-e^{-2\tilde{\omega}_{k_N}t_{k+1}})a_{k_N} \times\prod_{m=1}^{l}\left(\frac{1}{2\pi}\frac{1}{2\omega_{k_m}}\delta_{k_m,-k_{s+m}}\right) \tag{5.16}$$

which we refer to as J_6. The terms in contractions with the exponent are

$$\sum_{s=0}^{N-1}\sum_{k_1+\cdots+k_N=-p}:\phi^N:_p:\phi_{k_1}\cdots\phi_{k_s}:e^{-(\sum_{i=1}^s\tilde{\omega}_{k_i})t_{k+1}} \times\int_0^{t_{k+1}}ds'e^{-A_{F_k}(t_{k+1}-s')}[\tilde{\phi}_{k_{s+1}}(s'),\hat{A}_{F_k}] \times e^{-A_{F_k}s'}:\phi_{k_{s+2}}\cdots\phi_{k_N}: \tag{5.17}$$

which we call J_7, and terms in contractions analagous to (5.16) we call J_8. Thus

$$(5.13)=(5.14)+J_5+J_6+J_7+J_8. \tag{5.18}$$

Equation (5.14) replaced in (5.9) yields

$$\sum_{\substack{p=p_1+p_2\\ |p_2|,|p_1|>R_{F_{k+1}}}}\sum_{k_1+\cdots+k_N=-p}vv:\phi^N:_p:\phi_{k_1}\cdots\phi_{k_N}: \times e^{-(\sum_{i=1}^N\tilde{\omega}_{k_i}+\mu_1+\mu_2)t_{k+1}}e^{-A_{F_k}(t_k+t_{k+1})}. \tag{5.19}$$

We now integrate s_k from s_{k-1} to s_{k+1} in (5.19) and obtain

$$\sum_{\substack{p=p_1+p_2\\ |p_2|,|p_1|>R_{F_{k+1}}}}\sum_{k_1+\cdots+k_N=-p}vv\left(\sum\tilde{\omega}_{k_i}+\mu_1+\mu_2\right)^{-1} \times\left(1-e^{-(\sum\tilde{\omega}_{k_i}+\mu_1+\mu_2)(t_k+t_{k+1})}\right):\phi^N:_p:\phi_{k_1}\cdots\phi_{k_N}: \times e^{-A_{F_k}(t_k+t_{k+1})}. \tag{5.20}$$

The first term in parentheses in (5.20) against the term in (5.5) is the renormalization cancellation. The second term in the parentheses in (5.20) yields a term we call J_9.

Now finally we take the Boson annihilation operators in J_5 and J_6 [(5.15) and (5.16)] and push them to the right until they either hit $e^{-N_\tau B/2}$ or contract out. At this stage the renormalization cancellation is manifest, and all terms sandwiched between exponentials from the Duhamel expansion involve only boson ϕ's and fermion operators. However, as performed in the next section some of the expressions in boson ϕ's must be dismembered before expressions are obtained submissive to domination by the $\tilde{H}_{0B}$ + : ϕ^{2M} : in the exponents.

6. MORE BOSON PULL-THROUGHS

We begin with the contribution from the first term in parentheses in (5. 20), omitting the exponent at the right:

$$\sum_{\substack{p=p_1+p_2 \\ |p|,|p|>R_{F_{k+1}} \\ 2 \quad 1}} \sum_{k_1+\cdots k_N=-p} vv(\sum \tilde{\omega}_{k_i} + \mu_1 + \mu_2)^{-1} : \times \phi^N :_p : \phi_{k_1} \cdots \phi_{k_N} :. \tag{6.1}$$

We normal order this expression to obtain

$$\sum_{\substack{p=p_1+p_2 \\ |p_2|,|p_1|>R_{F_{k+1}}}} \sum_{k_1+\cdots+k_N=-p} vv(\sum \tilde{\omega}_{k_i} + \mu_1 + \mu_2)^{-1} \times \sum_{l=0}^{N} l!\binom{N}{l}\binom{N}{l} : \phi^{N-l}_{p-\sum_{i=1}^{l} k_i} \phi_{k_{l+1}} \cdots \phi_{k_N} : \times \prod_{m=1}^{l}\left(\frac{1}{2\pi}\frac{1}{2\omega_{k_m}}\right). \tag{6.2}$$

This is in form to cancel with the renormalization counterterms except that the energy denominators include all the boson energies instead of just the contracted boson energies. We write (6. 2) as the sum of two terms:

$$K = \sum_{\substack{p=p_1+p_2 \\ |p_2|,|p_1|>R_{F_{k+1}}}} \sum_{k_1+\cdots+k_N=-p} vv \sum_{l=0}^{N} (l!)\binom{N}{l}\binom{N}{l} \times : \phi^{N-l}_{p-\sum_{i=1}^{l}k_i} \phi_{k_{l+1}} \cdots \phi_{k_N} : \times \prod_{m=1}^{l}\left(\frac{1}{2\pi}\frac{1}{2\omega_{k_m}}\right)\cdot\left(\sum_{i=1}^{l}\tilde{\omega}_{k_i} + \mu_1 + \mu_2\right)^{-1} \tag{6.3}$$

and

$$MJ = \sum_{\substack{p=p_1+p_2 \\ |p_2|,|p_1|>R_{F_{k+1}}}} \sum_{k_1+\cdots+k_N=-p} vv \sum_{l=0}^{N-1} (l!)\binom{N}{l}\binom{N}{l} \times : \phi^{N-l}_{p-\sum_{i=1}^{l}k_i} \phi_{k_{l+1}} \cdots \phi_{k_N} : \prod_{m=1}^{l}\left(\frac{1}{2\pi}\frac{1}{2\omega_{k_m}}\right). \times \left[\left(\sum_{i=1}^{N}\tilde{\omega}_{k_i} + \mu_1 + \mu_2\right)^{-1} - \left(\sum_{i=1}^{l}\tilde{\omega}_{k_i} + \mu_1 + \mu_2\right)^{-1}\right]. \tag{6.4}$$

In MJ we decompose $\phi_{k_{l+1}}, \cdots, \phi_{k_N}$ into creation and annihilation operators, and then push the annihilation operators to the right until contraction or they hit $e^{-N_\tau B/2}$, and creation operators to the left until they contract or hit the left $e^{-N_\tau B/2}$. This pushing of creation operators to the left leads to more complicated estimates than of the annihilation operators to the right, since the number of terms to the left increases with order in the Duhamel expansion. Similarly $\phi_{k_1}, \cdots, \phi_{k_N}$ from the second term in parentheses in (5. 20), $\phi_{k_1}, \cdots, \phi_{k_s}$ in (5. 15), $\phi_{k_{l+1}}, \cdots, \phi_{k_s}$ in (5. 16), $\phi_{k_1}, \cdots, \phi_{k_s}$ in (5. 17), and $\phi_{k_{l+1}}, \cdots, \phi_{k_s}$ from J_8 are decomposed into creation and annihilation operators that are pushed to the left and right, respectively.

At the end of the "pull acrosses" above, preliminary algebra is now concluded and the stage is set to begin making estimates.

It is to be noted that it is possible to decompose still other ϕ's in the above manner and still control later estimates. In a sense we have decomposed a minimal number of ϕ's.

7. UNIT BLOCK ESTIMATES

A. Classification of blocks

Having performed a Duhamel expansion, pulled some fermion operators across a single exponent, pulled some boson operators across a single exponent, integrated over some of the time variables to reveal the renormalization cancellation, pulled some boson annihilation operators to the right, and finally pulled some more boson annihilation operators to the right and creation operators to the left, we are left with a sum of products of operators to be integrated over t variables. The operator products will be broken into products of unit blocks, and the operator norm of the product estimated as the product of the operator norms of the unit blocks, t-dependent operator norms. These product estimates will then be integrated over the t. The estimate for the sum of such products will be taken to be merely the sum of the norms for individual terms, by the triangle inequality; the sum includes sums over momenta of pulled-across boson operators. We proceed to consider the unit blocks.

Each product of unit blocks involves fermion operators, boson ϕ's, and boson kinetic energy terms in the exponentials. A matrix element of such a product can be realized as an integral over boson path space, with an expression over path space depending only on ϕ's; the boson kinetic energies removed from the exponents. This set-up is viewed as an expression in fermion operators "fibered" over path space. The operator norm is estimated over each fiber treating the ϕ as numerical quantities, the integral over path space then considered. This integral over path space is then replaced by an operator expression in boson ϕ operators and boson kinetic energies. A sample inequality derived by this method is the following:

$$\left| \sum_{k,j} \alpha_j \phi_j e^{-(H_{0B} + H_{0F} + s(\phi))t} \phi_j b^*_k b'^*_{-k} \beta_k \right| \leq \left(\sum_j |\alpha_j \phi_j^2|\right)^{1/2} e^{-(H_{0B})t - s(\phi)t} \left(\sum_j |\alpha_j \phi_j^2|\right)^{1/2} \times \left(\sum \frac{1}{\omega} |\beta_k|^2\right)^{1/2} \frac{1}{\sqrt{t}}. \tag{7.1}$$

It is possible that most if not all of these estimates are derivable without path space methods. Some of our unit blocks will be disconnected terms united by some sum, as the sum on j above, which can be estimated as a unit. The path space method will enable an expression like

$$\sum_j \alpha_j \phi_j J \phi_j \tag{7.2}$$

to be estimated by

$$(\sum \alpha_j |\phi_j|^2)^{1/2} |J| \left(\sum_j \alpha_j |\phi_j|^2\right)^{1/2}, \tag{7.3}$$

where we will think of an expression like

$$\sum_j (\alpha_j \phi_j) \sim (\phi_j) \tag{7.4}$$

as a (disconnected) unit block.

In fact, as stated before we imagine working with an upper cutoff to the fermion momenta, and obtaining estimates independent of this upper cutoff. With such an upper cutoff the fermion operators are finite matrices, and the fibering of fermion space over path space presents no analytic complexities.

Elements of the unit blocks will be subscripted by four possible subscripts T_1, T_2, T_3, and T_4.

T_1 assumes the value n if the term arises from either A_n or B_n in the Duhamel expansion.

T_2 contains two sets of momenta representing contractions with pulled-through creation and annihilation operators. Thus $(V)^{\langle T_2 \rangle}$ if $T_2 = \{(k_1, k_2), (k_3, k_4)\}$ is

$$(\sqrt{2\omega_{k_1}}\, a^*_{k_1}, \sqrt{2\omega_{k_2}}\, a^*_{k_2}, \sqrt{2\omega_{k_3}}\, a_{k_3}, \sqrt{2\omega_{k_4}}\, a_{k_4}, V), \tag{7.5}$$

where successive commas indicate successive commutators. We define $|T_2|$ to be the total number of momenta in the two subsets of T_2.

T_3 contains two sets of momenta representing creation and annihilation operators originating in the term, that have been pulled through. Thus

$$\sum_{i,j,l} c_{ijl}\, \phi_{k_i} a_{k_j} a_{k_l} = \sum_{i,j,l} \left(c_{ijl} \frac{1}{\sqrt{2\pi}} \sqrt{2\omega_{k_j}} \sqrt{2\omega_{k_j}} \right)_{T_3'} + \sum_{i,j,l} \left(c_{ijl} \frac{1}{\sqrt{2\pi}} \sqrt{2\omega_{k_j}} \sqrt{2\omega_{k_l}} \right)_{T_3''} \tag{7.6}$$

with $T_3' = \{(\ ), (k_i, k_j, k_l)\}$ and $T_3'' = \{(-k_i), (k_j, k_l)\}$ if ϕ_{k_i}, a_{k_j}, and a_{k_l} are all pulled across. As with T_2 we define $|T_3|$, and $|T_3|_a$ and $|T_3|_c$ in an obvious fashion, with $|T_3| = |T|_a + |T|_c$.

T_4 labels momenta of creation and annihilation operators that have reached the outside of the Duhamel expansion.

We proceed to enumerate the types of unit blocks.

(1) U_{1L,T_4} and U_{1R,T_4}. If $T_4 = \{k_1, \ldots, k_s\}$ then

$$U_{1R,T_4} = e^{-A_1 t} a_{k_1} \cdots a_{k_s} e^{-N_\tau B/2} \tag{7.7}$$

and similarly for U_{1L,T_4}.

(2) U_{2a,T_1,T_2} and U_{2c,T_1,T_2}.

$$U_{2a,T_1,T_2} = e^{-A_{T_1+1} t} (V_{pa} - V_{pa}^{(T_1)})^{\langle T_2 \rangle} e^{-A_{T_1} t'} \tag{7.8}$$

and similarly for U_{2c,T_1,T_2}.

(3) U_{3,T_1,T_2}.

$$U_{3,T_1,T_2} = e^{-A_{T_1+2} t} (\Delta^{(T_1)} - D^{(T_1)})^{\langle T_2 \rangle} e^{-A_{T_1} t'}. \tag{7.9}$$

(4) U_{4,T_1,T_2}, arising from J_1.

$$U_{4,T_1,T_2} = e^{-A_{T_1+1} t} (A_{T_1+1} - A_{T_1})^{\langle T_2 \rangle} e^{-A_{T_1} t'}. \tag{7.10}$$

(5) U_{5a,T_1,T_2} and U_{5c,T_1,T_2}, arising from J_2.

$$U_{5c,T_1,T_2} = e^{-A_{T_1} t} (V_{pc} - V_{pc}^{(T_1)})_{J_2}^{\langle T_2 \rangle} e^{-A_{T_1} t'}, \tag{7.11}$$

where the subscript J_2 indicates limits on the fermion momenta stated in the definition of J_2.

(6) U_{6a,T_1,T_2} and U_{6c,T_1,T_2}, arising from J_3.

$$U_{6c,T_1,T_2} = \sum_{\substack{p = p_1 + p_2 \\ |p_2|, |p_1| > R_{T_1+1}}} v e^{-A_{T_1+2} t_i} e^{-H_{0F}^{(-T_1)} T/2} \times e^{-[(\mu_1 + \mu_2)/2] T}\, b'^*_{p_1} b^*_{p_2}$$
$$\sim e^{-(A_{T_1} - H_{0F}^{(-T_1)}) t_k} \cdot (:\phi^N:_p)^{\langle T_2 \rangle} \tag{7.12}$$

with $T = s_{i-1} - s_{k-1}$. This is a disconnected unit block.

(7) $U_{7,T_1,T_2,T_2'}$, arising from J_4.

$$U_{7,T_1,T_2,T_2'} = \sum_{\substack{p_1 = p_2 + p_3 + p_4 \\ |p_1|, |p_4|, |p_3 - p_1| > R_{T_1+1}}} vve^{-A_{T_1+2} t_i} b^*_{p_1}$$
$$\sim (:\phi^N:_{p_2})^{\langle T_2 \rangle} e^{-A_{T_1} t_j}$$
$$\sim e^{-A_{T_1} t_k} (:\phi^N:_{p_3})^{\langle T_2' \rangle} \sim b_{p_4} e^{-A_{T_1} t_l} e^{-\mu_1 T - \mu_4 \tilde{T} - \mu_c \tilde{\tilde{T}}} \tag{7.13}$$

with $\mu_c = \mu_{p_3 - p_1}$, $T = s_{i-1} - s_{k-1}$, $\tilde{T} = s_j - s_l$, $\tilde{\tilde{T}} = s_j - s_{k-1}$.

(8) U_{8,T_1,T_2,T_3} arising from J_5.

If $T_3 = \{(\ ), (k_1, \ldots, k_s, k_N)\}$, for example,

$$U_{8,T_1,T_2,T_2',T_3} = \sum_{\substack{p_1 + p_2 = p \\ k_{s+1} + \cdots + k_{N-1} = -k_N - k_1 \cdots - k_s - p \\ |p_1|, |p_2| > R_{T_1+1}}} \times vve^{-A_{T_1+2} t_{i_1}/2} (:\phi^N:_p)^{\langle T_2 \rangle}$$
$$\sim e^{-A_{T_1} t_{i_2}/2} (:\phi_{k_{s+1}} \cdots \phi_{k_{N-1}}:)^{\langle T_2' \rangle} \left(\frac{1}{2\pi}\right)^{(s+1)/2}$$
$$\times (1 - e^{-2\tilde{\omega}_{k_N} T}) \cdot e^{-(\sum_{i=1}^{s} \tilde{\omega}_{k_i} + \mu_1 + \mu_2) T}, \tag{7.14}$$

where

$$T = s_{i_1 - 1} - s_{i_2 - 1}. \tag{7.15}$$

(9)–(15) arise from J_6, J_7, J_8, J_9, MJ, (6.3) $- (\Delta - \Delta^{(i)})$, and contractions with the exponent, respectively.

(16) U_{16,T_2} arising from V_s.

$$U_{16,T_2} = e^{-A_{T_1+1} t} (V_s)^{\langle T_2 \rangle} e^{-A_{T_1} t'} \tag{7.16}$$

the only term depending on V_s.

B. Numerical constants and the generic estimate

The numerical estimates of our calculation involve several constants: $\alpha, \beta, \gamma, \epsilon, \delta, \tau$, and $\tau!$. β is required to satisfy

$$N/2M < \beta < \tfrac{1}{2} \tag{7.17}$$

and

$$\tfrac{1}{2} - 1/4M < \beta. \tag{7.18}$$

ϵ must satisfy

$$(2N+4)\epsilon < \tfrac{1}{2} - \beta. \tag{7.19}$$

We pick

$$\tau' = 1 \tag{7.20}$$

and choose γ and τ satisfying

$$\gamma = \tfrac{1}{2} + \tfrac{1}{2}\tau + \epsilon > 1. \tag{7.21}$$

δ is required to satisfy

$$\delta < \epsilon/2\alpha. \tag{7.22}$$

$\beta, \gamma, \epsilon, \tau,$ and τ' can be picked satisfying (7.17)–(7.21) and and fixed; α will have to be picked suitably large later and δ then picked satisfying (7.22).

We estimate unit block (1) first, a unique type block; referring to (7.7),

$$|U_{1R,T_4}| \le c2^{s/2} \prod_{i=1}^{s} (\omega_{k_i})^{1/2} (s!)^{1/2}. \tag{7.23}$$

We shall use c to denote uninteresting constants.

The desired estimates for all the remaining blocks assume the following general form [if $T_3 = \{(k_1, \cdots, k_s)(k_{s+1}, \cdots, k_r)\}$]:

$$\sum_{k_1, \cdots, k_r} \left(\prod_{i=1}^{r} \frac{1}{\tilde{\omega}_{k_i}} \right)^{\gamma} |U_{T_1,T_2,T_3}| \le c(T_1)^{\epsilon} (T_1^{|T_3|\alpha c_1})^{-1} \frac{1}{t^a} \frac{1}{t'^b} \tag{7.24}$$

with c_1 and c constants, $|T_3| = r$. Here we have considered for simplicity a block depending on two t's; for a block such as (7) depending on four t's an analogous formula holds. a and b satisfy

$$a + b \le 1 - (2N+3)\epsilon. \tag{7.25}$$

The blocks are chosen so that the product of neighboring estimates involving the same t yields a total exponent

$$b + a' \le 1 - \delta. \tag{7.26}$$

[$|U_A| \le \phi(1/t^a)(1/t'^b)$, $|U_B| \le \phi'(1/t'^{a'})(1/t''^{b'})$, U_A and U_B successive blocks.] Equations (7.24), (7.25), and (7.26) are central, organizing in spirit the entire set of estimates in this section, and enabling the overall estimates of the next section. We now discuss the verification of (7.24) for a selected number of the unit blocks, introducing methods applicable to all the cases.

C. Estimates for U_2 and U_{16} in fermion variables

The calculations of this subsection apply virtually unchanged to unit blocks (5) and (6) and parts of (4) and (15). In basic essentials they apply to all the blocks except U_1. For simplicity we take T_2 to be the empty set. We first consider U_2. It is sufficient to look at the expression

$$e^{-A_{T_1+1}t} V_{pc} e^{-A_{T_1}t'} = e^{-A_{T_1+1}t} \left(\sum_{i+j+k=0} :\phi^N:_i v b'^*_{-j} b^*_{-k} \right) \cdot e^{-A_{T_1}t'}. \tag{7.27}$$

We break the sum into four parts:

$$\begin{aligned} &(a) \quad |j| < |k|, \quad |k| > (T_1+1)^\alpha, \\ &(b) \quad |j| < |k|, \quad |k| \le (T_1+1)^\alpha, \\ &(c) \quad |j| \ge |k|, \quad |j| > (T_1+1)^\alpha, \\ &(d) \quad |j| \ge |k|, \quad |j| \le (T_1+1)^\alpha. \end{aligned} \tag{7.28}$$

It is clearly sufficient to consider (a) and (b) only. Case (a) is the most interesting. We define

$$\hat{\mu}_i = \begin{cases} \mu_i & \text{if } |i| > (T_1+1)^\alpha \\ 0 & \text{if } |i| \le (T_1+1)^\alpha \end{cases} \tag{7.29}$$

and study the contribution of (a) to (7.27).

$$\begin{aligned} \mathrm{II}_a &= e^{-A_{T_1+1}t} \sum_{\substack{i+j+k=0 \\ |k|>|j| \\ |k|>(T_1+1)^\alpha}} :\phi^N:_i v b'^*_{-j} b^*_{-k} e^{-A_{T_1}t'} \\ &= \exp[-(A_{T_1+1} - \tfrac{1}{2} H_{0F}^{(-(T_1+1))})t] \sum \\ &\quad \times :\phi^N:_i v b'^*_{-j} b^*_k e^{(\hat{\mu}_j + \hat{\mu}_k)t/2} \\ &\quad \times \exp[-\tfrac{1}{2} H_{0F}^{(-(T_1+1))} t] e^{-A_{T_1}t'}. \end{aligned} \tag{7.30}$$

We have used the fact that $H_{0F}^{(-(T_1+1))}$ commutes with A_{T_1+1}; the fact that H_{0F} does not commute with A_{T_1+1} is what forces us to consider (a) and (b) separately.

We view II_a as an operator in fermion variables with numerical valued boson operators merely functions in path space (removing $\tilde{H}_{0B}$) and take the norm as a fermion operator, using an N_τ estimate:

$$\begin{aligned} |\mathrm{II}_a|_{\mathrm{p.s.}} &\le c \exp[-\textstyle\int(\int :\phi^{2M}:)dt] \left\{ \frac{1}{t^{1/2}} \sum |v|^2 \frac{1}{\hat{\mu}_j + \hat{\mu}_k} \right. \\ &\quad \times e^{-(\hat{\mu}_j+\hat{\mu}_k)t} :\phi^N:_i :\phi^N:_{-i} \Big\}^{1/2} \exp[-\textstyle\int(\int :\phi^{2M}:)dt'] \end{aligned} \tag{7.31}$$

$$\begin{aligned} &\le c \frac{1}{t^{1/2}} (\ln t)^{1/2} \exp[-\textstyle\int(\int :\phi^{2M}:)dt] \\ &\quad \times \left\{ \sum_i e^{-1/2|i|t} :\phi^N:_i :\phi^N:_{-i} \right\}^{1/2} \\ &\quad \times \exp[-\textstyle\int(\int :\phi^{2M}:)dt']. \end{aligned} \tag{7.32}$$

This is a positive function on path space.

The N_τ estimate is as follows:

$$\begin{aligned} |e^{-At}J| &= \left| e^{-At}\sqrt{A} \cdot \frac{1}{\sqrt{A}} \cdot \sqrt{H_0} \cdot \frac{1}{\sqrt{H_0}} \cdot J \right| \\ &\le |e^{-At}\sqrt{A}| \cdot \left| \frac{1}{\sqrt{A}} \cdot \sqrt{H_0} \right| \cdot \left| \frac{1}{\sqrt{H_0}} \cdot J \right| \\ &\le \frac{1}{\sqrt{t}} \cdot 1 \cdot \left| \frac{1}{\sqrt{H_0}} \cdot J \right| \end{aligned}$$

provided $A \ge H_0$. In this derivation the fact that the fermion variables in the exponential are "time ordered" has been suppressed. See the Appendix for a treatment of this point.

Turning to II_b,

$$\mathrm{II}_b = e^{-A_{T_1+1}t} \sum_{\substack{i+j+k=0 \\ |k|>|j| \\ |k|\le(T_1+1)^\alpha}} :\phi^N:_i v b'^*_{-j} b^*_{-k} e^{-A_{T_1}t'}. \tag{7.33}$$

In this case we directly use an N_τ estimate:

$$|\mathrm{II}_b|_{\mathrm{p.s.}} \le c\,\frac{1}{t^{1/2}}\,[\ln(T_1+1)]^{1/2}\exp[-\textstyle\int(\int:\phi^2:)dt]$$
$$\times\Big\{\sum_{|i|\le 2(T_1+1)^\alpha}:\phi^N:_i:\phi^N:_{-i}\Big\}^{1/2}$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt']. \qquad (7.34)$$

Before performing the boson estimates we turn to U_{16}.

We consider

$$e^{-A_{T_1+1}t}V_s e^{-A_{T_1+1}t'} = e^{-A_{T_1+1}t}\sum:\phi^N:_i b^*_{-j}b_k v e^{-A_{T_1}t'}. \qquad (7.35)$$

(The antiparticle terms are of course identical.) We break the sum over momenta into a number of regions, where it is sufficient to consider the following cases:

$$\begin{aligned}&\text{(a)}\quad |j|\ge|k|,\quad |j|\le(T_1+1)^\alpha,\\ &\text{(b)}\quad |j|\ge|k|,\quad |k|\le(T_1+1)^\alpha<|j|,\qquad (7.36)\\ &\text{(c)}\quad |j|\ge|k|,\quad |k|>(T_1+1)^\alpha.\end{aligned}$$

We note the operator inequality

$$N^{\alpha_1+\alpha_2}_{[(\alpha_1\tau_1+\alpha_2\tau_2)/(\alpha_1+\alpha_2)]}\le N^{\alpha_1}_{\tau_1}N^{\alpha_2}_{\tau_2}. \qquad (7.37)$$

In case (a) we proceed as follows:

$$|e^{-At}V^a_s e^{-At'}|_{\mathrm{p.s.}}$$
$$\le\frac{c}{t^{1/2-\delta}t'^{\delta}}\exp[-\textstyle\int(\int:\phi^{2M}:)dt]\Big|\frac{1}{H_{0F}^{1/2-\delta}}V^a_s\frac{1}{N_F^\delta}\Big|$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt']$$
$$\le\frac{c}{t^{1/2-\delta}t'^{\delta}}\exp[-\textstyle\int(\int:\phi^{2M}:)dt]\Big|\frac{1}{N_{F\rho}^{1/2}}V^a_s\Big|$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt'], \qquad (7.38)$$

with $\rho=1-2\delta$. So

$$|\mathrm{XVI}_a|_{\mathrm{p.s.}}\le\frac{c}{t^{1/2-\delta}t'^{\delta}}\exp[-\textstyle\int(\int:\phi^{2M}:)dt]$$
$$\times\Big\{\sum_{|i|\le 2(T_1+1)^\alpha}:\phi^N:_i:\phi^N:_{-i}\Big\}^{1/2}$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt']\cdot T_1^{\alpha 2\delta}. \qquad (7.39)$$

In case (b) we proceed similarly but pull $\exp[-H_{0F}^{(-(T_1+1))}t]$ across V to attach $e^{-\mu_j t}$ to the kernel to obtain

$$|\mathrm{XVI}_b|_{\mathrm{p.s.}}\le\frac{c}{t^{1/2-\delta}t'^{\delta}}\exp[-\textstyle\int(\int:\phi^{2M}:)dt]$$
$$\times\Big\{\sum_i:\phi^N:_i:\phi^N:_{-i}e^{-|i|t/4}\Big\}^{1/2}$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt']T_1^{2\alpha\delta}. \qquad (7.40)$$

In case (c) we also pull $\exp(-H_{0F}^{(-(T_1+1))}t)$ across V this time attaching $e^{-(\mu_j-\mu_i)t}$ to the kernel. A simple estimate of the kernal gives

$$|\mathrm{XVI}_c|_{\mathrm{p.s.}}\le\frac{c}{t^{1/4}t'^{1/4}}\exp[-\textstyle\int(\int:\phi^{2M}:)dt]$$
$$\times\Big\{\sum_i:\phi^N:_i:\phi^N:_{-i}\Big(\frac{1}{|i|+1}+e^{-1/4|i|t}\Big)\Big\}^{1/2}$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt']. \qquad (7.41)$$

Here the N_τ estimate used attaches $1/\sqrt{H_{0F}}$ on the right side of V where H_{0F} is multiplied by $(t+t')$ in the exponent.

D. Estimates in boson variables

Looking at (7.32), (7.34), (7.39), (7.40), and (7.41) we see that we are left with the task of performing an estimate in path space for an expression of the form

$$\exp[-\textstyle\int(\int:\phi^{2M}:)dt]\Big\{\sum_i\alpha_i:\phi^N:_i:\phi^N:_{-i}\Big\}^{1/2}$$
$$\times\exp[-\textstyle\int(\int:\phi^{2M}:)dt']. \qquad (7.42)$$

At this stage we trade in path space for Foch space and consider

$$\exp[-(\tilde H_{0B}+\textstyle\int:\phi^{2M}:)t]\Big\{\sum_i\alpha_i:\phi^N:_i:\phi^N_{-i}:\Big\}^{1/2}$$
$$\times\exp[-(\tilde H_{0B}+\textstyle\int:\phi^{2M}:)t']. \qquad (7.43)$$

We pick d such that

$$\tilde H_{0B}+\textstyle\int:\phi^{2M}:+d\ge 1 \qquad (7.44)$$

and normal order the expression:

$$\sum_i\alpha_i:\phi^N:_i:\phi^N:_{-i}=\sum_{i,s\le N}\beta_{i,s}:\phi_i^s\phi_{-i}^s:. \qquad (7.45)$$

By a trivial argument it is sufficient to consider expressions like

$$G=\exp[-(\tilde H_{0B}+\textstyle\int:\phi^{2M}:)t]\Big\{\Big|\sum_i\beta_i:\phi_i^s\phi_{-i}^s:\Big|\Big\}^{1/2}$$
$$\times\exp[-(\tilde H_{0B}+\textstyle\int:\phi^{2M}:)t'] \qquad (7.46)$$

with $1\le s\le N$. We assume that as with II_a we want the boson expression to be dominated entirely by the right side exponent. Equation (7.26) dictates in each case what ratio of each exponent is to be used. We seek the following estimate

$$|G|\le\Big(\sup_i|\beta_i|/t'^{\beta}\Big)c. \qquad (7.47)$$

We define $\bar\beta=\sup_i|\beta_i|$ and observe that (7.47) would be implied by

$$\Big\|\Big\{\Big|\sum_i\beta_i:\phi_i^s\phi_{-i}^s:\Big|\Big\}^{1/2}[(\tilde H_{0B}+\textstyle\int:\phi^{2M}:+d)^\beta]^{-1}\Big\|\le c\bar\beta, \qquad (7.48)$$

in turn implied by

$$[(\tilde H_{0B}+\textstyle\int:\phi^{2M}:+d)^\beta]^{-1}\Big\{\Big|\sum_i\gamma_i:\phi_i^s\phi_{-i}^s:\Big|\Big\}$$
$$\times[(\tilde H_{0B}+\textstyle\int:\phi^{2M}:+d)^\beta]^{-1}\le c \qquad (7.49)$$

if $|\gamma_i|\le 1$. Putting $\gamma_i=1$ does not essentially change any of the succeeding estimates. We now use (3.4) to see that (7.49) is implied by

$$\tilde H_{0B}+\textstyle\int:\phi^{2M}:+d-c\Big\{\Big|\sum_i:\phi_i^s\phi_{-i}^s:\Big|\Big\}^{1/2\beta}\ge 0 \qquad (7.50)$$

(for c small enough).

This last inequality we show by E. Nelson's method of Ref. 3. To an upper cutoff L we associate

$$V_L=\textstyle\int:\phi_L^{2M}:-c\{|\int:\phi_L^{2s}:|\}^{1/2\beta}, \qquad (7.51)$$

$$R_L=\textstyle\int:\phi^{2M}:-c\{|\int:\phi^{2s}:|\}^{1/2\beta}-V_L. \qquad (7.52)$$

It is then required to find a suitable estimate for a constant c_L with

$$V_L \geq C_L \tag{7.53}$$

and an estimate for

$$\langle 0 | (R_L)^n | 0 \rangle . \tag{7.54}$$

We unwick:

$$: \phi_L^{2s} : = \sum_{r=0}^{s} \omega_r \phi^{2r}, \tag{7.55}$$

$$| \int : \phi^{2s} : | = \left| \int \left(\sum_{r=0}^{s} \omega_r \phi^{2r} \right) \right| \leq \sum_{r=0}^{s} |\omega_r| \int \phi^{2r}, \tag{7.56}$$

$$\begin{aligned} | \int : \phi^{2s} : |^{1/2\beta} &\leq \left(\sum_{r=0}^{s} |\omega_r| \int \phi^{2r} \right)^{1/2\beta} \\ &\leq (s+1)^{1/2\beta} \sum_{r=0}^{s} |\omega_r|^{1/2\beta} (\textstyle\int \phi^{2r})^{1/2\beta} \\ &\leq c \sum_{r=0}^{s} |\omega_r|^{1/2\beta} \int \phi^{r/\beta}. \end{aligned} \tag{7.57}$$

(Yes, ω_r depends on L.) r/β is less than $2M$, the crucial fact.

To study R_L we look at

$$\Delta_L = | \int : \phi_L^{2s} : |^{1/2\beta} - | \int : \phi^{2s} : |^{1/2\beta}. \tag{7.58}$$

We use the inequality for $x, y \geq 0,\ 1 < \alpha < 2$

$$\begin{aligned} |x^\alpha - y^\alpha| &\leq |x - y| (\alpha x^{\alpha-1} + \alpha y^{\alpha-1}) \\ &\leq c |x - y| (1 + x^2 + y^2) \end{aligned} \tag{7.59}$$

to get

$$|\Delta_L| \leq c | \textstyle\int : \phi_L^{2s} : - \int : \phi^{2s} : | (1 + (\int : \phi_L^{2s} :)^2 + (\int : \phi^{2s} :)^2). \tag{7.60}$$

From (7.57) and (7.60), (7.50) follows by direct calculation in imitation of Ref. 3.

8. OVERALL NUMERICAL ESTIMATES

The "phase space" volume integral we use is

$$I_n = \int_0^1 dt_1 \cdots \int_0^1 dt_n \delta(\textstyle\sum t_i - 1) \frac{1}{t_1^{1-\alpha_1} \cdots t_n^{1-\alpha_n}} = \frac{\pi \Gamma(\alpha_i)}{\Gamma(\sum \alpha_i)}. \tag{8.1}$$

By (7.26) the $\Gamma(\alpha_i)$ in the numerator will satisfy

$$\Gamma(\alpha_i) \leq \Gamma(\delta). \tag{8.2}$$

Each pulled-across creation or annihilation operator borrows $1/\tilde{\omega}_i^\epsilon$ from the first exponent it can by

$$\tilde{\omega}_i^\epsilon e^{-\tilde{\omega}_i t} \leq C/t^\epsilon. \tag{8.3}$$

Equation (7.25) allows for $2N$ borrowings of t^ϵ by N operators passing to the left and N passing to the right. Equation (7.21), the estimate for unit block (1), and the $1/\sqrt{\omega}$ associated to each, a, a^* in ϕ, assure each pulled-across operator finally contracted or reaching $e^{-N_T \beta/2}$ has associated to it a factor $1/\tilde{\omega}_i^\delta$. The integral in (8.1) is then estimated by

$$I_n \leq c^n/(n!)^{3\epsilon}. \tag{8.4}$$

In (7.24) we cancel $(T_1)^\epsilon$ by a power, leaving an estimate

$$\sum_{k_1, \cdots, k_r} \left(\prod_{i=1}^{r} \frac{1}{\tilde{\omega}_{k_i}} \right)^\gamma | U_{T_1, T_2, T_3} | \leq c \frac{1}{T_1^{|T_3| \alpha c_1}} \frac{1}{t^a} \frac{1}{t'^b}, \tag{8.5}$$

$$b + a' \leq 1 - \delta, \tag{8.6}$$

$$a + b \leq 1 - 2\epsilon \tag{8.7}$$

with each pulled-across operator having $1/\tilde{\omega}_i^\gamma$ associated to it. We pick $c \geq 1$.

We rewrite the product

$$\frac{1}{(n!)^{2\epsilon}} e_n \cdot e_{n-1} \cdots e_1 \tag{8.8}$$

as

$$\frac{1}{(n!)^\epsilon} f_n f_{n-1} \cdots f_1 \tag{8.9}$$

with

$$f_s = e_s \frac{1}{s^{\epsilon/s}} \frac{1}{(s+1)^{\epsilon/s+1}} \cdots \frac{1}{n^{\epsilon/n}}. \tag{8.10}$$

Note that this clever definition of f_s depends on n.

We proceed to the overall estimate; associating terms in B_i's second order in g with fermion contracted loops as we have. We have

$$|e^{-H}| \leq |e^{-N_T \beta/2} e^{-\bar{H}} e^{-N_T \beta/2}| \leq \sum_{n=0}^{\infty} E_n, \tag{8.11}$$

where nonexponentiated terms in E_n (before pulling across operations) are of order g^n. We write

$$E_n \leq G_n \cdot M_n \cdot \frac{1}{(n/2!)^\epsilon}, \tag{8.12}$$

where G_n is the number of types of decompositions into unit blocks, before the onset of boson annihilation operator pull-throughs or "more boson pull-throughs" (Sec. 6), grouping together terms in a given unit block having the same values of $|T_3|_a$ and $|T_3|_c$. G_n is easily overestimated by

$$G_n \leq 3^n \cdot 4^n \cdot (16 N^3)^n \tag{8.13}$$

(3 corresponding to V_s, V_a, or V_c; 4 corresponding to not pulling across fermions, contracting to a loop, a single contraction, or pulled-across fermions with no contractions; $16N$ overestimates types of contributions from closed fermion loops; N^2 overestimates values of $|T_3|_a$ and $|T_3|_c$). $(n/2!)^\epsilon$ in (8.12) comes from integrating (8.5) by (8.1), using decomposition (8.8)–(8.10), and recalling there are terms in which as many as $\frac{1}{2}$ the original t variables are integrated out to exhibit the renormalization cancellation.

Looking at the terms in E_n where

$$\begin{aligned} |T_3|_a &= r_i \leq N, \\ |T_3|_c &= l_i \leq N \end{aligned} \tag{8.14}$$

in the ith unit block, $i = 1, \cdots, n$. M_n is estimated using (8.1), (8.5), and (8.8)–(8.10) by

$$M_n \leq \sup_{\{r_i, l_i\}} (C^{N+1})^n \prod_{i=1}^{n} (f_i (3N_i i)^{r_i} (3iN(n-i))^{l_i}), \tag{8.15}$$

$$f_i = \frac{1}{i^{(r_i + l_i)\alpha c_1}} \cdot \frac{1}{i^{\epsilon/i}} \cdot \frac{1}{(i+1)^{\epsilon/i+1}} \cdots \frac{1}{n^{\epsilon/n}}. \tag{8.16}$$

The terms $(3Ni)^{r_i}(3N(n-i))^{l_i}$ generously count the number of terms generated by boson pull-throughs, or include the $(s!)^{1/2}$ from (7.23). The contracted over boson terms have been summed over, as have boson operators reaching $e^{-N_\tau B/2}$. In (8.16) split up (8.9)–(8.10) has been used, applied to unit block terms present before boson annihilation pull-throughs or "more boson pull-throughs." We get

$$M_n \leq \sup_{\{r_i, l_i\}} (c')^n \prod_i (i^{r_i(1-ac_1)} \cdot i^{l_i(1-\alpha c_1)} \cdot h^{l_i}(i)), \tag{8.17}$$

$$h(i) = \sup_{x>i} \frac{(x-i)}{i^{\epsilon/iN} \cdot (i+1)^{\epsilon/(i+1)N} \cdots x^{\epsilon/xN}} \tag{8.18}$$

$$= \sup_{x>i} \exp[\ln(x-i) - (\epsilon/N) \sum_{j=1}^{x} (1/j) \ln j]. \tag{8.19}$$

Estimating the sum in (8.19) by an integral, we have

$$h(i) \cong \sup_{x>i} \exp[\ln(x-i) - (\epsilon/N)((\ln x)^2 - (\ln(i))^2)] \tag{8.20}$$

$$\lesssim \sup_{x>i} \exp[\ln(x-i) - (2\epsilon/N)(\ln i)(\ln x - \ln i)]. \tag{8.21}$$

Differentiating to obtain the maximum:

$$\begin{aligned} &\frac{1}{x-i} - \frac{\epsilon}{N} \ln i \frac{1}{x} = 0, \\ &x - i = \left(\frac{N}{\epsilon} \frac{\ln i}{x}\right). \end{aligned} \tag{8.22}$$

Thus $h(i) \leq C''$ for some constant C''. Picking α to satisfy

$$\alpha \gtrsim 1/c_1, \tag{8.23}$$

we get

$$|e^{-H}| \leq \sum_{n=0}^{\infty} (\tilde{C})^n \frac{1}{(n/2!)^\epsilon}. \tag{8.24}$$

APPENDIX: TIME ORDERING IN FERMION VARIABLES

We consider an exponential of the type

$$e^{-t(H_{0F} + H_{0B} + V + V_+ + V_-)}, \tag{A1}$$

where (with modification of previous notation) V is a polynomial in φ variables, V_+ and V_- are terms in φ variables, and fermion creation and annihilation terms, respectively. The fermion momenta in V_+ and V_- are cutoff at i^α. Using the Trotter product formula equates (A1) with

$$\lim_{n\to\infty} (e^{-t/n H_{0B}} e^{-t/n (H_{0F} + V + V_+ + V_-)})^n. \tag{A2}$$

Passage to boson path space thus yields

$$F = \lim_{n\to\infty} T \prod_{j=1}^{n} \times \left\{ \exp\left[-t/n\left(H_{0F} + V\left(\frac{jt}{n}\right) + V_+\left(\frac{jt}{n}\right) + V_-\left(\frac{jt}{n}\right)\right)\right]\right\} \tag{A3}$$

where the arguments of the V are the time variables in the φ, and the T indicates that the product is written with j increasing to the left (time ordered). Since $H_{0F}^{(-i)}$ commutes with V_+ and V_-, we have

$$F = e^{-sH_{0F}^{(-i)}} G e^{-(t-s)H_{0F}^{(-i)}} \tag{A4}$$

for any $0 \leq s \leq t$, and

$$|G|_{p.s.} \leq 1 \exp\left[-\int_0^t dt P(\varphi(t))\right] \tag{A5}$$

provided $H_{0F}^{(i)}/2 + V + V_+ + V_- \geq P(\varphi)$ as an operator in fermion variables, as we assume.

We consider

$$G = \lim_{n\to\infty} T \prod_{j=1}^{n} \left\{ \exp\left[-t/n\left(V\left(\frac{jt}{n}\right) + V_+\left(\frac{jt}{n}\right) + V_-\left(\frac{jt}{n}\right) + H_{0F}^{(i)}\right)\right]\right\}. \tag{A6}$$

If the V_+ and V_- all commuted so (A6) could be written as

$$\exp\left[-\int_0^t dt (H_{0F}^{(i)} + V(t) + V_+(t) + V_-(t))\right], \tag{A7}$$

then it would follow

$$|(H_{0F}^{(i)})^\alpha G|_{p.s.} \leq (c/t^\alpha) \exp\left[-\int_0^t dt P(t)\right] \tag{A8}$$

if $0 \leq \alpha \leq \frac{1}{2}$, as we have used in the paper in the argument preceding (7.33). Relation (A5) is correct but presumably (A8) is not, so we proceed to replace it by an alternate estimate.

We write $V = V_1 + V_2$ where V_1 is quadratic in g and V_2 is independent of g. We also write $W = V_+ + V_- + V_1$. G is expandable in a Duhamel expansion

$$\begin{aligned} G = \sum_{n=0}^{\infty} (-1)^n \int_0^t ds_n \cdots \int_0^{s_3} ds_2 \int_0^{s_2} ds_1 \\ \times \exp\left[-\int_{s_n}^t dt (H_{0F}^{(i)} + V_2(t))\right] W(s_n) \\ \cdots W(s_1) \exp\left[-\int_0^{s_1} dt (H_{0F}^{(i)} + V_2(t))\right]. \end{aligned} \tag{A9}$$

Using estimates of the type of Sec. 7 an estimate for (A9) may be obtained replacing (A5) and (A8) with estimates adequate for the present calculation, and showing $e^{-N_\tau B/2} e^{-\bar{H}} e^{-N_\tau B/2}$ is in fact analytic in g. Thus the calculation could have been performed from the beginning with all terms in g in the B, the unexponentiated terms in the Duhamel expansion. This would have made the calculation somewhat simpler; but using cruder estimates, less useful possibly for later calculations.

We intend to use a better estimate for G than that obtained from (A9). In fact, what we require in addition to (A5) is a replacement for the doubtful inequality (A8). In virtue of (A4) the present calculation requires only an estimate for this, $|(H_{0F}^{(i)})^\alpha G|_{p.s.}$. We use the Duhamel expansion:

$$\begin{aligned} (H_{0F}^{(i)})^\alpha G = (H_{0F}^{(i)})^\alpha \sum_{n=0}^{\infty} (-1)^n \int_0^t ds_n \\ \cdots \int_0^{s_2} ds_1 \left[\exp - \int_{s_n}^t dt (H_{0F}^{(i)} + V_2(t))\right] V_-(s_n) \\ \cdots V_-(s_1)\left(\exp\left[-\int_0^{s_1} dt (H_{0F}^{(i)} + V_2(t))\right]\right. \\ - \int_0^{s_1} ds_0 \exp\left[-\int_{s_0}^{s_1} dt (H_{0F}^{(i)} + V_2(t))\right] V_+(S_0) \\ \left. \times \exp\left[-\int_0^{s_0} dt (H_{0F}^{(i)} + V(t) + V_+(t) + V_-(t))\right]\right). \end{aligned} \tag{A10}$$

It is easiest to find an estimate for

$$|(H_{0F}^{(i)})^{\alpha}(H_{0B}+V_2)^{\beta}e^{-(H_{0B}+H_{0F}^{(i)}+V+V_{+}+V_{-})t}|_{\text{v.p.s.}}=E_{\alpha,\beta}, \tag{A11}$$

where $\alpha+\beta<\frac{1}{2}$ and the v.p.s. norm is the total operator norm obtained by first passing to path space, taking the fermion norm, and then calculating the boson norm. This norm may be larger than the operator norm of

$$(H_{0F}^{(i)})^{\alpha}(H_{0B}+V_2)^{\beta}e^{-(H_{0B}+H_{0F}^{(i)}+V+V_{+}+V_{-})t}$$

since we have integrated over fermion norms rather than integrating and then taking norms. Using the methods of Sec. 7 and (A10) we find

$$E_{\alpha,\beta}\le\frac{1}{t^{\alpha+\beta}}(c_1+c_2 i^{\delta})e^{c_3(\ln i)}c_{4_t}c_5 \tag{A12}$$

with δ fixed but arbitrarily small. Replacing (A8) by (A12) does not substantially modify the present calculation. It is an interesting question as to whether there is a better estimate for (A11) than (A12), such as would be implied by (A8).

*This work was supported in part by NSF Grant GP-17523.

[1]J. Glimm and A. Jaffe, "Positivity and Self-Adjointness of the $P(\phi)_2$ Hamiltonian," to be published.
[2]J. Glimm and A. Jaffe, Fortsch. Phys. 21, 327 (1973).
[3]E. Nelson, "A. Quartic Interaction in Two Dimensions," in *Mathematical Theory of Elementary Particles*, edited by R. Goodman and I. Segal (M.I.T. Press, Cambridge, 1966).
[4]J. Glimm, Commun. Math. Phys. 5, 343 (1967); Commun. Math. Phys. 6, 120 (1967).
[5]J. Glimm and A. Jaffe, J. Math. Phys. 3, 1568 (1972).

Motion and expansion of wavepackets in periodic potentials

Y. Weissman and J. A. Sussmann*

Soreq Nuclear Research Centre, Yavne, Israel

Department of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel
(Received 14 May 1973)

The motion and expansion of single band wavepackets in a three-dimensioanl periodic potential are studied. Explicit expressions are derived for the expectation values $\langle \mathbf{x}(t) \rangle$ and $\langle \mathbf{x}^2(t) \rangle$ of an arbitrary combination of time-dependent Bloch functions belonging to a given nondegenerate band. Special attention is paid to spatial expansion of a Wannier function. It is shown that in this particular case $\langle \mathbf{x}^2(t) \rangle = \langle \mathbf{x}^2(0) \rangle + \langle \mathbf{v}^2 \rangle t^2$, where $\langle \mathbf{v}^2 \rangle$ is the average value of $[\nabla_k \omega(\mathbf{k})]^2$. This apparent generalization of the well known analogy between $\nabla_k \omega(\mathbf{k})$ and velocity does not hold in the more general case, where we get an additional term which is linear in time. The expressions derived for $\langle \mathbf{x}^2(t) \rangle$ are an important element in the calculation of the diffusion coefficient when the intermediate state of the jumping process is a wavepacket.

1. INTRODUCTION

The propagation of wavepackets in the presence of potential barriers has been studied extensively. This behavior has been used to exemplify quantum effects, such as reflection and tunneling. The propagation of a Gaussian wave packet in the presence of a parabolic potential barrier was recently studied in great detail by J. H. Weiner and Y. Partom,[1] in an effort to incorporate the above-mentioned quantum effects in chemical and other rate theories. However, the application of their results to diffusion of interstitials in solids is limited by the fact that only one single potential barrier is considered, while the diffusing particle is subject to a periodic potential, whenever the accommodation of the lattice to the migrating particle is negligible. A simple one-dimensional calculation shows[2] that under certain conditions jumps to nonnearest neighbor sites make a contribution which, regarding the diffusion process, cannot be neglected. Under these circumstances the periodicity of the potential must be taken into account.

The aim of the present paper is to study the features of wavepacket propagation which are relevant to diffusion in a general, three-dimensional periodic potential. The motion of the center of such wavepackets is usually described with the help of classicallike equations of motion with constant velocity. On the basis of the fact that the expectation value of the operator $\dot{\mathbf{x}}$ in a pure Bloch state with quasimomentum $\mathbf{k}_0$ is proportional to the $\mathbf{k}$-space energy gradient $\nabla_k \omega(\mathbf{k})$ evaluated at $\mathbf{k}_0$, one concludes that $\nabla_k \omega(\mathbf{k})$ plays the role of the classical velocity.[3] This interpretation is also supported by approximate calculations with wavepackets.[4] In this work we present an alternative derivation of the equivalence between velocity and $\nabla_k \omega(\mathbf{k})$ by calculating the motion of the center of the wavepacket directly. We further calculate the time evolution of the mean square displacement $\langle \mathbf{x}^2(t) \rangle$ of the wavepacket and are able to show that it consists of two time dependent terms: one quadratic in time and the other linear in time. The former can be considered as a generalization of the analogy between $\nabla_k \omega(\mathbf{k})$ and velocity.

In Sec. 2 we investigate the expansion of a Wannier function. An explicit expression for the time-dependent expectation value of the operator $\mathbf{x}^2$ is obtained. This expression is generalized in Sec. 3 to arbitrary wavepackets, and a simple expression for their time-dependent mean position is derived. The calculations are carried out in the $(\mathbf{k}, \mathbf{q})$ representation,[5] which appears to be the natural way of handling time-dependent problems with periodic potentials.

2. EXPANSION OF A WANNIER FUNCTION

The Wannier function $a(\mathbf{x})$ is defined by the following equation[3]:

$$a(\mathbf{x}) = [V/(2\pi)^3]^{1/2} \int \phi_{\mathbf{k}}(\mathbf{k}) d\mathbf{k}, \tag{2.1}$$

where $\phi_{\mathbf{k}}(\mathbf{x})$ is a Bloch eigenfunction and V the unit cell volume. Throughout this paper, the quasimomentum integrations are always performed over the first Brillouin Zone. In (2.1) the band index is omitted since we shall be dealing with Bloch functions belonging to one band only, which we assume to be nondegenerate.

Consider a wavepacket $\psi(\mathbf{x}, t)$ which satisfies the initial condition

$$\psi(\mathbf{x}, 0) = a(\mathbf{x}).$$

In order to find the time dependence of $\psi(\mathbf{x}, t)$ we have to expand $\psi(\mathbf{x}, 0)$ in terms of the Bloch functions. The expansion is given in (2.1). Consequently, we have

$$\psi(\mathbf{x}, t) = [V/(2\pi)^3]^{1/2} \int e^{-i\omega_{\mathbf{k}} t} \phi_{\mathbf{k}}(\mathbf{x}) d\mathbf{k}. \tag{2.2}$$

Our aim is to calculate the expectation value of the operator $\mathbf{x}^2$ for this wavepacket:

$$\langle \mathbf{x}^2(t) \rangle = \int \psi^*(\mathbf{x}, t) \mathbf{x}^2 \psi(\mathbf{x}, t) d\mathbf{x}. \tag{2.3}$$

We shall carry out the calculations in the $(\mathbf{k}, \mathbf{q})$ representation, introduced by Zak.[5] The use of this representation reduces the mathematical complications considerably. The basis functions in this representation are[5]

$$\theta_{\mathbf{k},\mathbf{q}}(\mathbf{x}) = [V/(2\pi)^3]^{1/2} \sum_n e^{i\mathbf{k}\mathbf{R}_n} \delta(\mathbf{x} - \mathbf{q} - \mathbf{R}_n), \tag{2.4}$$

where k and q are the quasimomentum and quasicoordinate, limited to the reciprocal and real space unit cells respectively. Zak has shown[5] that the functions $\theta_{\mathbf{k},\mathbf{q}}(\mathbf{x})$, being the eigenfunctions of the translation operators in both reciprocal and real spaces, form a complete orthonormal basis of the Hilbert space.

We now transform the Bloch functions to the $(\mathbf{k}, \mathbf{q})$ representation. First we express them in terms of the Wannier functions, inverting relation (2.1)[3]:

$$\phi_{\mathbf{k}}(\mathbf{x}) = [V/(2\pi)^3]^{1/2} \sum_n e^{i\mathbf{k}\mathbf{R}_n} a(\mathbf{x} - \mathbf{R}_n). \tag{2.5}$$

Copyright © 1973 by the American Institute of Physics

Denoting by $c_1(\mathbf{k},\mathbf{q})$ the transformed Block function corresponding to the quasimomentum quantum number 1, we obtain

$$c_1(\mathbf{k},\mathbf{q}) = \int \phi_1(\mathbf{x})\theta^*_{\mathbf{k},\mathbf{q}}(\mathbf{x})d\mathbf{x} = \frac{V}{(2\pi)^3}\sum_{n,n'}\int \delta(\mathbf{x}-\mathbf{q}-\mathbf{R}_n) \times e^{-i\mathbf{k}\mathbf{R}_n}e^{i\mathbf{l}\mathbf{R}_{n'}}a(\mathbf{x}-\mathbf{R}_n)d\mathbf{x} \quad (2.6)$$

and, introducing $\mathbf{R}_m = \mathbf{R}_n - \mathbf{R}_{n'}$ we get

$$c_1(\mathbf{k},\mathbf{q}) = \frac{V}{(2\pi)^3}\sum_{m,n} a(\mathbf{q}+\mathbf{R}_m)e^{i\mathbf{l}(\mathbf{R}_n-\mathbf{R}_m)}e^{-i\mathbf{k}\mathbf{R}_n} = \frac{V}{(2\pi)^3}\sum_{m,n} a(\mathbf{q}+\mathbf{R}_m)e^{-i\mathbf{l}\mathbf{R}_m}e^{i\mathbf{R}_n(\mathbf{l}-\mathbf{k})} = \sum_m a(\mathbf{q}+\mathbf{R}_m)e^{-i\mathbf{l}\mathbf{R}_m}\delta(\mathbf{l}-\mathbf{k}). \quad (2.7)$$

The scalar product $\langle a|\phi_{\mathbf{k}}\rangle$ is, of course, independent of the representation, so in view of (2.3) we have

$$\psi(\mathbf{k},\mathbf{q},t) = [V/(2\pi)^3]^{1/2}\int e^{-i\omega_1 t}c_1(\mathbf{k},\mathbf{q})d\mathbf{l} = [V/(2\pi)^3]^{1/2}e^{-i\omega_{\mathbf{k}}t}\int c_1(\mathbf{k},\mathbf{q})d\mathbf{l}. \quad (2.8)$$

We now introduce a function $\Lambda(\mathbf{k},\mathbf{q})$ given by

$$\Lambda(\mathbf{k},\mathbf{q}) = \int c_1(\mathbf{k},\mathbf{q})d\mathbf{l}. \quad (2.9)$$

$\Lambda(k,q)$ has three relevant properties:

$$\int |\Lambda(\mathbf{k},\mathbf{q})|^2 d\mathbf{q} = 1, \quad (2.10)$$

$\mathrm{Im}\int \nabla\Lambda\Lambda^* d\mathbf{q} = 0$ (throughout this paper all gradients are with respect to the quasi-momentum), (2.11)

$$|\Lambda(\mathbf{k},\mathbf{q})|^2 = |\Lambda(\mathbf{k},-q)|^2. \quad (2.12)$$

In order to demonstrate the first property we use the orthonormality of the Wannier functions[3]:

$$\int a(\mathbf{x}-\mathbf{R}_n)a(\mathbf{x}-\mathbf{R}_m)d\mathbf{x} = \delta_{n,m}.$$

Thus

$$\int |\Lambda(\mathbf{k},\mathbf{q})|^2 d\mathbf{q} = \sum_{n,n'} e^{i\mathbf{k}(\mathbf{R}_{n'}-\mathbf{R}_n)}\int a^*(\mathbf{q}+\mathbf{R}_{n'})a(\mathbf{q}+\mathbf{R}_n)d\mathbf{q} = \sum_{m,n} e^{i\mathbf{k}\mathbf{R}_m}\int a^*(\mathbf{q}+\mathbf{R}_m+\mathbf{R}_n)a(\mathbf{q}+\mathbf{R}_n)d\mathbf{q} = \sum_m e^{i\mathbf{k}\mathbf{R}_m}\int a^*(\mathbf{x}+\mathbf{R}_m)a(\mathbf{x})d\mathbf{x} = 1,$$

where we introduced a new variable $R_m = R_{n'} - R_n$. In the third step we used the following identity, which holds for any function $f(x)$:

$$\sum_n \int f(\mathbf{q}+\mathbf{R}_n)d\mathbf{q} = \int f(\mathbf{x})d\mathbf{x},$$

where the integration on the right-hand side covers the whole space.

In order to demonstrate the properties (2.12) and (2.13), we have to assume that the Wannier function $a(\mathbf{x})$ has the property

$$a(\mathbf{x}) = \pm a^*(-\mathbf{x}). \quad (2.13)$$

This equality is true whenever the potential $V(\mathbf{x})$ satisfies $V(\mathbf{x}) = V(-\mathbf{x})$, provided there are no degeneracies. This restriction does not reduce the generality of our treatment too seriously, since we do not require the space coordinates x, y, z to be orthogonal. In fact, the coordinate axes are always chosen in the direction of the symmetry axes of the potential. Assuming (2.13), we proceed now with the demonstration of (2.11):

$$\int \nabla\Lambda\Lambda^* d\mathbf{q} = \sum_{n,n'}\int(-i\mathbf{R}_n)a(\mathbf{q}+\mathbf{R}_n)e^{-i\mathbf{k}\mathbf{R}_n} \times a^*(\mathbf{q}+\mathbf{R}_{n'})e^{i\mathbf{k}\mathbf{R}_{n'}}d\mathbf{q} = \sum_{m,m'}\int i\mathbf{R}_m a(\mathbf{q}-\mathbf{R}_m)e^{i\mathbf{k}\mathbf{R}_m}a^*(\mathbf{q}-\mathbf{R}_{m'})e^{-i\mathbf{k}\mathbf{R}_{m'}}d\mathbf{q} = \sum_{m,m'}\int i\mathbf{R}_m a^*(\mathbf{q}'+\mathbf{R}_m)e^{i\mathbf{k}\mathbf{R}_m}a(\mathbf{q}'+\mathbf{R}_{m'})e^{-i\mathbf{k}\mathbf{R}_{m'}}d\mathbf{q}' = \int \nabla\Lambda^*\Lambda d\mathbf{q}',$$

where in the second step we introduced new variables $\mathbf{R}_m = -\mathbf{R}_n$, $\mathbf{R}_{m'} = -\mathbf{R}_{n'}$, and in the third we introduced $\mathbf{q}' = -\mathbf{q}$ and used the property (2.13). Consequently, we have

$$\int \nabla\Lambda^*\Lambda d\mathbf{q} = \int \nabla\Lambda\Lambda^* d\mathbf{q},$$

from which (2.11) follows immediately.[6]

Finally, the demonstration of (2.12) is straightforward:

$$|\Lambda(\mathbf{k},\mathbf{q})|^2 = \sum_{n,n'} e^{i\mathbf{k}(\mathbf{R}_n-\mathbf{R}_{n'})}a^*(\mathbf{q}+\mathbf{R}_n)a(\mathbf{q}+\mathbf{R}_{n'}) = \sum_{n,n'} e^{-i\mathbf{k}(\mathbf{R}_n-\mathbf{R}_{n'})}a(\mathbf{q}+\mathbf{R}_n)a^*(\mathbf{q}+\mathbf{R}_{n'}) = \sum_{m,m'} e^{i\mathbf{k}(\mathbf{R}_m-\mathbf{R}_{m'})}a(\mathbf{q}+\mathbf{R}_m)a^*(\mathbf{q}-\mathbf{R}_{m'}) = \sum_{m,m'} e^{i\mathbf{k}(\mathbf{R}_m-\mathbf{R}_{m'})}a^*(-\mathbf{q}+\mathbf{R}_m)a(-\mathbf{q}+\mathbf{R}_{m'}) = |\Lambda(\mathbf{k},-\mathbf{q})|^2,$$

where again, in the third step, we introduced $\mathbf{R}_m = -\mathbf{R}_n$, $\mathbf{R}_{m'} = -\mathbf{R}_{n'}$, and in the fourth step we used (2.13).

We can now turn to the calculation of $\langle \mathbf{x}^2(t)\rangle$. In the basis (2.4) the operator $\mathbf{x}$ is represented by[5]

$$\mathbf{x} = i\nabla + \mathbf{q}. \quad (2.14)$$

Transforming Eq. (2.3) into the $(\mathbf{k},\mathbf{q})$ representation we obtain

$$\langle \mathbf{x}^2(t)\rangle = \int \psi^*(\mathbf{k},\mathbf{q},t)(i\nabla+\mathbf{q})^2\psi(\mathbf{k},\mathbf{q},t)d\mathbf{k}d\mathbf{q} = -\int \psi^*\nabla^2\psi d\mathbf{k}d\mathbf{q} + 2i\int \psi^*\nabla\mathbf{q}\psi d\mathbf{k}d\mathbf{q} + \int \psi^*\mathbf{q}^2\psi d\mathbf{k}d\mathbf{q}. \quad (2.15)$$

The first term of (2.15) can be written

$$-\int \psi^*\nabla^2\psi d\mathbf{k}d\mathbf{q} = -[V/(2\pi)^3]\int \Lambda^* e^{i\omega_{\mathbf{k}}t}\nabla^2\Lambda e^{-i\omega_{\mathbf{k}}t}d\mathbf{k}d\mathbf{q} = [V/(2\pi)^3]\int |\nabla(\Lambda e^{-i\omega_{\mathbf{k}}t})|^2 d\mathbf{k}d\mathbf{q} \quad (2.16)$$

Here we used the fact that the surface integral over the surface of the Brillouin Zone of the functions involved vanishes. Moreover,

$$\int |\nabla(\Lambda e^{-i\omega_{\mathbf{k}}t})|^2 d\mathbf{k}d\mathbf{q} = \int [|\Lambda|^2 + t^2|\Lambda|^2(\nabla\omega_{\mathbf{k}})^2$$

$$- 2t\ \mathrm{Im}\Lambda^*\nabla\Lambda\nabla\omega_{\mathbf{k}}]d\mathbf{k}d\mathbf{q} = \int |\nabla\Lambda|^2 d\mathbf{k}d\mathbf{q} + t^2\int (\nabla\omega_{\mathbf{k}})^2 d\mathbf{k}d\mathbf{q}.$$

The second term of (2.15) yields

$$\begin{aligned} 2i\int \psi^*\nabla\mathbf{q}\psi d\mathbf{k}d\mathbf{q} &= [V/(2\pi)^3]2i\int \Lambda^* e^{i\omega_{\mathbf{k}}t}\mathbf{q}\nabla(\Lambda e^{-i\omega_{\mathbf{k}}t})d\mathbf{k}d\mathbf{q} \\ &= [V/(2\pi)^3]2i\int \Lambda^* e^{i\omega_{\mathbf{k}}t} \\ &\quad\times \mathbf{q}(e^{-i\omega_{\mathbf{k}}t}\nabla\Lambda - it\Lambda e^{-i\omega_{\mathbf{k}}t}\nabla\omega_{\mathbf{k}})d\mathbf{k}d\mathbf{q} \\ &= [V/(2\pi)^3]2i\int \Lambda^*\mathbf{q}\nabla\Lambda d\mathbf{k}d\mathbf{q} \end{aligned} \tag{2.17}$$

since, because of (2.12), the function $\int \mathbf{q}|\Lambda|^2 d\mathbf{q}$ vanishes for all $\mathbf{k}$.

Substituting (2.16) and (2.17) into (2.15), we finally obtain

$$\begin{aligned} \langle \mathbf{x}^2(t)\rangle &= [V/(2\pi)^3]\int [|\nabla\Lambda|^2 + 2i\Lambda^*\mathbf{q}\nabla\Lambda + \mathbf{q}^2|\Lambda|^2 \\ &\quad + t^2|\Lambda|^2(\nabla\omega_{\mathbf{k}})^2]d\mathbf{k}d\mathbf{q} \\ &= \int \psi^*(\mathbf{k},\mathbf{q},0)(i\nabla + \mathbf{q})^2\psi(\mathbf{k},\mathbf{q},0)d\mathbf{k}d\mathbf{q} \\ &\quad + [V/(2\pi)^3]t^2\int (\nabla\omega_{\mathbf{k}})^2 d\mathbf{k}. \end{aligned} \tag{2.18}$$

Equation (2.18) can be written compactly in the form

$$\langle \mathbf{x}^2(t)\rangle = \langle \mathbf{x}^2(0)\rangle + \langle \mathbf{v}^2\rangle t^2, \tag{2.19}$$

where

$$\langle \mathbf{v}^2\rangle = [V/(2\pi)^3]\int (\nabla\omega_{\mathbf{k}})^2 d\mathbf{k}. \tag{2.20}$$

Equation (2.19) is formally very similar to the classical equation of motion of a particle moving with constant velocity. However, it should be stressed that there is a contribution to $\langle \mathbf{x}^2(t)\rangle$ only from the spatial expansion of the wave packet, because for the Wannier wave packet we have $\langle \mathbf{x}(t)\rangle \equiv 0$.

3. EXPANSION AND MOTION OF GENERAL WAVEPACKETS

We shall consider here the general wavepacket which is expressed as an arbitrary combination of Bloch functions from a given band:

$$\psi(\mathbf{k},\mathbf{q},t) = \int g(\mathbf{l})c_1(\mathbf{k},\mathbf{q})e^{-i\omega_1 t}d\mathbf{l} = g(\mathbf{k})e^{-i\omega_{\mathbf{k}}t}\Lambda(\mathbf{k},\mathbf{q}). \tag{3.1}$$

The Wannier wave packet is a special case of (3.1) with

$$g(\mathbf{k}) = [V/(2\pi)^3]^{1/2}.$$

The treatment of Sec. II is easily generalized; in fact all one has to do is to substitute $g\Lambda$ for Λ. The following result is readily obtained:

$$\begin{aligned} \langle \mathbf{x}^2(t)\rangle &= \int \psi^*(\mathbf{k},\mathbf{q},0)(i\nabla + \mathbf{q})^2\psi(\mathbf{k},\mathbf{q},0)d\mathbf{k}d\mathbf{q} \\ &\quad + t^2\int |g|^2(\nabla\omega_{\mathbf{k}})^2 d\mathbf{k} - 2t \\ &\quad \times \mathrm{Im}\int \nabla(g\Lambda)g^*\Lambda^*\nabla\omega_{\mathbf{k}}\, d\mathbf{k}d\mathbf{q}. \end{aligned} \tag{3.2}$$

In the general case, the term linear in time does not vanish. However, it can be simplified:

$$\begin{aligned} \mathrm{Im}\int \nabla(g\Lambda)g^*\Lambda^*\nabla\omega_{\mathbf{k}}d\mathbf{k}d\mathbf{q} &= \mathrm{Im}\int (\Lambda\nabla g + g\nabla\Lambda)g^*\Lambda^*\nabla\omega_{\mathbf{k}}d\mathbf{k}d\mathbf{q} \\ &= \mathrm{Im}\int (|\Lambda|^2 g^*\nabla g\nabla\omega_{\mathbf{k}}d\mathbf{k}d\mathbf{q} + |g|^2\Lambda^*\nabla\Lambda\nabla\omega_{\mathbf{k}})d\mathbf{k}d\mathbf{q} \\ &= \mathrm{Im}\int g^*\nabla g\nabla\omega_{\mathbf{k}}d\mathbf{k} \end{aligned} \tag{3.3}$$

so that (2.22) can be put in the form

$$\langle \mathbf{x}^2(t)\rangle = \langle \mathbf{x}^2(0)\rangle - 2t\ \mathrm{Im}\int g^*\nabla g\nabla\omega_{\mathbf{k}}d\mathbf{k} + \langle \mathbf{v}^2\rangle t^2, \tag{3.4}$$

where $\langle \mathbf{v}^2\rangle$ is the generalized mean square velocity of the wavepacket:

$$\langle \mathbf{v}^2\rangle = \int |g|^2(\nabla\omega_{\mathbf{k}})^2 d\mathbf{k}. \tag{3.5}$$

We see that $\langle \mathbf{v}^2\rangle$ is obtained by averaging $(\nabla\omega_{\mathbf{k}})^2$ over the $\mathbf{k}$ space with the weight function $|g(\mathbf{k})|^2$, which is exactly the square of the modulus of the overlap between the wave packet and the Bloch function $\phi_{\mathbf{k}}$.

For the general wavepacket it is no longer true that $\langle \mathbf{x}(t)\rangle = 0$, so it is of interest to calculate it:

$$\begin{aligned} \langle \mathbf{x}(t)\rangle &= \int g^*\Lambda^* e^{i\omega_{\mathbf{k}}t}(i\nabla + \mathbf{q})g\Lambda e^{-i\omega_{\mathbf{k}}t}d\mathbf{k}d\mathbf{q} \\ &= i\int g^*\Lambda^*(\nabla(g\Lambda) - itg\Lambda\nabla\omega_{\mathbf{k}})d\mathbf{k}d\mathbf{q} \\ &= i\int g^*\nabla g d\mathbf{k} + t\int |g|^2\nabla\omega_{\mathbf{k}}d\mathbf{k}. \end{aligned} \tag{3.6}$$

Here we again used the property $|\Lambda(\mathbf{k},\mathbf{q})|^2 = |\Lambda(\mathbf{k},-\mathbf{q})|^2$.

Equation (3.6) can also be put in the form

$$\langle \mathbf{x}(t)\rangle = \langle \mathbf{x}(0)\rangle + \langle \mathbf{v}\rangle t, \tag{3.7}$$

where the average velocity is given by

$$\langle \mathbf{v}\rangle = \int |g|^2\nabla\omega_{\mathbf{k}}d\mathbf{k} \tag{3.8}$$

which is very similar to (3.5). Equations (3.7) and (3.8) express the well known analogy between $\nabla\omega_{\mathbf{k}}$ and velocity.

It is interesting to apply Eqs. (3.5) and (3.8) to a "wavepacket" consisting of one single Bloch function.[7] In this case we have

$$|g|^2 = \delta(\mathbf{k} - \mathbf{k}_0)$$

and

$$\langle \mathbf{v}^2\rangle = (\nabla\omega_{\mathbf{k}})^2|_{\mathbf{k}=\mathbf{k}_0}, \quad \langle \mathbf{v}\rangle = \nabla\omega_{\mathbf{k}}|_{\mathbf{k}=\mathbf{k}_0}.$$

This is as expected, since, as already stated, the expectation value of x in a pure Bloch state with quasimomentum $\mathbf{k}_0$ is equal to $\nabla\omega_{\mathbf{k}}|_{\mathbf{k}=\mathbf{k}_0}$.

4. SUMMARY

In the present paper a formalism was developed which permits calculation of the expectation values $\langle \mathbf{x}(t)\rangle, \langle \mathbf{x}^2(t)\rangle$ of the operators $\mathbf{x}$ and $\mathbf{x}^2$ for time dependent wavepackets subject to a three-dimensional periodic potential. We obtain expressions which, in general, justified the widely used analogy between velocity and $\nabla\omega_{\mathbf{k}}$.

An important application of the formalism developed here is in atomic diffusion theory, where it is possible to obtain an intermediate state in the jumping process which is a time dependent wavepacket. Such a wavepacket displaces itself and expands before decaying back to a localized ground state.[2] The evolution of the excited wavepacket thus becomes essential in determining the diffusion rates.

ACKNOWLEDGMENTS

We would like to thank Professor R. Englman for a critical reading of the manuscript and valuable suggestions.

* Deceased.

[1] J. H. Weiner and Y. Partom, Phys. Rev. **187**, 1134 (1969). J. H. Weiner and Y. Partom, Phys. Rev. B **1**, 1533 (1970).

[2] J. Amran Sussmann and Y. Weissman, Phys. Status Solidi **53**, 419 (1972). J. Amran Sussmann, Ann. Phys. (Paris) **6**, 135 (1971).

[3] C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963).

[4] W. A. Harrison, *Solid State Theory* (McGraw-Hill, New York, 1970).

[5] J. Zak, Phys. Rev. **168**, 686 (1968).

[6] From (2.10) it follows that $\nabla\int|\Lambda|^2 d\mathbf{q} = 0$, from which one obtains $\int\Lambda\nabla\Lambda^* d\mathbf{q} = -\int\Lambda^*\nabla\Lambda d\mathbf{q}$; so in fact the real part of $\Lambda\nabla\Lambda^* d\mathbf{q}$ vanishes as well.

[7] Formally speaking, this application involves certain mathematical complications, which we shall, however, disregard since they possess no physical significance.

Lie algebras connected with infinite momentum kinematics

A. Blasi and R. Collina

Istituto di Scienze Fisiche dell'Università-Genova

Istituto Nazionale di Fisica Nucleare-Sezione di Genova, Genova, Italy
(Received 1 March 1973)

In this paper we investigate the Lie algebras obtained from the infinite momentum limit of the Poincaré group. It will be shown that this limit depends upon a real, nonnegative number γ, and so do the structure constants of the resulting Lie algebra, which has three singular points $\gamma = 1, 0, \infty$. The last two singularities give algebras isomorphic to two different contractions of the algebra of the Poincaré group, while the case $\gamma = 1$ gives an algebra which cannot be obtained in this fashion. Suggestions are provided towards a physical interpretation of these results.

1. INTRODUCTION

The study of the kinematics in an infinite momentum reference system originated[1,2] with the hope that the dynamics too, in this limit, would become simpler. More recent mathematical treatments[3-5] of the subject, are essentially based on a "contraction" procedure[6,7] to be applied to the Lie algebra P of the Poincaré group to arrive at a new algebra which should reflect the kinematics of the infinite momentum systems. Since, in general, there is more than one way to contract a Lie algebra (which is the case for P), different results are obtained, with accordingly different physical interpretations. The infinite momentum limit is embodied in the contraction parameter, which in the limit, usually attained for the value zero, gives the new algebra.

Note that if the relativistic kinematics are described by the Poincaré group, then two limits have been performed; first to have the algebra P and second to contract P itself in order to arrive at the infinite momentum reference system. In this paper we shall first consider the infinite momentum limit on the Poincaré group, and then extract from this an eventual algebraic structure. This procedure requires the investigation of a Poincaré transformation near a branching point approached when the relative velocity between systems tends to c, and hence a definition of this limit is needed.

To this end we employ three reference systems and the linear mapping between two of them when their relative velocities with respect to the third one tend to the velocity of light $c = 1$ in a preassigned direction.

In Sec. 2 we carry out this step and show how the mapping depends upon a real nonnegative number γ which must be kept fixed when computing the infinitesimal transformations.

Section 3 is devoted to the study of the commutation rules of these infinitesimal transformations, rules which are γ-dependent and in general are not algebraically closed. The closure is obtained adjointing a new generator d.

The structure constants of this algebra have three singular points $\gamma = 1, 0, \infty$. The case $\gamma = 1$, treated in Sec. 4, is the only value of γ for which the commutation rules are closed without the adjunction of d, giving a new algebra with 9 generators.

In Sec. 5 the algebra for the singularity $\gamma = 0$ is shown to correspond to the one of Ref. 3, which can be obtained as a contraction, explicitly given in the Appendix, of the algebra P.

The case $\gamma = \infty$, treated in Sec. 6, yields an algebra identical to that of Ref. 5.

All cases contain some remarks about a possible physical interpretation and their common features, as it is further stressed in the conclusive Sec. 7.

2. THE INFINITESIMAL GENERATORS

Consider three reference systems S, S', S''; let $\mathbf{v}$ be the velocity of S' and $\mathbf{u}$ that of S'' with respect to S both measured in S, while $\mathbf{u}'$ is the velocity of S'' with respect to S' as measured in S'. We are interested in the connection between S' and S'' when both $|\mathbf{v}|$ and $|\mathbf{u}|$ approach the velocity of light $c = 1$ and, for the time being, we shall take into account only the relative motions of the origins. From the relativistic addition low for velocities we have

$$\mathbf{u}' = ((1 - v^2)^{1/2}\mathbf{u} + \{[1 - (1 - v^2)^{1/2}]\mathbf{u}\cdot\mathbf{v} - 1\}\mathbf{v})/(1 - \mathbf{v}\cdot\mathbf{u}) \quad (2.1)$$

and, denoting with $\parallel$ the direction parallel to $\mathbf{v}$ while $\perp$ indicates the plane perpendicular to $\mathbf{v}$, it follows that

$$\mathbf{u} = (1 - u_{\parallel}^2)^{1/2}\mathbf{u}_{\perp} + \mathbf{u}_{\parallel}, \quad (2.2)$$

$$\mathbf{u}' = (1 - u_{\parallel}'^2)^{1/2}\mathbf{u}_{\perp}' + \mathbf{u}_{\parallel}'. \quad (2.3)$$

The substitution of (2.2) into (2.1) yields

$$\mathbf{u}' = \{[(1 - v^2)(1 - u_{\parallel}^2)]^{1/2}\mathbf{u}_{\perp} + [(u_{\parallel}/v) - 1]\mathbf{v}\}/(1 - \mathbf{v}\cdot\mathbf{u}_{\parallel}), \quad (2.4)$$

which, compared with (2.3), gives

$$\mathbf{u}_{\parallel}' = (\mathbf{u}_{\parallel} - \mathbf{v})/(1 - \mathbf{v}\cdot\mathbf{u}_{\parallel}) \quad (2.5a)$$

$$\mathbf{u}_{\perp}' = \mathbf{u}_{\perp}. \quad (2.5b)$$

In order to perform the limits $|\mathbf{v}| = v \to 1$ and $|\mathbf{u}_{\parallel}| = u_{\parallel} \to 1$ we set

$$v = (1 - \alpha), \quad (2.6a)$$

$$u_{\parallel} = (1 - \beta), \quad (2.6b)$$

$$\alpha/\beta = \gamma, \quad (2.6c)$$

where α, β, γ are real nonnegative numbers; hence

$$u_{\parallel}' = (\gamma - 1)/(\gamma + 1), \quad (2.7a)$$

$$u'^2 = [(\gamma - 1)^2 + 4\gamma u_{\perp}'^2)/(\gamma + 1)^2. \quad (2.7b)$$

With the help of (2.7a) and (2.7b) we can write the finite Lorentz boost between S' and S'' as follows:

Copyright © 1973 by the American Institute of Physics

$$x_0'' = \frac{\gamma+1}{2[\gamma(1-u_\perp'^2)]^{1/2}}\left(x_0' - \frac{x_\parallel'(\gamma-1) + 2\mathbf{x}_\perp'\cdot\mathbf{u}_\perp'\gamma^{1/2}}{\gamma+1}\right), \tag{2.8a}$$

$$x_\parallel'' = x_\parallel' + (\gamma-1)\left[\frac{x_\parallel'(\gamma-1) + 2\mathbf{x}_\perp'\cdot\mathbf{u}_\perp'\gamma^{1/2}}{4\gamma u_\perp'^2 + (\gamma-1)^2}\left(\frac{\gamma+1}{2[\gamma(1-u_\perp'^2)]^{1/2}} - 1\right) - \frac{x_0'}{2[\gamma(1-u_\perp'^2)]^{1/2}}\right], \tag{2.8b}$$

$$\mathbf{x}_\perp'' = \mathbf{x}_\perp' + 2\gamma^{1/2}\mathbf{u}_\perp'\left[\frac{x_\parallel'(\gamma-1) + 2\mathbf{x}_\perp'\cdot\mathbf{u}_\perp'\gamma^{1/2}}{4\gamma u_\perp'^2 + (\gamma-1)^2}\left(\frac{\gamma+1}{2[\gamma(1-u_\perp'^2)]^{1/2}} - 1\right) - \frac{x_0'}{2[\gamma(1-u_\perp'^2)]^{1/2}}\right]. \tag{2.8c}$$

As expected, the limiting procedure for v and $u_\parallel$ is not exactly defined, but depends upon the ratio γ; we shall take γ as a real parameter and investigate the structure of the transformations (2.8) as a function of γ. Of course the group property of (2.8) may be retained only for some values of γ, i.e., for a specified way of performing the limits on v and $u_\parallel$, but we notice that when the free "group" parameter $\mathbf{u}_\perp'$ is zero, (2.8) become

$$x_\parallel'' = (2\gamma^{1/2})^{-1}\,[(\gamma+1)x_\parallel' - (\gamma-1)x_0'], \tag{2.9a}$$

$$\mathbf{x}_\perp'' = \mathbf{x}_\perp', \tag{2.9b}$$

$$x_0'' = (2\gamma^{1/2})^{-1}\,[(\gamma+1)x_0' - (\gamma-1)x_\parallel']. \tag{2.9c}$$

The transformation (2.9) has 4×4 matrix $D(\gamma)$

$$D(\gamma) = \begin{pmatrix} (\gamma+1)/2\sqrt{\gamma} & 0 & 0 & (1-\gamma)/2\sqrt{\gamma} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ (1-\gamma)/2\sqrt{\gamma} & 0 & 0 & (1+\gamma)/2\sqrt{\gamma} \end{pmatrix} \tag{2.10}$$

on the vector basis (x_0', x_1', x_2', x_3') where $x_\perp' = (0, x_1', x_2', 0)$, $x_\parallel' = x_3'$. This matrix becomes diagonal on the light-cone variables basis $y_2 = x_2'$, $y_1 = x_1'$, $y_0 = x_0' + x_3'$, $y_3 = x_0' - x_3'$, where it can be interpreted as an anisotropic dilatation, and form a one parameter group with multiplication rule $D(\gamma)D(\lambda) = D(\gamma\lambda)$. It is then desirable to factor out the dilatations from the limit of the Lorentz transformations since we would like to have an operation like (2.8) which reduces to the identity when $\mathbf{u}_\perp' \to 0$. The factorization is carried out for the infinitesimal transformations, i.e., when $\mathbf{u}_\perp' \to 0$, in which case (2.8) are described by the product

$$\begin{pmatrix} \frac{\gamma+1}{2\sqrt{\gamma}} & 0 & 0 & \frac{1-\gamma}{2\sqrt{\gamma}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1-\gamma}{2\sqrt{\gamma}} & 0 & 0 & \frac{\gamma+1}{2\sqrt{\gamma}} \end{pmatrix} \begin{pmatrix} 1 & -u_1' & -u_2' & 0 \\ -u_1' & 1 & 0 & u_1'\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} \\ -u_2' & 0 & 1 & u_2'\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} \\ 0 & -u_1'\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} & -u_1'\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} & 1 \end{pmatrix}, \tag{2.11}$$

where the l.h.s. matrix is a $D(\gamma)$ while the r.h.s. gives the two infinitesimal, γ-dependent transformations:

$$K_1 = \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & (\sqrt{\gamma}-1)/(\sqrt{\gamma}+1) \\ 0 & 0 & 0 & 0 \\ 0 & -(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1) & 0 & 0 \end{pmatrix}, \tag{2.12a}$$

$$K_2 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & (\sqrt{\gamma}-1)/(\sqrt{\gamma}+1) \\ 0 & 0 & -(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1) & 0 \end{pmatrix}. \tag{2.12b}$$

We remark that the extraction of the dilatation $D(\gamma)$ is not necessary when $\gamma = 1$ since $D(1) = 1$, but this procedure must be followed for all other values of γ; the factorization may be equivalently accomplished through a redefinition of the basis vectors on which (2.8) act.

In the above derivation of formulae (2.12), we have taken into account only the Lorentz boost between the two systems S' and S''; the appearance of the extra terms $(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1)$ in the generators K_1 and K_2 is due to the limiting process performed on the finite transformation.

Having found the generators of the Lorentz boosts, we now turn to the investigation of the remaining ones, i.e., the rotations between S' and S'', employing the 2 to 1 homomorphism of the Lorentz group with $SL(2,C)$.[8] We now must specify the direction of the axes and set those of S'' parallel to the axes of S. Let: $\tilde{x} = x_0 + x_k\sigma_k$, with σ_k the usual Pauli matrices, be the matrix on which the elements $A(\mathbf{u})$ of $SL(2,C)$ act; if $R(\mathbf{n},\theta) \in SU(2,C)$ is the rotation which brings the axes of S' parallel to those of S, and $R(\mathbf{n}',\theta') \in SU(2,C)$ the one which does the same for the axes of S' with respect to S'', we have

$$\tilde{x}'' = A(\mathbf{u})\tilde{x}A(\mathbf{u})^\dagger, \tag{2.13a}$$

$$\tilde{x}'' = A(\mathbf{u}')R(\mathbf{n}',\theta')\tilde{x}'R(\mathbf{n}',\theta')^\dagger A(\mathbf{u})^\dagger, \tag{2.13b}$$

$$\tilde{x} = A(-\mathbf{v})R(\mathbf{n},\theta)\tilde{x}'R(\mathbf{n},\theta)^\dagger A(-\mathbf{v})^\dagger, \tag{2.13c}$$

where, substituting (2.13c) into (2.13a) and comparing with (2.13b) we obtain

$$R(\mathbf{n}', \theta') = A(-\mathbf{u}')A(\mathbf{u})A(-\mathbf{v})R(\mathbf{n}, \theta). \tag{2.14}$$

To compute the generators, we first perform the limiting procedure, with the subsequent appearance of the parameter γ, and then derive with respect to the angles; note that no dilatation must be factored out here, since this has already been done for the Lorentz boost [see formula (2.11)]. Written in 4×4 notation for convenience, these generators are

$$\Lambda_1 = \begin{pmatrix} 0 & (1-\sqrt{\gamma})/\sqrt{\gamma} & 0 & 0 \\ (1-\sqrt{\gamma})/\sqrt{\gamma} & 0 & 0 & (1-\sqrt{\gamma})/\sqrt{\gamma}(\sqrt{\gamma}+1) \\ 0 & 0 & 0 & 0 \\ 0 & -(1-\sqrt{\gamma})/\sqrt{\gamma}(\sqrt{\gamma}+1) & 0 & 0 \end{pmatrix}, \tag{2.15a}$$

$$\Lambda_2 = \begin{pmatrix} 0 & 0 & -(1-\sqrt{\gamma})/\sqrt{\gamma} & 0 \\ 0 & 0 & 0 & 0 \\ -(1-\sqrt{\gamma})/\sqrt{\gamma} & 0 & 0 & (1-\sqrt{\gamma})/\sqrt{\gamma}(\sqrt{\gamma}+1 \\ 0 & 0 & -(1-\sqrt{\gamma})/\sqrt{\gamma}+1) & 0 \end{pmatrix}, \tag{2.15b}$$

$$\Lambda_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.15c}$$

It is apparent that only Λ_3 has the structure of the usual generator of a rotation around the third axis, while Λ_1 and Λ_2 contain extra terms which arise in the limit.

For the generators of the translations we write the transformation in homogeneous coordinates, i.e., with 5×5 matrices, extract the dilatation as shown in (2.11) and obtain

$$P_0 = \left(\begin{array}{c|c} & (\gamma+1)/2\sqrt{\gamma} \\ & 0 \\ \mathbf{0} & 0 \\ & (1-\gamma)/2\sqrt{\gamma} \\ \hline 0 & 0 \end{array}\right), \tag{2.16a}$$

$$P_1 = \left(\begin{array}{c|c} & 0 \\ & 1 \\ \mathbf{0} & 0 \\ & 0 \\ \hline 0 & 0 \end{array}\right), \tag{2.16b}$$

$$P_2 = \left(\begin{array}{c|c} & 0 \\ & 0 \\ \mathbf{0} & 1 \\ & 0 \\ \hline 0 & 0 \end{array}\right), \tag{2.16c}$$

$$P_3 = \left(\begin{array}{c|c} & (1-\gamma)/2\sqrt{\gamma} \\ & 0 \\ \mathbf{0} & 0 \\ & (1+\gamma)/2\sqrt{\gamma} \\ \hline 0 & 0 \end{array}\right). \tag{2.16d}$$

At this point we shall study the possible algebraic structures which can be obtained from (2.12), (2.15), and (2.16); naturally these structures will depend upon the parameter γ describing the limit.

3. THE GENERAL P-∞ ALGEBRA

The commutators among the 9 matrices $K_1, K_2, \Lambda_1, \Lambda_2, \Lambda_3, P_1, P_2, T_0 = P_0 - P_3, T_3 = P_0 + P_3$ are

$$\begin{aligned}
&[K_1, K_2] = \{1 - [(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1)]^2\}\Lambda_3, \\
&[K_1, \Lambda_2] = [K_2, \Lambda_1] = [(3-\gamma-2\sqrt{\gamma})/(\sqrt{\gamma}+1)^2]\Lambda_3, \\
&[K_1, \Lambda_3] = K_2, \quad [K_2, \Lambda_3] = -K_1, \\
&[\Lambda_1, \Lambda_2] = \{[\sqrt{\gamma}(1-\gamma) + 2(\sqrt{\gamma}-1)]/\sqrt{\gamma}(\sqrt{\gamma}+1)^2\}\Lambda_3, \\
&[\Lambda_1, \Lambda_3] = -\Lambda_2, \quad [\Lambda_2, \Lambda_3] = \Lambda_1, \\
&[K_1, T_0] = -[2\gamma/(\sqrt{\gamma}+1)]P_1, \\
&[K_1, T_3] = -[2/\sqrt{\gamma}(\sqrt{\gamma}+1)]P_1,
\end{aligned} \tag{3.1}$$

$$\begin{aligned}
&[K_1, P_1] = [K_2, P_2] = -(T_0/\sqrt{\gamma} + \gamma T_3)/(1+\sqrt{\gamma}), \\
&[K_2, T_0] = -2\gamma/(\sqrt{\gamma}+1)P_2, \\
&[K_2, T_3] = -2/[\sqrt{\gamma}(\sqrt{\gamma}+1)]P_2, \\
&[\Lambda_1, T_0] = [\sqrt{\gamma}(1-\sqrt{\gamma})/(1+\sqrt{\gamma})]P_1, \\
&[\Lambda_1, T_3] = [(2-\sqrt{\gamma}-\gamma)/\gamma(\sqrt{\gamma}+1)]P_1, \\
&[\Lambda_1, P_1] = [P_2, \Lambda_2] \\
&\qquad = [(1-\sqrt{\gamma})/2\sqrt{\gamma}(\sqrt{\gamma}+1)][(T_0/\sqrt{\gamma})(2+\sqrt{\gamma}) + \gamma T_3], \\
&[\Lambda_2, T_0] = [\sqrt{\gamma}(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1)]P_2, \\
&[\Lambda_2, T_3] = [(\gamma+\sqrt{\gamma}-2)/\gamma(\sqrt{\gamma}+1)]P_2, \\
&[\Lambda_3, P_1] = -P_2, \quad [\Lambda_3, P_2] = P_1. \\
&[K_1, \Lambda_1] = [\Lambda_2, K_2] = [(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1)]d,
\end{aligned} \tag{3.2}$$

all others being zero and where

$$d = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.3}$$

These commutators, for an arbitrary value of γ do not give an algebraically closed structure due to the appearance in (3.2) of the matrix d; it is evident though, that for the particular value $\gamma = 1$ the r. h. s. of (3.2) vanishes and we do obtain an algebra. The matrix d must then be added to the other generators in order to have an algebra for any value of γ; its physical interpretation may be that of a dilatation generator on the light cone variables (case $\gamma = 0$) or that of a Lorentz boost generator along the third direction (cast $\gamma = \infty$), but, although formally it corresponds to the generator of the transformation (2.10) if γ were treated as a free parameter, its presence in the general algebra can be justified only by the closure argument. Recall in fact that the limiting procedure on the Poincaré group is performed with γ fixed, and in the same way are derived the matrices (2.12), (2.15), (2.16).

The resulting commutators are

$$
\begin{aligned}
&[d,K_1] = [(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1)]K_1,\\
&[d,K_2] = [(\sqrt{\gamma}-1)/(\sqrt{\gamma}+1)]K_2,\\
&[d,\Lambda_1] = [1/(\sqrt{\gamma}+1)]\{[(\sqrt{\gamma}-1)(\sqrt{\gamma}+2)/\sqrt{\gamma}\,]K_1 - (3+\sqrt{\gamma})\Lambda_1\}, \qquad (3.4)\\
&[d,\Lambda_2] = [1/(\sqrt{\gamma}+1)]\{[(1-\sqrt{\gamma})(\sqrt{\gamma}+2)/\sqrt{\gamma}\,]K_2 - (3+\sqrt{\gamma})\Lambda_2\},\\
&[d,T_0] = -T_0, \quad [d,T_3] = T_3,\\
&[d,\Lambda_3] = [d,P_1] = [d,P_2] = 0.
\end{aligned}
$$

It is now interesting to study the γ dependence of the algebra (3.1), (3.2), (3.4); dependence which is all contained in the structure constants and no longer in the particular representation (2.12), (2.15), (2.16) of the generators. As functions of γ the structure constants have three singular, real, nonnegative points, $\gamma = 0, 1, \infty$. A singularity must be intended as a value of γ which does does not give an algebra isomorphic to the complete one (3.1), (3.2), (3.4). It is possible that near a singularity some structure constants may diverge; then this divergence must be removed by a redefinition of the generators, thus obtaining an algebra isomorphic to the old one away from the singularity, but whose structure constants converge to a finite non zero value at the singular point. As we shall see, this is the case for $\gamma = 0, \gamma = \infty$, while for $\gamma = 1$ the commutation rules (3.1), (3.2) are already closed with no divergent structure constants. The important fact is that the γ singularities are all assigned by the limiting procedure, and hence so are the algebraic structures finally obtained.

4. CASE $\gamma = 1$

Setting

$$
\begin{aligned}
&J_a = i\Lambda_a,\\
&L_a = iK_a, \qquad (4.1)\\
&J_3 = -i\Lambda_3,\\
&Q_a = P_a, \quad Q_0 = \tfrac{1}{2}(T_0 + T_3), \quad Q_3 = \tfrac{1}{2}(T_3 - T_0),\\
&a = 1, 2,
\end{aligned}
$$

we obtain, for $\gamma = 1$ the commutators

$$
\begin{aligned}
&[L_a, L_b] = -i\epsilon_{ab}J_3,\\
&[J_3, L_a] = i\epsilon_{ab}L_b,\\
&[J_3, J_a] = i\epsilon_{ab}J_b, \qquad (4.2)\\
&[L_a, Q_0] = -iQ_a,\\
&[L_a, Q_b] = -i\delta_{ab}Q_0,\\
&[J_3, Q_a] = i\epsilon_{ab}Q_b,
\end{aligned}
$$

while all others are zero.

As it is clear from (4.2) the operators $L_1, L_2, J_3, Q_0, Q_1, Q_2$ form a subalgebra isomorphic to that of the inhomogeneous L_{2+1} Lorentz group, the commuting operators J_1 and J_2 are an abelian ideal, and Q_3, commuting with all the other generators, can be adjoined as a direct product. Note further that J_1, J_2, J_3 form the algebra E_2 of the two-dimensional Euclidean group.

There are three independent Casimir operators[9] in the algebra (4.2):

$$Q_3, \qquad (4.3a)$$

$$Q_0^2 - Q_1^2 - Q_2^2, \qquad (4.3b)$$

$$J_1^2 + J_2^2. \qquad (4.3c)$$

If Q_0, Q_1, Q_2, Q_3 are given the meaning of components of the four momentum, then Q_3 selects the particular class of reference systems in which the phenomenon is described, while (4.3b) indicates that there are only two independent components for the spatial momentum. Once the value of Q_3 is assigned, we actually deal with a two-dimensional motion taking place in a plane perpendicular to the direction of Q_3 itself, as it is further confirmed from the sole presence, in (4.2), of the operators L_1 and L_2 which correspond to the transverse Lorentz boost generators. The only admissible rotation takes place about the third axis with generator J_3. The harmonic analysis of the rotational properties of physical systems, in this limit, must be carried out according to the algebra of J_1, J_2, J_3 which is that of E_2 and this may lead to a description of scattering processes in terms of the impact parameter.[5,10]

5. CASE $\gamma = 0$

As remarked in the introduction, when $\gamma = 0$ some structure constants diverge and it is necessary to redefine the generators; in particular we must change $\Lambda_a \to \sqrt{\gamma}\,\Lambda_a$, $T_0 \to T_0/\sqrt{\gamma}$, $T_3 \to \sqrt{\gamma}\,T_3$. The divergence is essentially brought about by the dilatation (2.10) for $\gamma = 0$, and the above redefinition counteracts its effects on the light-cone variables basis. For convenience we shall also consider the following linear, nonsingular combinations of the newly defined generators:

$$
\begin{aligned}
&J_1 = -\tfrac{1}{2}i(K_2 - \sqrt{\gamma}\Lambda_2), \quad J_2 = \tfrac{1}{2}i(K_1 + \sqrt{\gamma}\Lambda_1), \quad J_3 = -i\Lambda_3,\\
&G_1 = \tfrac{1}{2}i(K_1 - \sqrt{\gamma}\Lambda_1), \quad G_2 = \tfrac{1}{2}i(K_2 + \sqrt{\gamma}\Lambda_2), \quad D = -id, \qquad (5.1)\\
&Q_0 = T_0/\sqrt{\gamma}, \quad Q_1 = P_1, \quad Q_2 = P_2, \quad Q_3 = \sqrt{\gamma}\,T_3.
\end{aligned}
$$

In the limit $\gamma = 0$ we obtain the commutators:

$$
\begin{aligned}
&[J_a, J_3] = -i\epsilon_{ab}J_b,\\
&[G_a, J_3] = -i\epsilon_{ab}G_b,\\
&[G_a, Q_3] = -2iQ_a,\\
&[G_a, Q_b] = -i\delta_{ab}Q_0, \qquad (5.2)\\
&[J_3, Q_a] = i\epsilon_{ab}Q_b,
\end{aligned}
$$

$$[D, G_a] = iG_a,$$
$$[D, Q_0] = iQ_0, \quad [D, Q_3] = -iQ_3, \quad a, b = 1, 2,$$

all others being zero.

The subalgebra generated by $G_1, G_2, J_3, Q_1, Q_2, Q_3, D$ is isomorphic to that of the two-dimensional, nonzero mass Galilei group plus the dilatation D, while J_1, J_2 form an Abelian ideal in (5.2).

The complete algebra (5.2) can be obtained by a contraction of the Poincaré group algebra P, as shown in the Appendix, arriving at the same results of Ref. 3.

Note that, similarly to the case $\gamma = 1$, the generators J_1, J_2, J_3 form the algebra E_2, with consequences, for the harmonic analysis of the rotational properties of physical systems, analogous to those already mentioned in Sec. 4, once for J_1 and J_2 are substituted the correct linear combinations (5.1) in terms of the generators obtained from the infinite momentum limit of the Poincaré group as specified in Sec. 2.

This algebra has two Casimir operators $Q_1^2 + Q_2^2 - Q_0Q_3$, $Q_3^2(J_1^2 + J_2^2)$ which can be assigned a physical meaning in accordance with the "Galilean invariance" interpretation of Refs. 3 and 4.

6. CASE $\gamma = \infty$

In this instance too it is necessary to redefine some generators, namely $T_0 \to T_0/\gamma^{1/2}, T_3 \to \gamma^{1/2}T_3$, to have finite structure constants.

We then set

$$A_a = (-1)^a iK_a, \quad B_a = (-1)^a i(\Lambda_a - \tfrac{1}{2}K_a), \quad a = 1,2,$$
$$J_3 = i\Lambda_3, \quad K_3 = id,$$
$$A = \sqrt{\gamma}T_3, \quad B = T_0/2\sqrt{\gamma}, \quad Q_a = -(-1)^a P_a \tag{6.1}$$

and obtain the following commutation rules:

$$[J_3, A_a] = i\epsilon_{ab}A_b,$$
$$[K_3, A] = iA, \quad [K_3, A_a] = iA_a, \quad [K_3, B] = -iB,$$
$$[K_3, B_a] = -iB_a, \quad [A_a, B_b] = i\delta_{ab}K_3 - i\epsilon_{ab}J_3, \tag{6.2}$$
$$[A_a, P_b] = i\delta_{ab}A, \quad [B_a, P_b] = i\delta_{ab}B,$$
$$[A_a, B] = iP_a, \quad [B_a, A] = iP_a,$$

all others being zero.

The algebra they define is identical (with the same notation) to the one mentioned in Ref. 5 and hence it is isomorphic to P; the generators A_1, A_2, J_3 form the subalgebra E_2 with a physical interpretation similar to that proposed in Ref. 5.

Obviously the Casimir operators of (6.2) are deducible from those of P through the isomorphic mapping which links the two algebras.

7. CONCLUSIONS

The questions posed in the introduction have been answered at the level of structure of the Lie algebras involved; the procedure here adopted of first finding the infinite momentum limit of the Poincaré group and then computing the infinitesimal generators, yields an algebra with commutation rules (3.1), (3.2), (3.4) which contains the cases treated in the literature, i.e., $\gamma = 0$ and $\gamma = \infty$, and which furthermore gives a new algebraic structure for the remaining singularity at $\gamma = 1$. We may mention that this last case, besides being the most direct since it does not need a redefinition of the generators to have a finite limit, is also the only one which cannot be obtained as a contraction of the Poincaré group algebra P.

The more interesting fact common to all three alternatives is that, in an infinite momentum frame, we essentially deal with a motion in two dimensions, which is reflected by the presence of the E_2 subalgebra in all three algebras.

This subalgebra is connected with the rotational properties of physical systems in such frames, where then the harmonic analysis must be carried out according to this result.

It is our hope that further insights could be gained by a similar procedure applied to the irreducible, unitary representations of the Poincaré group of its algebra P, especially for what concerns the physical interpretation of phenomena described in an infinite momentum reference frame.

ACKNOWLEDGMENTS

We would like to thank Professors E. Eberle and G. Viano for helpful discussions.

APPENDIX

Let M_i, N_i, Π_μ where $i = 1, 2, 3, \mu = 0, 1, 2, 3$ be the generators of the rotations, Lorentz boosts, and translations, respectively, in the Poincaré group algebra with commutation rules

$$[M_i, M_j] = i\epsilon_{ijk}M_k, \quad [N_i, \Pi_j] = i\delta_{ij}\Pi_0,$$
$$[N_i, N_j] = -i\epsilon_{ijk}M_k, \quad [N_i, \Pi_0] = i\Pi_i,$$
$$[M_i, N_j] = i\epsilon_{ijk}N_k, \quad [\Pi_\mu, \Pi_\nu] = 0.$$
$$[M_i, \Pi_j] = i\epsilon_{ijk}\Pi_k,$$
$$[M_i, \Pi_0] = 0,$$

We perform the following nonsingular change of basis:

$$J_a = M_a - \epsilon_{ab}N_b, \quad G_a = N_a - \epsilon_{ab}M_b, \quad a,b = 1,2,$$
$$J_3 = M_3, \quad D = N_3,$$
$$Q_0 = \Pi_0 - \Pi_3, \quad Q_1 = \Pi_1, Q_2 = \Pi_2, \quad Q_3 = \Pi_0 + \Pi_3,$$

and contract following Inonü and Wigner,[6] with respect to the generators J_a substituting J_a with ηJ_a and letting η tend to zero. It is easily verified that the new commutators thus obtained are identical with (5.2).

[1]S. Fubini and G. Furlan, Physics (N.Y.) **4**, 229 (1965); R. Dashen and M. Gell-Mann, "Coral Gable Conference Report" (1966).
[2]S. Weinberg, Phys. Rev. **150**, 1313 (1966).
[3]L. Susskind, Phys. Rev. **165**, 1535 (1967).
[4]M. P. Halpern, Lect. Theor. Phys. **11A**, 345 (1969).
[5]H. Bacry and N. P. Chang, Ann. Phys. (N.Y.) **47**, 407 (1968).
[6]E. Inonü and E. P. Wigner, Proc. Natl. Acad. Sci. USA **39**, 510 (1953).
[7]E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[8]A. S. Wightman, in *Dispersion Relations and Elementary Particles*, edited by De Witt and Omnes (Wiley, New York, 1960), p. 161.
[9]E. G. Beltrametti and A. Blasi, Phys. Lett. **20**, 62 (1966).
[10]G. A. Viano, Commun. Math. Phys. **26**, 290 (1972).

Existence and uniqueness of solutions of the Hamiltonian constraint of general relativity on compact manifolds* †

Niall O'Murchadha‡ and James W. York Jr.§

Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540

(Received 9 April 1973)

The Hamiltonian constraint "$G^{00} = 8\pi T^{00}$" of general relativity is written as a quasilinear elliptic differential equation for the conformal factor of the metric of a three-dimensional spacelike manifold. It is shown that for "almost every" configuration of initial data on a compact manifold, with or without boundary, a solution exists. Dirichlet boundary conditions are assumed if the boundary is not empty. The solution is *unique*.

1. INTRODUCTION AND SUMMARY

An outstanding problem in general relativity has been to identify a minimal set of Cauchy data for the gravitational field. These data may be defined as those "coordinate" and "momentum" variables which can be freely given on an initial spacelike surface, but once given, completely define the field on the initial surface and for a finite time into the future.[1] The identification of these variables is of importance in several areas, for example: (1) in attempts to produce a quantum theory of gravity; (2) in astrophysical or cosmological situations in which exact solutions are not known; (3) in describing the properties of the energy of a gravitational field, and (4) in investigating the stability of solutions of Einstein's equations.

It has been demonstrated that the initial value data are most naturally described by the intrinsic geometry of a spacelike slice, i.e., as a three-dimensional manifold V, with a Riemannian metric g_{ij} defined on V; and by the extrinsic curvature K_{ij}. In a Hamiltonian formulation of general relativity g_{ij} can be treated as the "position" variable, and the "momentum" π^{ij} conjugate to it is closely related to the extrinsic curvature:

$$\pi^{ij} = g^{1/2}(Kg^{ij} - K^{ij}). \tag{1}$$

However, g_{ij} and π^{ij} cannot be given independently on V, but must obey four constraints. In vacuum, the initial value equations are

$$\nabla_j \pi^{ij} = 0, \tag{2}$$

$$g^{1/2}R - g^{-1/2}\left(\pi_{ij}\pi^{ij} - \frac{1}{2}\pi^2\right) = 0, \tag{3}$$

where Eq. (2) is known as the "momentum constraint," and Eq. (3) as the "Hamiltonian constraint."

The Hamiltonian constraint is so-called because it generates the time development of the g_{ij}'s and the π^{ij}'s. It has been shown[2] that the independent data are certain conformally invariant fields defined on the initial manifold V. The conformal factor ϕ is determined by the Hamiltonian constraint. The purpose of this paper is to show that this equation can always be solved and that the solution is unique.

Suppose that $\sigma^{ij}_{\mathrm{TT}}$ is a transverse tracefree tensor density of unit weight with respect to g_{ij}, i.e.,

$$\nabla_j \sigma^{ij}_{\mathrm{TT}} = 0, \quad g_{ij}\sigma^{ij}_{\mathrm{TT}} = 0. \tag{4}$$

Then under the conformal transformation

$$g'_{ij} = \phi^4 g_{ij}, \tag{5}$$

$$\sigma^{ij\prime}_{\mathrm{TT}} = \phi^{-4}\sigma^{ij}_{\mathrm{TT}}, \tag{6}$$

where ϕ is any strictly positive real function, then $\sigma^{ij\prime}_{\mathrm{TT}}$ is transverse and traceless with respect to g'_{ij}, as can be readily verified.

Therefore, if we are given a manifold with a Riemannian metric g_{ij}, with a transverse, traceless tensor density $\sigma^{ij}_{\mathrm{TT}}$ as the momentum, then the momentum constraint is satisfied but the Hamiltonian constraint will not be, in general. Now we conformally transform to a new initial data set. Under the conformal transformation (5) and (6) the momentum constraint remains satisfied for any ϕ. We try to pick a ϕ such that the Hamiltonian constraint is satisfied, i.e.,

$$R' = (g')^{-1}\left(\pi^{ij\prime}\pi'_{ij} - \frac{1}{2}\pi'^{1/2}\right) = (g')^{-1}\sigma^{ij\prime}_{\mathrm{TT}}\sigma^{\mathrm{TT}\prime}_{ij}. \tag{7}$$

(Note that here $\pi' = 0$.)

Since

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

the Hamiltonian constraint becomes [I][2]:

$$8\nabla^2\phi = R\phi - M\phi^{-7}, \quad M \equiv g^{-1}\sigma^{ij}_{\mathrm{TT}}\sigma^{\mathrm{TT}}_{ij}. \tag{9}$$

So far we have assumed that the momentum is traceless, but this is not necessary. In general the momentum will have a trace:

$$\pi^{ij} = \sigma^{ij} + \frac{1}{2}g^{1/2}g^{ij}\tau, \tag{10}$$

where $\tau = (2/3)\pi g^{-1/2}$ is a scalar. The momentum π^{ij} must be transverse. A particularly simple way to satisfy this requirement is to pick any transverse traceless tensor density $\sigma^{ij}_{\mathrm{TT}}$ and any constant τ. We can then regard $\sigma^{ij}_{\mathrm{TT}}$ and τ as independent initial value data.[3] Given a metric g_{ij} and such a choice of $\sigma^{ij}_{\mathrm{TT}}$ and τ, then for arbitrary ϕ,

$$\pi^{ij\prime} = \sigma^{ij\prime}_{\mathrm{TT}} + \frac{1}{2}(g')^{1/2}g^{ij\prime}\tau \tag{11}$$

continues to satisfy the momentum constraint for metric g'_{ij} and $\sigma^{ij\prime}_{\mathrm{TT}}$ as in (5) and (6). Notice that τ is the fixed constant originally chosen.

The Hamiltonian constraint becomes the "scale equation" for the conformal factor [IV][2]:

$$8\nabla^2\phi = R\phi - M\phi^{-7} + \tfrac{3}{8}\tau^2\phi^5. \tag{12}$$

In the presence of external sources, the initial value equations become [IV][2]:

Copyright © 1973 by the American Institute of Physics

$$-2\nabla_j \pi^{ij} = 16\pi S^i, \tag{13}$$

$$R - g^{-1}\left(\pi^{ij}\pi_{ij} - \frac{1}{2}\pi^2\right) = 16\pi T^*_*, \tag{14}$$

where S^i is the current density of the sources and T^*_* is the energy density[4]. As before we can choose $\tau =$ constant. In general, the source must also scale in a definite way when we make a conformal transformation. For example we shall here illustrate the case of a charge-free electromagnetic field, where it may be readily verified that the Poynting vector-density scales as $S^{i\prime} = \phi^{-4} S^i$ and the energy per unit proper volume scales $T^{*\prime}_* = \phi^{-8} T^*_*$. This behavior follows from that fact the electric and magnetic vectors *must* remain divergence-free under the transformation.

The momentum constraints remain independent of the Hamiltonian constraint, just as above, i.e., they can readily be solved in a conformally invariant manner. The scale equation[5] is now [IV][2]

$$8\nabla^2\phi = R\phi - M\phi^{-7} + \tfrac{3}{8}\tau^2\phi^5 - 16\pi T^*_*\phi^{-3}. \tag{15}$$

In this paper we will show that:

(1) A positive, bounded solution ϕ to the scale equation "almost always" exists, i.e., the initial data sets for which a solution does not exist occupy a "set of measure zero" in the function space of all initial data sets (that is, all choices of $g_{ij}, \sigma^{ij}, \tau$, and T^*_*). These exceptional cases correspond to physically unrealizable configurations of gravitational waves and matter.

(2) If a solution exists, it is unique. (The only exception is trivial; see Sec. 3.)

In the proofs of existence given in this paper, it has been assumed that the initial manifold is "compact." We are using the term "compact" in the sense usually intended by physicists. That is, we call a compact manifold *without* boundary "closed" to distinguish it from a compact manifold *with* boundary. By the latter term we mean that the boundary is included as part of the manifold and that the resulting structure is compact. Thus "compact" in this work refers to "closed" manifolds (boundary is empty) *or* to "nonclosed" compact manifolds (boundary is not empty). If the boundary is not empty, we have assumed the physically natural Dirichlet boundary condition, i.e., ϕ is specified on the boundary. This terminology agrees with that, of Choquet-Bruhat in Ref. 5. In Sec. 2 and Appendix B we deal with the proof of existence when the manifold is closed. In Appendix I we treat the case when the boundary is not empty.

The conformal method and its attendant physical implications[2] are also valid for asymptotically flat "open" initial manifolds, i.e., in the case where the boundary is "at infinity." In this case one requires that $\phi \to 1 + O(r^{-1})$ as $r \to \infty$. The coefficient of the $O(r^{-1})$ term is essentially the total mass of the system, assumed to be finite. The further technical requirements on the asymptotic behavior of the physical fields that are sufficient to establish our results concerning *existence* of solutions in this case are mentioned in Appendix A. The proof of *uniqueness* in Sec. 3 is valid as given for both compact and open manifolds.

2. EXISTENCE

The scale equation has the form

$$\nabla^2\phi = P(\phi, x) \tag{16}$$

on some manifold V, where $P(\phi, x)$ is a polynomial in ϕ whose coefficients may be position-dependent. The following general existence theorem may be proved about such equations[6]:

Theorem I: The equation $\nabla^2\phi = P(\phi, x)$ has a positive bounded solution ϕ on a closed manifold if there exist two positive constants ϕ_-, ϕ_+ ($\phi_- < \phi_+$) such that

$$\left.\begin{aligned} P(\phi_-, x) &< 0 \\ P(\phi_+, x) &> 0 \end{aligned}\right\} \forall x \in V. \tag{17}$$

The solution lies in the interval (ϕ_-, ϕ_+). We can prove this theorem in two independent ways:

(i) By using an iterative technique. This is shown in Appendix A.

(ii) By using Leray–Schauder degree theory. This is shown in Appendix B.

Here we will only offer an argument as to the reasonableness of this result. The existence of ϕ_-, ϕ_+ shows that for each point $x \in V$, $P(\phi, x)$ has one real positive root (or an odd number of such roots) lying in the interval (ϕ_-, ϕ_+). Let us restrict our attention to the case where $P(\phi, x)$ has one real root only, as this is sufficient for our purposes. Therefore, the theorem in essence says that if the polynomial has, for each $x \in V$, a root in some bounded, fixed interval (ϕ_-, ϕ_+), then Eq. (16) has a solution in the same interval.

Firstly, if $P(\phi, x)$ had *no* real root for any x then Eq. (16) on a closed manifold does not have any solution, because the nonexistence of a real root implies that $P(\phi, x)$ is either always positive or always negative. However,

$$\int_V \nabla^2\phi \, dv = 0 \tag{18}$$

since V is closed. Therefore,

$$\int_V P(\phi, x) dv = 0, \tag{19}$$

which is a contradiction. Returning to the case where P has one root, if $P(\phi, x)$ has a fixed root ϕ_0 for all $x \in V$, then obviously ϕ_0 is a solution to Eq. (16). Now, let $P(\phi, x)$, instead of having a fixed root, have a root in a small interval $\delta\phi$ around ϕ_0, as x varies over V. It seems reasonable to conclude that $\nabla^2\phi = P(\phi, x)$ again has a solution. Elementary considerations show that this solution must lie in the interval $\delta\phi$. Of course, this discussion is only meant to be suggestive. The treatment of the problem in the appendixes shows that this heuristic reasoning leads to a correct conclusion.

To apply Theorem I to the scale equation, we have to look at the polynomial

$$P(\phi, x) = -M\phi^{-7} - 16\pi T^*_*\phi^{-3} + R\phi + \tfrac{3}{8}\tau^2\phi^5 \tag{20}$$

to see whether or not it has a real, positive root in some fixed interval (ϕ_-, ϕ_+) as x varies over V. The roots of $P(\phi, x)$ coincide with those of $Q(\phi, x) = \phi^7 P(\phi, x)$, since we are only interested in $\phi > 0$. We have

$$Q(\phi, x) = -M - 16\pi T^*_*\phi^4 + R\phi^8 + \tfrac{3}{8}\tau^2\phi^{12}. \tag{21}$$

This polynomial is a cubic in ϕ^4. Let us investigate the simplest case $M > 0$, $\tau \neq 0$. An arbitrary cubic can have zero, one, two, or three positive roots. However, $Q(0, x) =$

$-M < 0$ and $Q(\phi, x)$ tends towards $+\frac{3}{8}\tau^2\phi^{12} > 0$ as ϕ becomes large, for all x. Therefore it can have only one or three positive roots. Say it has three positive roots α, β, γ; therefore

$$Q(\phi, x) = \tfrac{3}{8}\tau^2(\phi^4 - \alpha)(\phi^4 - \beta)(\phi^4 - \gamma). \tag{22}$$

The term in ϕ^4 is $\frac{3}{8}\tau^2(\alpha\beta + \beta\gamma + \gamma\alpha)\phi^4 > 0$. However, the term in ϕ^4 is $-16\pi T_*^*\phi^4 \leq 0$. Therefore, the polynomial has one and only one positive root. So long as M, T_*^*, R, and τ are bounded this root will lie in some positive, finite interval for all x. Therefore, so long as $M > 0$, $\tau \neq 0$, the scale equation always has a positive, bounded solution, irrespective of the behavior of T_*^* and R, on a closed manifold.

Since only the conformal equivalence class of the initial metric is significant, we have the choice of switching to any representation of the conformal geometry we please before attempting to solve the scale equation. In other words, any Riemannian metric conformally related to the original one may be used. This freedom is represented by the fact that the scale equation maintains the same form under such a transformation. Let $\bar{g}_{ij} = \theta^4 g_{ij}$. Then Eq. (15) becomes

$$8\bar{\nabla}^2\bar{\phi} = \bar{R}\bar{\phi} - \bar{M}\bar{\phi}^{-7} + \tfrac{3}{8}\tau^2\bar{\phi}^5 - 16\pi\bar{T}_*^*\bar{\phi}^{-3}, \tag{23}$$

$$\bar{\phi} = \phi\theta^{-1}; \bar{M} = M\theta^{-12}, \bar{T}_*^* = T_*^*\theta^{-8}; \bar{R} = R\theta^{-4} - 8\theta^{-5}\nabla^2\theta. \tag{24}$$

If for some particular choice of these variables, one can prove that a solution to the scale equation exists, then clearly, for any other conformally equivalent set of data $\bar{g}_{ij} = \theta^4 g_{ij}$, $\bar{M} = M\theta^{-12}$, $\bar{T}_*^* = T_*^*\theta^{-8}$, $\bar{\tau} = \tau$, a solution will also exist. We want to use this initial freedom of choosing the initial data in order to simplify the application of Theorem I to the scale equation in cases for which $M \geqslant 0$, rather than the stronger assumption $M > 0$ made above. For example, we can prove that if $M + T_*^*$ is not zero everywhere on V, and $\tau \neq 0$, then a solution to the scale equation must exist. The key idea is that, on a closed manifold, we can always find a conformal transformation that makes the scalar curvature negative on any given proper subset of the manifold, though not, of course, on the entire closed manifold. This is shown in Appendix C.

Since $M + T_*^*$ is not everywhere zero, suppose it vanishes only on a proper subset V_0 of V. Choose a conformal transformation θ that makes $\bar{R}$ negative on V_0. Then $P(\bar{\phi}, x)$ has a single, positive root for each $x \in V$, and so Theorem I shows that $\bar{\nabla}^2\bar{\phi} = P(\bar{\phi}, x)$ has a positive, bounded solution. Hence the scale equation has a solution in any conformally equivalent set of initial data, also.

If we consider yet more special cases, e.g., such that either $M + T_*^*$ or τ vanishes everywhere, we can no longer always prove that a solution exists. We can use the existence theorem, however, to place nontrivial necessary and sufficient conditions on the conformal intrinsic geometry for a solution to exist in these cases.

(i) If $\tau = 0$, i.e., the slice is "maximal," the scale equation has a solution if a conformal factor exists that transforms the given intrinsic geometry into one which has vanishing scalar curvature wherever $M + T_*^*$ vanishes, and has positive scalar curvature wherever $M + T_*^* > 0$. This follows from examining the roots of $P(\phi, x)$ in this case.

(ii) Let both M and T_*^* be zero everywhere on V. (a) If $\tau \neq 0$, the conformal intrinsic geometry must have a representation where $\bar{R} < 0$ everywhere, for a solution to exist. This is not true of every conformal equivalence class of Riemannian metrics.[7] (b) If $\tau = 0$, i.e., the slice is mass-free and time-symmetric, a solution in general does not exist.[7]

The existence theorem was only discussed for closed manifolds but it is also true for manifolds with a boundary, with Dirichlet boundary conditions (see Appendix I). The only added condition is that if ϕ_B is the boundary value, then we require that

$$\phi_- < \phi_B < \phi_+. \tag{25}$$

In the particular case of the scale equation, this is very easily accomplished and a solution exists "almost always" in both cases, i.e., except in very special circumstances such as noted in (i) and (ii). The "generic" case $\tau \neq 0$ and M and T_*^* not vanishing everywhere always leads to a solution for any conformal equivalence class of initial metrics.

3. UNIQUENESS

The existence theorem suggests that each root of $P(\phi, x)$ signals another solution to $\nabla^2\phi = P(\phi, x)$. In the case of the scale equation $P(\phi, x)$ has one and only one positive root, therefore we should not be too surprised at the following:

Theorem II: On a closed manifold, any positive bounded solution to the scale equation is unique except in the (trivial) case of the vacuum at a moment of time-symmetry: $M = T_*^* = \tau = 0$ everywhere on V.

Proof: Let us assume that we start off with a set of variables g_{ij}, σ^{ij}, T_*^*, τ, for which the scale equation $\nabla^2\phi = P(\phi, x)$ has a solution. Let us denote this solution by ϕ_S. Using ϕ_S as a conformal factor and transforming the initial data as in Eq. (24) we get, of course, a set of initial data that obeys the Hamiltonian constraint:

$$-\bar{M} - 16\pi\bar{T}_*^* + \bar{R} + \tfrac{3}{8}\tau^2 = 0. \tag{26}$$

Now, if the scale equation does not have a unique solution, then there must exist a positive, bounded ϕ, not identical to ϕ_S, which is also a solution of $\nabla^2\phi = P(\phi, x)$. Using ϕ_S as a conformal transformation this equation for ϕ becomes [see Eq. (23)]

$$8\bar{\nabla}^2\bar{\phi} = -\bar{M}\bar{\phi}^{-7} - 16\pi\bar{T}_*^*\bar{\phi}^{-3} + \bar{R}\bar{\phi} + \tfrac{3}{8}\tau^2\bar{\phi}^5. \tag{27}$$

If the uniqueness property of the scale equation does not hold, then Eq. (27) has a solution ϕ not identically equal to 1. We can prove that this does not happen by use of the following result.

Lemma 1: If $P(\phi, x)$ obeys the following conditions:

$$P(\phi, x) < 0 \text{ when } \phi < 1,$$
$$P(\phi, x) > 0 \text{ when } \phi > 1, \tag{28}$$

for each $x \in V$, where V is closed, then $\nabla^2\phi = P(\phi, x)$ has a unique solution $\phi = 1$.

Proof of Lemma: Let there exist a solution ϕ not identical to 1. Denote $V_- = \{x \in V \mid \phi(x) < 1\}$. The set V_- is either empty, identical to V, or a proper subset of V. It cannot be identical to V because then $P(\phi, x) < 0\ \forall\ x \in V$ and

$$\int_V \nabla^2\phi\, dv = 0 = \int_V P(\phi, x)\, dv$$

which is a contradiction.

If V_- is a proper subset of V, it is easy to show that $\phi(x) = 1 \; \forall \; x \in \partial \, V_-$ (∂ = boundary operator). Now we can write

$$\int_{V_-} \phi \nabla^2 \phi dv = \int_{\partial V_-} \phi(\nabla \phi) \cdot dS - \int_{V_-} (\nabla \phi)^2 dv$$
$$= \int_{\partial V_-} (\nabla \phi) \cdot dS - \int_{V_-} (\nabla \phi)^2 dv$$
$$= \int_{V_-} \nabla^2 \phi dv - \int_{V_-} (\nabla \phi)^2 dv$$

Therefore,

$$\int_{V_-} (\nabla \phi)^2 dv = \int_{V_-} (1 - \phi) \nabla^2 \phi \, dv = \int_{V_-} (1 - \phi) P(\phi, x) dv.$$

In V_-, $P(\phi, x) < 0$ and $(1 - \phi) > 0$. Therefore the left-hand side is positive and the right-hand side is negative, which is a contradiction. Therefore V_- is empty. A similar argument will lead to the conclusion $V_+ = \{x \in V | \phi(x) > 1\}$ is also empty. Therefore the only solution is $\phi \equiv 1$.

To return to the main theorem, it is clear that $P(\overline{\phi}, x)$ obeys the conditions of Lemma 1 so long as at each point $x \in V$ either $\overline{M} \neq 0$, or $\overline{T}^*_* \neq 0$, or $\tau \neq 0$.

To cover the cases in which *all* of these coefficients vanish on some subset V_2 of V, i.e., $P(\overline{\phi}, x) = 0, \forall \; \phi$, $\forall \; x \in V_2$, we need an additional result:

Lemma 2: If there exists a proper subset V_1 of a closed manifold V such that

$$\left.\begin{array}{l} P(\phi, x) < 0 \text{ when } \phi < 1 \\ P(\phi, x) > 0 \text{ when } \phi > 1 \end{array}\right\} \text{ for each } x \in V_1,$$

and

$$P(\phi, x) = 0, \forall \; \phi, \forall \; x \in V_2 = C(V_1), C = \text{complement}, \tag{29}$$

then $\nabla^2 \phi = P(\phi, x)$ has a unique solution $\phi = 1$.

In this case we can show, just as in Lemma 1, $V_1 \cap V_-$ is empty and $V_1 \cap V_+$ is empty. Therefore $\phi \equiv 1$ on V_1. Therefore $\phi \equiv 1$ on ∂V_1. However on $C(V_1)$ the equation becomes $\nabla^2 \phi = 0$ with boundary condition $\phi = 1$ on ∂V_2. The only bounded solution of this is $\phi = 1$ on V_2. Therefore, the equation has a unique solution. Applying Lemma 2 to our basic equation

$$8 \overline{\nabla}^2 \overline{\phi} = -\overline{M} \overline{\phi}^{-7} - 16\pi \overline{T}^*_* \overline{\phi}^{-3} + \overline{R} \overline{\phi} + \tfrac{3}{8} \tau^2 \overline{\phi}^5, \tag{27}$$

with

$$-\overline{M} - 16\pi \overline{T}^*_* + \overline{R} + \tfrac{3}{8} \tau^2 = 0, \tag{26}$$

we see that the only case in which either Lemma 1 or Lemma 2 does not hold is trivial one when $P(\overline{\phi}, x) = 0$ for every x in V. In this case, the vacuum, time-symmetric configuration, the scale equation reduces to $\overline{\nabla}^2 \overline{\phi} = 0$, and on a closed manifold, any positive constant is a solution.

This very strong uniqueness property carries over to the case where V has a boundary. In this case, since we have to specify the value of ϕ on the boundary, we seek a second solution to the scale equation different from $\phi = 1$, but with $\phi = 1$ on the boundary.[8] We require the additional condition that $\oint_{\partial V} \nabla \phi \cdot dS$ is finite. This is sufficient to permit

$$\oint_{\partial V} (\nabla \phi) \cdot dS = \oint_{\partial V} \phi (\nabla \phi) \cdot dS$$

since $\phi = 1$ on ∂V.

Then, using this in conjunction with

$$\int_V \phi \nabla^2 \phi dv = \oint_{\partial V} \phi \nabla \phi \cdot dS - \int_V (\nabla \phi)^2 dv,$$

we obtain

$$\int_V (\nabla \phi)^2 dv = \int_V (1 - \phi) P(\phi, x) dv.$$

The right-hand side is always negative, whereas the left-hand side is positive. In the case where V has a boundary, even a "boundary at ∞," we *always* have uniqueness, even when $M = 0$, $T^*_* = 0$, $\tau = 0$, because in this case the equation is $\nabla^2 \phi = 0$, $\phi = 1$ on ∂V, and the only bounded solution is $\phi = 1$.

The requirement that $\oint_{\partial V} \nabla \phi \cdot dS$ be finite is not purely a mathematical convenience but has physical content. Arnowitt, Deser, and Misner constructed an expression for the energy of an asymptotically flat gravitational field, which depends only on the behavior of the *intrinsic* 3-geometry at infinity.[9] If the 3-geometry is changed by a conformal transformation ϕ, the change in energy satisfies[10]

$$\Delta E = -\frac{1}{2\pi} \oint_{\partial V} (\nabla \phi) \cdot dS. \tag{30}$$

Therefore we have the following result: A solution to the scale equation on a manifold with a boundary is always unique when ΔE is finite.

APPENDIX A: AN ITERATIVE PROOF OF THE EXISTENCE THEOREM ON COMPACT MANIFOLDS[11]

In this appendix we will give a detailed proof of the existence of a solution to $\nabla^2 \phi = P(\phi, x)$ on a nonclosed compact manifold, with Dirichlet boundary conditions, i.e., we fix the value of ϕ on the boundary. Of course, we are especially interested also in the case of asymptotically flat *open* manifolds, i.e., ones with boundary "at infinity." To extend existence to this case, one needs conditions on the asymptotic behavior of the fields. (The result on uniqueness in Sec. 3 is valid in the case of either compact or open manifolds.) These additional requirements amount essentially to the statement that the total mass of the system of gravitational waves, matter, and other fields is finite. Of course, in this case one would not use τ = constant, but $\tau = 0$ would still be quite useful. Nonconstant choices of τ would also be valid, with $\tau \sim 0(r^{-2})$ as $r \to \infty$ being quite sufficient as regards asymptotic behavior. (The method of solution of the momentum constraints for nonconstant τ is dealt with in IV[2] and is treated in more detail in a forthcoming paper. We shall not present here the details of the asymptotic behavior of all the relevant variables since the mathematically sufficient conditions turn out to be the physically natural ones discussed by Arnowitt, Deser, and Misner[12] who used, however, techniques different from the present ones. Also, a recent mathematically rigorous discussion of asymptotic behavior in the context of perturbation theory on open manifolds has been given by Choquet-Bruhat and Deser.[13] Similar considerations should prove to be quite adequate for the exact nonperturbative techniques employed here and in IV[2]. The authors would like to thank Professor Choquet-Bruhat for discussions on this subject.

We proceed to discuss the case of a compact manifold with boundary. The present existence theorem could be generalized to the open case by considering a sequence of compact manifolds whose boundaries tend towards infinity. If a solution exists for each compact manifold, and if all the functions are well behaved in terms of the considerations mentioned above, then we expect a solution to exist in the open case also. The solution will be unique. An alternative treatment would be to apply the compactification technique of Geroch.[14]

Theorem III: $\nabla^2\phi = P(\phi, x)$ has a positive, bounded solution on a compact manifold V, which has the positive value ϕ_B on ∂V, when two positive constant ϕ_-, ϕ_+ exist, such that

$$0 < \phi_- < \phi_B < \phi_+ \tag{A1}$$

and

$$\left.\begin{array}{l} P(\phi_-, x) < 0 \\ P(\phi_+, x) > 0 \end{array}\right\} \forall\ x \in V. \tag{A2}$$

The solution is bounded by ϕ_-, ϕ_+.

Proof: It is convenient to introduce new variables:

$$\theta = \phi - \frac{1}{2}(\phi_+ + \phi_-), \tag{A3}$$

$$\tilde{\theta} = \phi_B - \frac{1}{2}(\phi_+ + \phi_-), \tag{A4}$$

$$\alpha = \frac{1}{2}(\phi_+ - \phi_-). \tag{A5}$$

Substituting these variables into the original equation we get

$$\nabla^2\theta = Q(\theta, x) \tag{A6}$$

with $\theta = \tilde{\theta}$ on ∂V, and also

$$\left.\begin{array}{l} Q(\alpha, x) > 0 \\ Q(-\alpha, x) < 0 \end{array}\right\} \forall\ x \in V, \tag{A7}$$

$$|\tilde{\theta}| < \alpha. \tag{A8}$$

Now we will construct a sequence of functions θ_m, which as m goes towards infinity, tends towards a solution of Eq. (A6). Each member of this sequence will have $\tilde{\theta}$ as boundary value, and will lie in the interval $(-\alpha, \alpha)$.

Define

$$k = \max_{-\alpha<\theta<\alpha} \left|\frac{d}{d\theta} Q(\theta, x)\right|. \tag{A9}$$

Consider the sequence of functions θ_m (with $\theta_1 = \alpha$) given by

$$\nabla^2\theta_{m+1} - k\theta_{m+1} = Q(\theta_m, x) - k\theta_m, \tag{A10}$$

with $\theta_m = \theta_{m+1} = \tilde{\theta}$ on ∂V. Since $k > 0$, this is a well-defined sequence, because given θ_m, the theory of linear elliptic equations shows that θ_{m+1} uniquely exists. In addition, the maximum principle shows that

$$|\theta_{m+1}| \le \max |\tilde{\theta}, k^{-1}[Q(\theta_m, x) - k\theta_m]|. \tag{A11}$$

Firstly,

$$\nabla^2\theta_2 - k\theta_2 = Q(\alpha, x) - k\alpha. \tag{A12}$$

Now Eq. (A11) shows

$$|\theta_2| \le \max |\tilde{\theta}, k^{-1}[Q(\alpha, x) - k\alpha]|.$$

Since $|\tilde{\theta}| < \alpha, Q(\alpha, x) > 0$,

$$-\alpha \le \theta_2 \le \theta_1 = \alpha. \tag{A13}$$

Now try induction. Assume θ_m given, $-\alpha \le \theta_m \le \theta_{m-1} \le \alpha$, where θ_m is a solution of

$$\nabla^2\theta_m - k\theta_m = Q(\theta_{m-1}, x) - k\theta_{m-1}. \tag{A14}$$

Consider θ_{m+1} as a solution of

$$\nabla^2\theta_{m+1} - k\theta_{m+1} = Q(\theta_m, x) - k\theta_m. \tag{A10}$$

Subtract (A10) from (A14):

$$\nabla^2(\theta_m - \theta_{m+1}) - k(\theta_m - \theta_{m+1}) = Q(\theta_{m-1}, x) - Q(\theta_m, x) - k(\theta_{m-1} - \theta_m) \le 0. \tag{A15}$$

Since $k > 0$, we can use the maximum principle, which gives $\theta_m - \theta_{m+1} \le 0$ since $\theta_m - \theta_{m+1} \to 0$ on ∂V.

In addition, it is easy to show

$$|k^{-1}[Q(\theta_m, x) - k\theta_m]| \le \alpha,\ |\tilde{\theta}| \le \alpha.$$

Therefore, Eq. (A11) shows $|\theta_{m+1}| \le \alpha$.

Therefore,

$$-\alpha \le \theta_{m+1} \le \theta_m \le \theta_{m-1} \le \ldots \le \alpha.$$

This is a bounded sequence. It is demonstrated in Courant and Hilbert[15] that this sequence has a limit θ. Obviously θ is the function we require. The existence of ϕ_-, ϕ_+ and the relevant properties of $P(\phi, x)$ have already been discussed in the text.

APPENDIX B: LERAY-SCHAUDER DEGREE THEORY AND THE EXISTENCE THEOREM

The Leray–Schauder degree[16] is an extension of the concept of the local degree of a mapping to infinite-dimensional function spaces. The local degree of a mapping θ is a function of the subset D of its domain and of a point P of its range. There are many equivalent definitions. One of them is

$$d(\theta, D, P) = \sum_{x \in \theta^{-1}(P) \cap D} \operatorname{sgn} \det J_\theta(x). \tag{B1}$$

J_θ is the Jacobian of the mapping. We require $J_\theta \ne 0$. The local degree has two properties of interest: (i) an existence theorem: if $d(\theta, D, P) \ne 0$, then $\exists\ q \in D$ such that $\theta(q) = P$, i.e., P has a pre-image in D. (ii) The local degree is an invariant measure of the number of pre-images that P has under θ in D. It is unchanged by a smooth change of θ, or D, or P, except when a point on the boundary of D gets mapped into P.

Leray and Schauder showed that an equivalent concept could be defined for compact transformations between Banach spaces. Both the existence and invariance properties carry over from the local degree to the Leray–Schauder degree. As a preliminary, let us recall a number of definitions: A Banach space is a normed linear space that is complete in the norm. A compact space is one in which every infinite sequence has a limit point. A compact transformation T is a continuous transformation with the added property that $T(M)$ is compact for every bounded set M. Let T_t be a family of compact transformations, $t \in [0,1]$, then T is a homotopy if T_t is uniformly continuous in t.

Basic Theorem of Leray and Schauder: Let T_t be a homotopy of compact transformations between a Banach space $\mathcal{B}$ and itself. Let E be a subset of $\mathcal{B}$. Then if $(I - T_t)(x) \neq \overline{O}\ \forall\ t \in [0,1], \forall\ x \in \partial E$ where I is the identity map and $\overline{O}$ is the origin of $\mathcal{B}$, then $d[I - T_t, E, O]$ exists and has the same value for each $t \in [0,1]$. The proof of this theorem may be found in Ref. 16.

Application to Theorem I: Consider the Banach space $\mathcal{B}$ of all C^3 functions on V. Choose any number ϕ_0 such that $\phi_- < \phi_0 < \phi_+$. Now consider the differential equation

$$\nabla^2 U - U = -\phi_0 + t[P(v,x) - v + \phi_0], \tag{B2}$$

where $v \in \mathcal{B}, t \in [0,1]$. This equation is to be solved for $U \in \mathcal{B}$. Given $(v,t) \in \mathcal{B}\ x[0,1]$, the right-hand side of Eq. (B2) becomes a well-defined C^3 function of position $f_{v,t}(x)$, so long as the initial data are C^3 functions also. Therefore, Eq. (B2) reduces to

$$\nabla^2 U - U = f_{v,t}(x). \tag{B3}$$

It immediately follows from the theory of linear elliptic equations that this equation has a unique solution. In addition, the maximum principle shows that

$$\max_V |U| \leqslant \max_V |f(x)|. \tag{B4}$$

Therefore Eq. (B2) is a transformation of the form $\mathcal{B} \times [0,1] \to \mathcal{B}, T_t(V) = U$. We wish to apply the basic theorem to this transformation. Firstly Eq. (B4) shows that T is a continuous transformation. The requirement that everything is C^3 is more than enough to guarantee compactness.[17] Finally we need to show that T_t is uniformly continuous in t. Change variables to $U_0 = U - \phi_0$. The equation for U_0 becomes

$$\nabla^2 U_0 - U_0 = t[P(v,x) - V + \phi_0]. \tag{B5}$$

Define

$$A = \max_{\substack{v \in E \\ x \in V}} |P(v,x) - V + \phi_0| < \infty. \tag{B6}$$

Given ϵ, define $\delta = \epsilon/A$. Then if $|t_1 - t_2| < \delta$,

$$\max_{x \in V} |U_0(v,t_1) - U_0(v,t_2)| < \epsilon$$

for any $v \in E$. This is because $U_0(v,t) = tU_0(v,1)$, and $|U(v,1)| \leqslant A$. Therefore, if $|t_1 - t_2| < \delta$, then $|T_{t1}(v) - T_{t2}(v)| < \epsilon$. Therefore T is uniformly continuous in t.

To make use of the basic theorem we have to select some $E \subset \mathcal{B}$. We choose E to be the set of all C^3 functions on V, bounded by ϕ_- and ϕ_+. It is easy to show that if $v \equiv \phi_-$ or ϕ_+, $(I - T_t)v \neq \overline{O}$ for any $t \in [0,1]$. This depends on the fact that $P(\phi_-, x) < 0$, $P(\phi_+, x) > 0$. Therefore $d[I - T_t, E, \overline{O}]$ is well defined and $d[I - T_1, E, \overline{O}] = d[I - T_0, E, \overline{O}]$.

The map T_0 takes every point of E into a single point, the constant function ϕ_0. Therefore $I - T_0$ takes one and only one point of E into $\overline{O}$; that point is again the constant function ϕ_0. It is easy to show that $d[I - T_0, E, \overline{O}] = 1$. Therefore,

$$d[I - T_1, E, \overline{O}] = +1. \tag{B7}$$

Hence, there exists a function $\phi \epsilon E$, which is mapped by $I - T_1$ into $\overline{O}$. It immediately follows that this function obeys the equation $\nabla^2\phi = P(\phi, x)$.

APPENDIX C: SIGN CHANGES IN THE SCALAR CURVATURE UNDER CONFORMAL TRANSFORMATIONS

It is impossible to conformally relate a metric whose scalar curvature is everywhere positive to one whose scalar curvature is everywhere negative on a closed manifold. On the other hand, it is always possible to find a conformal transformation that will make the scalar curvature negative on any proper subset V_0 of V. Firstly, it is always possible to change the scalar curvature by using a constant conformal factor α, i.e.,

$$R' = \alpha^{-4} R \quad (\nabla^2 \alpha = 0). \tag{C1}$$

We can choose α large enough so that $|R'| < \frac{1}{2}$ on V_0, i.e., choose $\alpha^4 > 2 \max |R|$ on V_0.

Now in this new metric solve

$$8\nabla'^2 U - U = 0 \tag{C2}$$

on V_0, with $U = 1$ on ∂V_0. The theory of linear elliptic equations shows that this equation has a strictly positive solution U. Choose as conformal factor ϕ' any strictly positive, bounded function that matches U on V_0. Then,

$$R'' = (\phi')^{-4} R' - 8(\phi')^{-5} \nabla'^2 \phi'. \tag{C3}$$

But on V_0, $8\nabla'^2\phi' = \phi'$. Therefore, on V_0, $R'' = (\phi')^{-4}(R' - 1)$. Since $R' < \frac{1}{2}$ on V_0, $R'' < 0$ on V_0.

*Based in part on N. O'Murchadha, "Existence and Uniqueness of Solutions to the Hamiltonian Constraint of General Relativity," doctoral thesis, Princeton University, February 1973.

†Supported in part by National Science Foundation Grant GP30799X to Princeton University.

‡Address for 1973–74: Department of Physics, The University of North Carolina, Chapel Hill, N. C. 27514.

§Permanent address beginning August 1973: Department of Physics, The University of North Carolina, Chapel Hill, N. C. 27514.

[1]It is known from the work of A. Lichnerowicz and Y. Choquet-Bruhat that once a complete set of initial data compatible with the constraints has been obtained, the time evolution of this data, and therefore the structure of spacetime, is uniquely determined for a finite time. See, for example, the article by Y. Choquet-Bruhat in *Gravitation*, edited by L. Witten (Wiley, New York, 1962). The heretofore unresolved issue is then to determine which part of the initial data is specifiable in such a way that the rest of the initial data may be constructed using the equations of constraint. The authors thank Professor Choquet-Bruhat for reading this paper.

[2]J. W. York, Phys. Rev. Lett. **26**, 1656 (1971); J. Math. Phys. **13**, 125 (1972); Phys. Rev. Lett. **28**, 1082 (1972); J. Math. Phys. **14**, 456 (1973). These papers are referred to as I, II, III, and IV, respectively.

[3]The form of the scale equation [(12) or (15)] remaines the same where or not τ = constant. Therefore, our results include the cases $\tau = \tau(x)$, a prescribed function. See Ref. 2, (IV), for a discussion

[4]In the right hand sides of (13) and (14), π denotes the number 3.14 . . . , not the trace of π^{ij}!

[5]A scale equation for the case $\tau = 0$, $T^{\cdot\cdot}_{\cdot\cdot} \neq 0$, was proposed by Y. Choquet-Bruhat, C.R. Acad. Sci. A **27**, 682 (1972). However, the energy density $T^{\cdot\cdot}_{\cdot\cdot}$ was treated as a fully prescribed function rather than as a quantity that must undergo a conformal transformation, as mentioned in the text. That equation did not lead to the strong

existence and uniqueness properties of (15).

[6]General existence theorems for wide classes of quasilinear elliptic equations are proved by Y. Choquet-Bruhat and J. Leray, C. R. Acad. Sci. (Paris) A **274**, 81 (1972).

[7]See the discussion of Yamabe's "theorem" in Ref. 2 (IV).

[8]Of course, ϕ could be chosen to have any positive value at infinity. However, the normalization $\phi = 1$ at infinity leads to the physical interpretation of ϕ as a generalized "Newtonian" potential defining the total mass of an isolated system. This interpretation was given in the time-symmetric, asymptotically flat case by D. R. Brill, Ann. Phys. (N.Y.) **7**, 466 (1959).

[9]R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **118**, 1100 (1960).

[10]N. O'Murchadha and J. W. York, "Gravitational Energy" (to be published).

[11]The authors thank Professor Jerry L. Kazdan for suggesting that an iterative technique would be feasible.

[12]R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **121**, 1556 (1960); Phys. Rev. **122**, 997 (1961).

[13]Y. Choquet-Bruhat and S. Deser, "On the Stability of Flat Space," Ann. Phys. N. Y. (to be published).

[14]R. Geroch, J. Math. Phys. **13**, 956 (1972), has given a one-point conformal compactification of spacelike infinity.

[15]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. II, p. 371.

[16]Leray–Schauder degree theory is discussed at length in J. Cronin-Scanlon, *Fixed Points and Topological Degree in Nonlinear Analysis* (American Mathematical Society, Providence, R. I., 1964). Also very helpful are O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and Quasi-linear Elliptic Equations* (Academic, New York, 1968), C. Miranda, *Partial Differential Equations of Elliptic Type* (Springer, Berlin, 1970), and M. Berger and M. Berger, *Perspectives in Nonlinearity* (Benjamin, New York, 1970).

[17]We technically require only that all the variables be Hölder continuous of order $C^{2,\alpha}(\alpha < 1)$. This is a special type of continuity that lies between C^2 and C^3. It is discussed in detail in O. A. Ladyzhenskaya and N. N. Ural'tseva (Ref. 16).

Continuum ambiguity in the construction of unitary analytic amplitudes from fixed-energy-scattering data*

G. R. Bart

Mayfair College, Chicago, Illinois 60630

P. W. Johnson and R. L. Warnock

Illinois Institute of Technology, Chicago, Illinois 60616
(Received 25 April 1973)

The problem of determining the scattering amplitude for a given fixed-energy elastic differential cross section is discussed in the spinless case. We show that when the energy is above the inelastic threshold, one may construct an infinte family of unitary scattering amplitudes, by appropriate variation of the elasticity parameters. These amplitudes are analytic in the cosine of the scattering angle throughout the Lehmann ellipse, and all correspond to the same cross section. Hence, even if the cross section is known exactly, there are infinitely many sets of phase shifts. Similar results have been obtained in earlier work, under conditions (on the cross section and elasticities) which seem to be physically unrealistic. In the present paper, the outstanding unrealistic assumptions are avoided. In particular, a finite number of zeros of the dispersive part are now allowed. Each zero reduces the continuum ambiguity by one elasticity parameter, but leaves infinitely many parameters to be varied independently.

1. INTRODUCTION

The problem of phase-shift analysis may be idealized as follows. Take the case of spinless particles, and suppose that the differential cross section is known exactly at a fixed energy. The latter is equivalent to knowing the modulus $g(z)$ of the scattering amplitude $f(z) = g(z)e^{i\phi(z)}$ (here, $z = \cos\theta$ is the cosine of the scattering angle, and the energy variable is suppressed). The task of phase-shift analysis is to determine the phase $\phi(z)$ by imposing the unitarity condition. Unitarity is imposed by means of the partial-wave expansion

$$f(z) = \sum_l (2l+1) f_l P_l(z), \tag{1.1}$$

$$f_l = (\eta_l e^{2i\delta_l} - 1)/2i \tag{1.2}$$

In a phase-shift analysis, η_l and δ_l are determined so as to give the experimental value of $g(z) = |f(z)|$. The determination of η_l and δ_l is subject to the constraint of the optical theorem, and the unitarity requirement, $0 \le \eta_l \le 1$.

The actual problem of energy-independent phase-shift analysis is more involved than this idealization, since one usually has particles with spin, and there are experimental errors to contend with. It is, nevertheless, important to study the idealized case, since the ambiguities of the actual problem are expected to be at least as severe as those encountered here.

In Ref. 1 it was shown that there is a continuum ambiguity in phase-shift analysis, provided the energy is above the particle production threshold, and that $g(z)$ and the elasticities obey certain bounds. In other words, there are infinitely many choices of the δ_l and η_l, which are compatible with a given differential cross section and unitarity, when the conditions of Ref. 1 are satisfied. The reason that this situation is normally overlooked is that one habitually truncates the partial-wave series (1.1) at a finite number of terms. Consequently, the chi-squared of the phase-shift fitting program has a finite number of minima, and one gets the false impression that any ambiguity is at worst finite-dimensional. The existence of a continuum ambiguity depends essentially on there being an infinite number of partial waves,[2] Of course, a model with a finite number of waves is unphysical, since it violates basic requirements of analyticity and crossing symmetry.

The bounds on the cross section and the elasticities required in Ref. 1 may not be physically realistic. For instance, if there is a deep and sharp dip in the cross section, or too much inelasticity, the bounds will be violated. The bounds of Ref. 1 are artifically restrictive, however, and have more to do with limitations of mathematical technique than with intrinsic properties of the unitarity equation. Indeed it was pointed out that the solution of the unitarity equation obtained in Ref. 1 is a continuous function of the η_l in some region surrounding their original values. The extent of this region was underestimated in Ref. 1, perhaps by a substantial amount, since one knows (by the implicit function theorem) that the solution of the equation varies continuously when the η_l are varied, until a singularity of the unitarity equation is encountered. By presence of a singularity of the equation, we mean that the Fréchet derivative (the infinite-dimensional Jacobian) of an appropriate nonlinear integral operator, defined through the unitarity equation, does not possess an inverse. That is, the continuum ambiguity extends at least to the first singularity of the equation. It then becomes important to understand the possible singularities, if one is to deal with the continuum ambiguity in practical problems of phase shift analysis.

The object of the present work is to generalize the discussion of Ref. 1 in such a way that unphysical assumptions are removed. We ask if the elasticities of a *given* unitary scattering amplitude can be varied continuously from their initial values, in such a way that the cross section remains the same and unitarity is maintained. The given amplitude could be the result of a phase shift analysis, so this is a question of practical importance. In previous work we did not assume that an amplitude was given, but rather gave sufficient conditions [on $g(z)$ and on elasticities] for a unitary amplitude to exist. These sufficient conditions were too restrictive to cover the majority of realistic cases.

As was noted above, the main task in removing the restrictions of our previous method is to study the singular points of the unitarity equation. One such singular point occurs when $\cos\phi(z)$ has a zero. We given special attention to this case, which has not been dealt with

 Copyright © 1973 by the American Institute of Physics

heretofore, and which is expected to arise frequently in practice.

In conjunction with considering the removal of unphysical restrictions, one naturally asks whether the continuum ambiguity remains when other physically realistic constraints such as analyticity are imposed. It is generally believed that amplitudes for strong interactions should be analytic in a region of the z plane which includes the Lehmann ellipse.[3] This analyticity is intimately related to the short range nature of the strong interactions. In Ref. 1, $g(z)$ and $\phi(z)$ were required to be continuous but not necessarily analytic. Atkinson, Mahoux, and Yndurain[4] have found that the requirement of analyticity in an ellipse of the z plane does not remove the ambiguity. The conditions under which their proof holds are similar in restrictiveness to those of Ref. 1; in particular, zeros of $\cos\phi(z)$ are not allowed. In the present study, analyticity of the amplitude is incorporated.

The conclusion of our analysis, now valid under assumptions which seem quite acceptable from the physical point of view, is that phase-shift analysis in the inelastic region is subject to an ambiguity of a serious nature. Namely, even with exact scattering data, the amplitude resulting from a particular phase-shift search is merely one of a continuous infinity of acceptable amplitudes, each satsifying the condition of unitarity and analyticity in $\cos\theta$. As was mentioned above, this ambiguity appears only when all partial waves are included. It should not be confused with the ambiguity noted by Gersten,[5] which becomes at worst a discrete ambiguity when the constraint of the optical theorem is met. Our ambiguity is not removed by the optical theorem.

How, exactly, does the continuum ambiguity relate to the practical problem of phase-shift analysis? Suppose that $f(z)$ is an amplitude obtained from a particular phase-shift analysis. The amplitude might be composed of a finite number of partial waves, as in a traditional phase-shift analysis, or it might contain an infinite number of waves (but depend on only a finite number of parameters), as in a Cutkosky–Deo analysis.[2] Then one may construct explicitly any number of additional unitary analytic amplitudes which will fit the same data. The construction is done by solving a certain differential equation, Eq. (3.18). This equation may be solved numerically by standard methods for initial value problems. It should be interesting to solve the equation in specific cases, in order to determine the severity of the ambiguity in practical situations. Of course, the present treatment does not apply to all cases of interest in that spin is omitted, but preliminary studies indicate that the ambiguity is equally serious when spin is included. As in the spinless case, one may formulate a differential equation to compute arbitrarily many amplitudes to fit given data.

In Sec. 2 we obtain theorems on the existence of unitary analytic amplitudes having given cross sections and given elasticities and a theorem on existence of the continuum ambiguity. This discussion, which is in the spirit of Refs. 1 and 4, is preliminary to the business at hand. We include it because it introduces our mathematical method, and shows how the proofs of Ref. 1 are generalized to the case of analytic functions. Our technique is different from that of Ref. 4.

In Sec. 3 we are concerned with the nature of the continuum ambiguity when $\cos\phi(z)$ has a zero. We assume existence of an amplitude with one such zero, and apply the implicit function theorem to obtain the behavior of ϕ when elasticities are varied. When only one of the elasticities η_l is varied, the Jacobian of the system (more exactly, the Fréchet derivative with respect to ϕ) is singular, and the implicit function theorem cannot be applied. If two or more elasticities are allowed to vary, then the problem becomes nonsingular, but only one of the two elasticities may be varied independently. The dependent elasticity, along with ϕ, is obtained as a function of the independent one through a differential equation. In fact, the effect of each zero of $\cos\phi$ is to reduce the continuum ambiguity by one dimension. Infinitely many of the η_l remain to be varied at will.

The discussion of Sec. 3 is carried out under the assumption that $g(z)$ has no zero in the ellipse of analyticity. In Secs. 4 and 5 we give alternative methods which avoid this assumption, while giving results similar to those of Sec. 3.

2. EXISTENCE OF UNITARY ANALYTIC AMPLITUDES WITH A GIVEN CROSS SECTION

Suppose that $f(z)$ is a scattering amplitude which is analytic in the open domain S bounded by the unifocal ellipse (foci at ± 1) with semimajor axis of length z_0. The function

$$\sigma(z) = f^*(z^*)f(z) \tag{2.1}$$

is analytic in S, and is equal to $k\,d\sigma/d\Omega$ for $-1 \le z \le 1$, where k is the barycentric momentum and $d\sigma/d\Omega$ the differential cross section. If experimental values of the differential cross section are known, one may always construct a function $\sigma(z)$ which is analytic in S and which fits the experimental data. Consequently, we regard the analytic function $\sigma(z)$ as given, and seek to find a function $f(z)$, analytic in S, which obeys (2.1) and the unitarity condition. For the work of this and the following section, we require that $\sigma(z)$ be free of zeros in S. One may then define $g(z) = [\sigma(z)]^{1/2}$, and define an analytic phase function $\phi(z)$ by

$$f(z) = g(z)e^{i\phi(z)}. \tag{2.2}$$

The unitarity condition, stated for the physical region $-1 \le z \le 1$, is

$$\sin\phi(z) = \frac{1}{4\pi}\int_{-1}^{1} du \int_0^{2\pi} dw \frac{g(u)g(\zeta)}{g(z)}\, e^{i[\phi(\zeta)-\phi(u)]} + \frac{I(z;\eta)}{g(z)}, \tag{2.3}$$

where

$$\zeta = zu + [(1-z^2)(1-u^2)]^{1/2}\cos w \tag{2.4}$$

and

$$I(z;\eta) = \sum_{l=0}^{\infty}(2l+1)\left(\frac{1-\eta_l^2}{4}\right)P_l(z). \tag{2.5}$$

The series (2.5) is assumed to be uniformly convergent in S, so that $I(z)$ is analytic in S. When g and I are given, Eq. (2.3) may be regarded as an integral equation for $\phi(z)$, $-1 \le z \le 1$. Continuous solutions of this equation were discussed in Ref. 1. We shall now prove that an analytic solution of (2.3) exists by applying a fixed point theorem in a certain Banach space $\mathcal{C}$ of analytic functions. Let $T \subset S$ be the interior of a unifocal ellipse of semimajor axis $z_0' < z_0$. We define the Banach space $\mathcal{C}$ of functions $\psi(z)$, defined and analytic in T, real on the real axis, and such that

$$\|\psi\| = \sup_{z\in T}|\psi(z)| < \infty. \tag{2.6}$$

Note that the scalars of the Banach space are required to be real numbers. For the proof that $\mathcal{C}$ is complete with respect to the norm (2.6), see Ref. 6, Vol. 1, Theorem 7.10.2, p. 193.

We abbreviate (2.3) as

$$\sin\phi = B(\phi) + g^{-1}I(\eta) \tag{2.7}$$

and consider B and $g^{-1}I$ at complex z. We show that the integral operator B maps the space $\mathcal{C}$ into itself. Let ϕ belong to $\mathcal{C}$; then $f = ge^{i\phi}$ is analytic in T, and the Legendre series converges in T. Now choose a $z \in T$ in Eq. (2.3), and note that the Legendre series for the functions $f^*(u)$ and $f(\zeta)$ which occur in (2.3) converge throughout the region of integration.[7] We may then integrate term by term, and use the addition theorem for Legendre polynomials to obtain

$$B(\phi) = \frac{1}{g(z)} \sum_{l=0}^{\infty} (2l+1)|f_l|^2 P_l(z), \tag{2.8}$$

where

$$f_l(\phi) = \tfrac{1}{2}\int_{-1}^{1} dx\, P_l(x) g(x) e^{i\phi(x)}. \tag{2.9}$$

We may employ the Cauchy representation of $ge^{i\phi}$, with a contour ∂E which is the boundary of an open unifocal elliptical domain E of semimajor axis $z_0'' < z_0'$. Upon introducing the Cauchy representation in (2.9) and reversing integration order, we obtain

$$|f_l(\phi)| \le \frac{1}{2\pi}\int_{\partial E} dz\, |Q_l(z)||g(z)|e^{|\mathrm{Im}\phi(z)|}, \tag{2.10}$$

where Q_l is the Legendre function of the second kind. A standard bound on the Q_l yields[8]

$$|f_l(\phi)| \le Me^{-\Delta l} \sup_{z\in\partial E} |g(z)| e^{\|\phi\|}, \tag{2.11}$$

$$\Delta = \ln[z_0'' + (z_0''^2 - 1)^{1/2}]. \tag{2.12}$$

The bound (2.11) implies that the series (2.8) converges in an open ellipse of semimajor axis $2z_0''^2 - 1$, and consequently represents an analytic function in that region. Since we may choose $2z_0''^2 - 1 > z_0$, the function B is analytic in S, and hence analytic and bounded in T.

Note that, whereas the integrand of the representation for B in (2.3) does involve a branch cut in z from -1 to 1, the integral does not. The estimate of the Legendre sum (2.8) proves that this branch cut is absent from B.

In order to apply Schauder's fixed point theorem,[9] we define V, a closed convex subset of $\mathcal{C}$:

$$V = \{\phi : \phi \in \mathcal{C}, \sup_{z\in T}|\sin\phi(z)| \le b < 1\}. \tag{2.13}$$

To see that V is convex, define $x = 2\mathrm{Re}\phi$, $y = 2\mathrm{Im}\phi$. Then the inequality which specifies V is

$$\cosh y - \cos x \le 2b^2. \tag{2.14}$$

In terms of the variables $\cosh y$ and $\cos x$, this region is a triangle. By checking signs of derivatives, one can show that the corresponding region of the x-y plane is convex. The region is inscribed in the rectangle

$$\sin|\mathrm{Re}\phi| \le b, \qquad \sinh|\mathrm{Im}\phi| \le b. \tag{2.15}$$

A graph of its boundary is shown in Fig. 1, for b slightly less than 1.

We apply Schauder's theorem to the mapping A, defined by

$$A(\phi) = \sin^{-1}[B(\phi) + g^{-1}I(\eta)]. \tag{2.16}$$

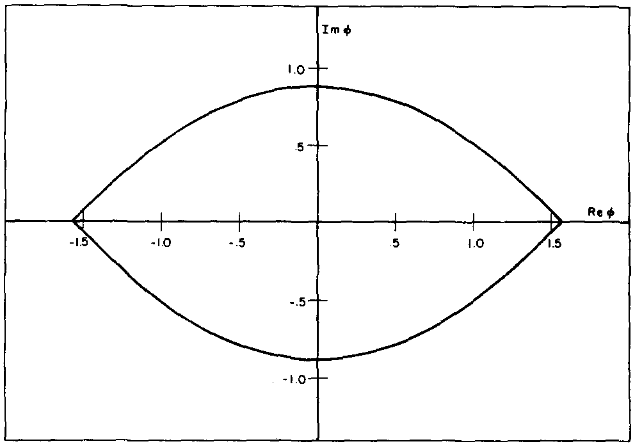

Fig. 1 The boundary of the domain specified by the inequality (2.14), for a value of b slightly less than 1.

The theorem asserts that if A is a continuous mapping of V into a compact subset of itself, then there is at least one solution in V of the unitarity equation $\phi = A(\phi)$. We first show that A maps V into itself provided b and $\sup|g^{-1}I|$ are sufficiently small. If $\phi \in V$, it follows from (2.3) and (2.15) that

$$|B(\phi) + g^{-1}I(\eta)| \le J\chi(b) + K, \tag{2.17}$$

where

$$J = \sup_{z\in T}\left|\frac{1}{4\pi}\int_{-1}^{1} du \int_0^{2\pi} dw\, \frac{g(u)g(\zeta)}{g(z)}\right|, \tag{2.18}$$

$$K = \sup_{z\in T} |g^{-1}(z)I(z;\eta)|, \tag{2.19}$$

and

$$\chi(b) = \exp\left(\sup_{z\in T}|\mathrm{Im}\phi(z)|\right) \le \exp(\sinh^{-1}b)$$
$$= b + (1+b^2)^{1/2}. \tag{2.20}$$

Let us restrict g and I to satisfy the condition

$$J[b + (1+b^2)^{1/2}] + K \le b. \tag{2.21}$$

We choose the sheet of the inverse sine in (2.15) so that $\sin^{-1}0 = 0$. Since $b < 1$, (2.19) implies that $A(z;\phi)$ is analytic in T and $A(V) \subset V$.

The continuity of $A(\phi)$ is obvious, and it remains only to show that $A(V)$ is compact. The compactness follows immediately from the following[10]:

Theorem: Let $\{\psi_n(z)\}$ be a sequence of functions analytic in a domain S, such that

$$|\psi_n(z)| \le M(U)$$

on every compact subset U of S. Then there is a subsequence $\{\psi_{n_k}(z)\}$ which converges uniformly on every compact subset of S to a function which is analytic in S.

Let $\{\psi_n(z)\}$ be a sequence of functions in V, and let $\psi_n(z) = A(z;\phi_n)$. The functions $\psi_n(z)$ are analytic in the domain S, and $|\psi_n(z)| \le M$, for $z \in U$, where U is any compact subset of S (in fact, the functions are even uniformly bounded on any compact subset of the bigger ellipse of semimajor axis $2z_0''^2 - 1$). According to the theorem, there is then a subseqence $\{\psi_{n_k}(z)\}$ which converges uniformly on $\bar{T}$ (the closure of T) to a function

$\psi(z)$ which is analytic in S. It is easy to see that $\psi \in V$, and that $\|\psi_{n_k} - \psi\|$ tends to zero. Hence, $A(V)$ is compact. We conclude from Schauder's theorem that there is at least one solution of the unitarity equation in the set V, if I and g obey the condition (2.21).

To clarify the restriction on J and K implied by (2.21), we write it as

$$J[1 + (1 + 1/b^2)^{1/2}] + K/b \le 1. \tag{2.22}$$

Since the left side of (2.2) is monotonically decreasing with b, the inequality can be satisfied for some $b < 1$ if, and only if,

$$J(1 + 2^{1/2}) + K < 1. \tag{2.23}$$

For the real-variable formulation of Ref. 1, quantities analogous to J and K were defined as suprema over $[-1, 1]$. If one replaces suprema over T by suprema over $[-1, 1]$, a sufficient condition for existence of a solution real and continuous on $[-1, 1]$ is $J + K < 1$.

By consideration of the Fréchet derivative of the operator equation (2.16), one establishes that $A(\phi)$ is a contraction mapping[9] of V into itself if, in addition to (2.23), one has

$$2J\chi[1 - (J\chi + K)^2]^{-1/2} < 1. \tag{2.24}$$

A simple sufficient condition upon J and K to guarantee (2.24) for any $b < 1$ is

$$5J^2(1 + 2^{1/2})^2 + 2KJ(1 + 2^{1/2}) + K^2 < 1. \tag{2.25}$$

For sufficiently small J and K the conditions (2.23) and (2.25) are satisfied. When these conditions are met, $\phi = A(\phi)$ has a unique solution in V. Furthermore, the Fréchet derivative $F_\phi(\phi_0)$ of $F(\phi, \eta) = \sin\phi - B(\phi) - g^{-1}I(\eta)$ has an inverse on $\mathcal{C}$ provided $\phi_0 \in V$. We can then apply the implicit function theorem, as in Ref. 1, Sec. 6, to prove that the unique solution ϕ in V varies continuously as the elasticities are changed within the limits implied by (2.23) and (2.25). Thus, the continuum ambiguity persists in the present formulation of the problem. As was remarked in Ref. 1, the constraint of the optical theorem does not remove the continuum ambiguity; see also Sec. 3 of the present paper.

3. CONTINUUM AMBIGUITY WHEN THE DISPERSIVE PART HAS ZEROS

The operator F, defined by

$$F(\phi, I) = g[\sin\phi - B(\phi)] - I \tag{3.1}$$

maps the space $\mathcal{C}$ into itself. The inelasticity term I, defined in Eq. (2.5), will be written as

$$I(z) = \sum_{l=0}^{\infty} I_l P_l(z), \tag{3.2}$$

$$I_l = (2l + 1)(1 - \eta_l^2)/4.$$

We are interested in the dependence on I of the solution ϕ of the unitarity equation:

$$F(\phi(I), I) = 0. \tag{3.3}$$

If ϕ is continuously differentiable with respect to I_l, it must satisfy the differential equation

$$F_\phi \frac{d\phi}{dI_l} + F_{I_l} = 0. \tag{3.4}$$

When written out explicitly, this equation reads as follows [we suppress the variable I in $\phi(z; I)$]:

$$g(z)\cos\phi(z)\frac{d\phi(z)}{dI_l} + \frac{i}{2\pi}\int_{-1}^{1} du \int_0^{2\pi} dw$$

$$g(u)g(\zeta)e^{i[\phi(\zeta)-\phi(u)]}\frac{d\phi(u)}{dI_l} = P_l(z). \tag{3.5}$$

Let us regard (3.5) as a linear integral equation for $d\phi/dI_l$. We shall show presently that the integral operator is compact on $\mathcal{C}$, if $\phi \in \mathcal{C}$. Equation (3.5) is of Fredholm type on $\mathcal{C}$, therefore, provided $\cos\phi(z)$ has no zero for $z \in \overline{T}$. If $\cos\phi$ has a zero, Eq. (3.5) is analogous to the "linear integral equation of the third kind", studied recently by Bart and Warnock.[11] By an equation of the third kind we mean an equation of the form

$$a(z)h(z) + \int_{-1}^{1} dy\, K(z, y)h(y) = b(z), \tag{3.6}$$

where $a(z)$ has a zero in the domain where $h(z)$ is defined. In general, the Fredholm theorems do not hold for an equation of this form, because of the zero of $a(z)$. In Ref. 11, it was shown that the Fredholm theorems could be retained (under certain continuity conditions on a, K, and b) provided that one seeks solutions in an appropriate space D_τ of generalized functions. The space D_τ contains continuous functions, but the solution $h(z)$ will be continuous only in the special case in which the inhomogeneous term $b(z)$ satisfies a linear constraint.

By analogy with the work of Ref. 11, we expect that the operator F_ϕ of (3.4) will not have an inverse when $\cos\phi(z)$ has a zero in $\overline{T}$. The equation

$$F_\phi \psi = \xi \tag{3.7}$$

may still have a solution ψ in $\mathcal{C}$, however, if ξ is appropriately constrained. The inhomogeneous term in (3.4) will not meet the constraint, in general. We are then led to consider the variation of two or more of the I_l. For that purpose it is convenient to set up a differential equation in which one of the I_l is considered as a dependent variable. Through solution of the differential equation, the dependent I_l will automatically be determined so as to satisfy the constraint.

Henceforth, let χ denote the dependent inelastic term, say I_m, and let λ be another inelastic term, I_l, which we shall vary independently. We now denote F of (3.1) by $F(\phi, \chi, \lambda)$ and attempt to determine $\phi(\lambda)$ and $\chi(\lambda)$ to satisfy the equation

$$F(\phi(\lambda), \chi(\lambda), \lambda) = 0. \tag{3.8}$$

We suppose that there exists a solution (ϕ^0, χ^0) of the unitarity equation for $\lambda = \lambda^0$, with a simple zero of the real part of f lying in the interval $(-z_0, z_0)$:

$$F(\phi^0, \chi^0, \lambda^0) = 0, \tag{3.9}$$

$$\cos\phi^0(x_0) = 0, \quad \cos\phi^{0\prime}(x_0) \ne 0, \quad -z_0 < x_0 < z_0, \tag{3.10}$$

$$0 < \chi < (2m + 1)/4. \tag{3.11}$$

We also suppose that ϕ^0 is analytic in S, that $\cos\phi^0(z)$ has no other zero in S, and that $P_m(\chi_0) \ne 0$. Later, we shall drop the requirement that there be only one zero, located on the real axis.

We first apply the implicit function theorem[9] to determine conditions under which (3.8) has a solution for λ close to λ_0; of course, one is interested in a solution which approaches (ϕ^0, χ^0) continuously when λ tends to λ^0. Define $\psi = (\phi, \chi) \in \mathcal{C} \times R$, R being the real line. We abbreviate $F(\phi, \chi, \lambda)$ as $F(\psi, \lambda)$ and note that, by Sec. 2, F maps $\mathcal{C} \times R \times R$ into $\mathcal{C}$; the second R refers to the real parameter λ. Let us recall the implicit function theorem[9]:

Let X, Y, and Z be Banach spaces, and suppose that F maps a neighborhood Ω of $(\psi^0, \lambda^0) \in Y \times X$ into Z, and is continuous at (ψ^0, λ^0). Also, suppose that

(1) $F(\psi^0, \lambda^0) = 0$,

(2) the Fréchet derivative $F_\psi(\psi^0, \lambda^0)$ exists in Ω and is continuous at (ψ^0, λ^0),

(3) $F_\psi(\psi^0, \lambda^0)$: $Y \to Z$ has a linear inverse,

$$F_\psi^{-1}(\psi^0, \lambda^0): Z \to Y.$$

Then there exists a function $\psi(\lambda)$, defined on a certain neighborhood $G \subset X$ of λ^0, which maps G into Y, and such that

(a) $F(\psi(\lambda), \lambda) = 0, \ \lambda \in G$,

(b) $\psi(\lambda^0) = \psi^0$,

(c) $\psi(\lambda)$ is continuous at λ^0.

Furthermore, ψ is unique in the sense that any other function with the above properties coincides with ψ if $\|\lambda - \lambda^0\| < \delta$, for some $\delta > 0$.

We identify $\mathcal{C} \times R$ with Y, R with X, and $\mathcal{C}$ with Z. It is seen easily that the unitarity operator $F(\psi, \lambda)$ meets all of the conditions of the implicit function theorem, save condition (3). We must make one further assumption to be sure that (3) is satisfied. Condition (3) will be met if the following equation has a unique solution (h, k) in $\mathcal{C} \times R$ for every right side $\xi \in \mathcal{C}$:

$$F_\psi(\psi^0, \lambda^0)(h, k) = F_\phi(\psi^0)h + F_\chi(\psi^0)k = \xi. \tag{3.12}$$

We write out (3.12) as an explicit integral equation:

$$g(z) \cos\phi^0(z)h(z) + \frac{i}{2\pi}\int_{-1}^{1} du \int_0^{2\pi} dw$$
$$f^0(\zeta) f^{0*}(u)h(u) - P_m(z)k = \xi(z), \tag{3.13}$$

where $f^0(z) = g(z) \exp[i\phi^0(z)]$. We also make use of (3.12) evaluated at x_0:

$$\frac{i}{2\pi}\int_{-1}^{1} du \int_0^{2\pi} dw\, f^0(\zeta_0) f^{0*}(u)h(u) - P_m(x_0)k = \xi(x_0). \tag{3.14}$$

Since we have assumed $P_m(x_0) \neq 0$, we may solve (3.14) for k and substitute in (3.13) to obtain

$$h(z) + \frac{i}{2\pi}\int_{-1}^{1} du \int_0^{2\pi} dw \left(\frac{f^0(\zeta) - [P_m(z)/P_m(x_0)] f^0(\zeta_0)}{g(z) \cos\phi^0(z)}\right) f^{0*}(u)h(u)$$
$$= \frac{\xi(z) - [P_m(z)/P_m(x_0)]\xi(x_0)}{g(z) \cos\phi^0(z)}. \tag{3.15}$$

Our procedure will be to show that (3.15) is a regular Fredholm equation in $\mathcal{C}$. We shall then make the *additional assumption* that the corresponding homogeneous equation has no nontrivial solution. Equation (3.15) will then possess a unique solution $h(z)$ in $\mathcal{C}$. If k is obtained by (3.14) in terms of this h, we see [by retracing the steps that led to (3.15)] that the pair (h, k) satisfies (3.12). Furthermore, this solution of (3.12) is unique in $\mathcal{C} \times R$.

To show that (3.15) is a regular Fredholm equation in $\mathcal{C}$, we first note that its right side belongs to $\mathcal{C}$. It remains only to show that the integral operator appearing on the left side is completely continuous (compact)[9] on $\mathcal{C}$; namely, that it maps any bounded set in $\mathcal{C}$ into a compact set in $\mathcal{C}$. Let K denote the integral operator, and let $\{\psi_n\}$ be a bounded sequence of functions in $\mathcal{C}$: $\psi_n \in \mathcal{C}$, $\|\psi_n\| \le M$. To show that the set of $K\psi_n$ is compact, we shall apply the theorem on compactness of sets of analytic functions quoted in Sec. 2.[10] We have merely to demonstrate that the $K\phi_n$ are analytic in S and uniformly bounded in $\overline{T}$. The function $K\phi_n$ may be written as

$$\frac{2i}{g(z) \cos\phi^0(z)} \sum_{l=0}^{\infty} (2l+1)\left[P_l(z) - \frac{P_m(z)}{P_m(x_0)} P_l(x_0)\right] f_l^0(f^{0*}\phi_n)_l. \tag{3.16}$$

The series converges uniformly in S to a function bounded uniformly with respect to n, as is seen from estimates such as (2.11). The series thus represents a function analytic in S with a zero at x_0. Furthermore, the derivative of the function at x_0 is uniformly bounded with respect to n. Thus, the $K\phi_n$ are analytic in S and uniformly bounded in $\overline{T}$.

Thus, we have established that (3.15) is a Fredholm integral equation in the Banach space $\mathcal{C}$. The solution of (3.12) will be unique unless the kernel K has -1 as an eigenvalue. We have ruled out this latter possibility by assuming that the homogeneous equation, $h + Kh = 0$, has no nontrivial solution in $\mathcal{C}$. It follows from the implicit function theorem[9] that there is a solution $[\psi(\lambda), \chi(\lambda)]$ of (3.8) for λ sufficiently close to λ^0. In other words, we have constructed a continuum of solutions of (3.3) near a solution ϕ^0, for the case in which $\cos\phi^0(z)$ has one linear zero at a real point $x_0 \in T$. Consequently, there exists a continuum of unitary scattering amplitudes $f = ge^{i\phi}$, all corresponding to a given cross section, which are obtained by varying the inhomogeneous term I. In contrast to Sec. 2, the variation of the inhomogeneous term cannot be arbitrary; but it must be subject to one constraint.

For the case in which all but a finite number L of the Legendre coefficients of the initial amplitude vanish, it follows from (3.16) that the kernel K is of rank $L - 1$. With just s and p waves nonzero, K is of rank 1 and *never* has eigenvalue -1.

If K has -1 as an eigenvalue, then $\lambda = \lambda_0$ may be a bifurcation point of the equation, i.e., two or more solution curves $[\phi(\lambda), \chi(\lambda)]$ may pass through such a point. Such a point is not necessarily a bifurcation point, however; it could happen that a solution curve would end at λ_0. In any case, it should be illuminating to investigate points where K has eigenvalues of -1.

Since the function $\cos\phi(z)$ is real-analytic in z, its zeros are either real or occur in conjugate pairs. It is straightforward to treat the case in which $\cos\phi^0(z)$ has a finite number of zeros in S. We find that for every (simple) zero of $\cos\phi^0$, one must constrain the variation of one of the inelasticities I_l in solving (3.3) with a given cross section.

The discussion up to now has not included any experimental information on the total cross section. A measurement of the differential cross section and the total cross section gives us the total inelastic cross section,

$$\sigma_{\text{in}} = \sigma_{\text{tot}} - \sigma_{\text{el}} = \sum_{l=0}^{\infty} (2l+1)I_l. \tag{3.17}$$

We may maintain (3.17), the constraint of the optical theorem, while simultaneously solving Eq. (3.8). The relation (3.17) is a linear constraint on the variation of the elasticities, and is independent of ϕ and z. The discussion now parallels that of Eq. (3.8), except that, for the case in which $\cos\phi$ has one zero in $\overline{T}$, two elasticities must be taken as dependent variables instead of one. The continuum ambiguity is thus present, even when the optical theorem constraint is satisfied. This is in contrast to the analysis involving a finite number of waves, which is given in Ref. 5.

In order to solve Eq. (3.8) numerically, it is appropriate to consider an initial value problem based on the differential equation

$$F_\phi \frac{d\phi}{d\lambda} + F_\chi \frac{d\chi}{d\lambda} + F_\lambda = 0. \tag{3.18}$$

Given the initial values ϕ^0 and χ^0, one may calculate the solution curve $[\phi(\lambda), \chi(\lambda)]$ by standard methods for numerical treatment of ordinary differential equations. The solution curve will extend to the first singularity of the Fréchet derivative F_ψ. At a singularity, numerical investigation of possible bifurcation phenomena would be required.

4. THE ABSORPTIVE PART MAPPING

In previous sections we considered a mapping of the phase function ϕ and required that the scattering amplitude f have no zeros in $\overline{T}$. Indeed, if f has a zero in $\overline{T}$, g and ϕ as defined in Sec. 2 need not be analytic in $\overline{T}$. In order to bypass this difficulty, we write the analytic unitarity equation as a mapping of the absorptive part $A(z)$, defined by

$$A(z) = (1/2i)[f(z) - f^*(z^*)]. \tag{4.1}$$

Our approach parallels that of Ref. 4, except that we shall work with the integral form of the unitarity equation.

For a given cross section $\sigma(z)$ and inelastic contribution $I(z)$, both analytic in S, the absorptive part $A(z)$ satisfies the equation

$$A(z) = B(z;A) + I(z)$$
$$= \frac{1}{4\pi}\int_{-1}^{1} du \int_0^{2\pi} dw\ [A(u)A(\zeta) + D(u)D(\zeta)] + I(z), \tag{4.2}$$

where

$$D(z) = [\sigma(z) - A^2(z)]^{1/2}, \tag{4.3}$$

ζ is defined by (2.4), and $I(z)$ by (2.5). The scattering amplitude $f(z) = D(z) + iA(z)$ can be constructed from a solution of (4.2). We first show that (4.2) has solutions in a regime for which $\sigma(z)$ has no zeros in S and for which zeros of D in S are prevented. Then, we will establish the existence of a continuum of unitary, analytic amplitudes, allowing the case where σ and D do have zeros in S. In analogy to Sec. 3, we prove that the solution A varies continuously with respect to the inelasticities I_l; as before, one must place one constraint on the variation of the I_l for each zero of $D(z)$.

We proceed to apply Schauder's theorem to show that Eq. (4.2) has at least one solution in the space $\mathfrak{C}$ of Sec. 2. We shall require that $\sigma(z)$ be analytic in S, and

$$\inf_{z\in S} |\sigma(z)| = m^2 > 0, \tag{4.4}$$

$$\sup_{z\in S} |\sigma(z)| = M^2. \tag{4.5}$$

If we require that $\|A\| \le m$, then the function $D(z)$ in (4.3) is also real-analytic in T. Furthermore, one may express $B(z;A)$ as a convergent Legendre series, as was done with the analogous integral in Eq. (2.3). From straightforward estimates on this Legendre series, it follows that $B(z;A)$ is real-analytic and bounded for $z \in S$. For an appropriate choice of b, the ball

$$V = \{A: A \in \mathfrak{C},\ \|A\| \le b,\ b < m\} \tag{4.6}$$

is mapped into itself by the operator $B + I$. If $A \in V$, it follows directly from (4.2) that

$$\|B(A) + I\| \le 2b^2 + M^2 + L, \tag{4.7}$$

where

$$L = \|I\|. \tag{4.8}$$

Consequently, $B + I$ maps V into itself if

$$2b^2 + M^2 + L \le b. \tag{4.9}$$

We may satisfy (4.9) for some $b < m$ if

$$[(M^2 + L)/m] + 2m < 1. \tag{4.10}$$

To complete the existence proof, we need note only that $B + I$ is continuous on V, and that it maps V into a compact subset of V. The compactness is proved as in Sec. 2.

Thus we can guarantee that (4.2) has at least one solution in V if condition (4.10) is satisfied. Under somewhat more restrictive conditions on σ and I, one can show that $B(z;A)$ is a contraction mapping of an appropriate ball V into itself, so that (4.2) will have a unique solution in the ball V. If follows here, as in Sec. 2 that the solution in V varies continuously with I.

In the above analysis, we have assumed that $\sigma(z)$ has no zero in S, and we were required to rule out zeros of $D(z)$ in $\overline{T}$. Here we will relax both of these conditions. We now suppose that a solution $f(z)$ of the unitarity equation is given, which is analytic in T, and such that $D(z)$ has a finite number of zeros inside T. We will show that under appropriate conditions, such a solution is one of a continuum of solutions of (4.2), all of which correspond to the same cross section. The solutions of this continuum are analytic in T; they are produced by a continuous variation of the I_l.

Unless zeroes of $D(z)$ in $\overline{T}$ are specifically excluded, one cannot guarantee that the image $B(z;A)$ of $A \in \mathfrak{C}$ in (4.2) is analytic for $z \in T$, since $D(z)$ obtained in (4.3) may not be analytic in T. We shall consider, instead, an operator $C(z;A)$, which is identical with $B(A)$ if $D(z)$ is analytic in T. Specifically, we replace $A(z)$ and $D(z)$ by their Cauchy integrals over the boundary ∂E of the elliptical domain, E, slightly smaller than T; we define

$$C(z;A) = \frac{1}{(2\pi i)^2}\int_{\partial E} dx \int_{\partial E} dy\ K(x,y,z)[A(x)A(y) + D(x)D(y)]. \tag{4.11}$$

Here, $K(x,y,z)$ is the Mandelstam kernel,[12] which is represented by the Legendre series

$$K(x,y,z) = \sum_{l=0}^{\infty} (2l+1)P_l(z)Q_l(x)Q_l(y). \tag{4.12}$$

The domain E is chosen so that all zeros of D lie inside it. For any values x and y on ∂E, it follows from the uniform convergence of (4.12) that $K(x,y,z)$ is analytic in z for $z \in S$. Furthermore, a standard estimate of (4.12) establishes the existence of a constant M such that

$$|K(x,y,z)| \leq M, \tag{4.13}$$

for all x and y on ∂E and all $z \in S$. Thus, when σ and A are bounded on ∂E, the image $C(z;A)$ is real analytic, bounded in S, and thus bounded in $\overline{T}$.

The operator $F(A,I)$, defined by

$$F(A,I) = A - C(A) - I, \tag{4.14}$$

maps $A \in \mathcal{C}$ into an element of $\mathcal{C}$ if σ and I are analytic in S.

Let us examine the variation with I of the solution A of

$$F(A,I) = 0 \tag{4.15}$$

with a fixed cross section σ. If the solution A is continuously differentiable with respect to I_l, then it must satisfy the differential equation

$$F_A \frac{dA}{dI_l} + F_{I_l} = 0. \tag{4.16}$$

The operator F_A, unlike the corresponding operator F_ϕ of Eq. (3.4), is a regular Fredholm operator on $\mathcal{C}$, even if $D(z)$ has zeros inside the ellipse E. Specifically, F_A applied to a function $h \in \mathcal{C}$ has the form

$$F_A(A)h = h(z) + C_A(z;A)\cdot h$$

where

$$C_A(z;h) = \frac{1}{2\pi^2}\int_{\partial E} dx \int_{\partial E} dy\, K(x,y,z) \times \left[A(x) - \frac{D(x)}{D(y)}A(y)\right]h(y). \tag{4.17}$$

Note: It follows from (4.16) that if $h \in \mathcal{C}$ and $\|h\| \leq b$, $C_A(z,h)$ is analytic and uniformly bounded for $z \in S$. Thus, by the criterion of Ref. 9, C_A is a completely continuous operator on $\mathcal{C}$.

A solution of (4.15) is a solution of the unitarity equation (4.2) provided that $D(z)$ is analytic inside E. The solution A of (4.15) varies continuously with the I_l, but the corresponding $D(z)$ is analytic inside the domain E only for appropriately constrained variations of the I_l. Thus, the situation is similar to that encountered in Sec. 2, even though an integral equation of the third kind does not arise in the present formulation.

For our given scattering amplitude, $D(z) = [\sigma(z) - A^2(z)]^{1/2}$ is analytic in T. Consequently, all zeros of $\sigma(z) - A^2(z)$ in T are of even order. If we solve (4.15) while enforcing the constraint that zeros remain of even order, then the solution of (4.15) will also be a solution of the unitarity equation (4.2). For simplicity, we discuss the case in which there is only one real, simple zero x_0 of D in E, and assume that $\sigma(x_0) \neq 0$. Since $\sigma - A^2$ has a second order zero at x_0, we have that

$$\sigma(x_0) - [C(x_0;A) + I(x_0)]^2 = 0 \tag{4.18}$$

and

$$\sigma_z(x_0) - 2[C(x_0;A) + I(x_0)][C_z(x_0;A) + I_z(x_0)] = 0, \tag{4.19}$$

where the subscript z denotes partial differentiation with respect to z. If we enforce (4.18) and (4.19) while varying the inelasticity I and solving (4.15), D will remain analytic in E and A will be a solution of the unitarity equation. Let us define $g(z) = [\sigma(z)]^{1/2}$, such that $g(x_0) = C(x_0) + I(x_0)$. Then $g(z)$ is analytic in z in a neighborhood of x_0. We can express the above constraints as

$$g(x_0) - C(x_0;A) - I(x_0) = 0, \tag{4.20}$$

$$g_z(x_0) - C_z(x_0;A) - I_z(x_0) = 0. \tag{4.21}$$

To impose (4.20) and (4.21) while varying I in (4.15), we define the independent variable $\lambda = I_l$ and the dependent variables $A(\lambda)$, $x_0 = \mu(\lambda)$, and $I_m = \chi(\lambda)$, where $P_m(x_0) \neq 0$. The system of equations (4.15), (4.20), and (4.21) is viewed as a mapping of $\mathcal{C} \times R \times R$ into $\mathcal{C} \times R \times R$, where $A(\lambda) \in \mathcal{C}$, $\chi(\lambda) \in R$, and $\mu(\lambda) \in R$, and it is represented as

$$F(\Psi(\lambda), \lambda) = 0, \tag{4.22}$$

where 0 is the zero element of $\mathcal{C} \times R \times R$ and $\Psi(\lambda) = [A(\lambda), \chi(\lambda), \mu(\lambda)]$. For our initial solution of the unitarity equation, $\Psi^0 = [A^0, I_m^0, x_0]$, we have

$$F(\Psi^0, \lambda^0) = 0. \tag{4.23}$$

We shall apply the implicit function theorem to establish that there is a solution $\Psi(\lambda)$ of (4.22) for λ sufficiently close to λ^0. This solution varies continuously with λ. It is trivial to verify all but one of the suppositions of the implicit function theorem. The one condition which is not obviously satisfied is that the Fréchet derivative F_Ψ evaluated at Ψ^0 have an inverse. That is, we must show that the set of equations for $(h, k_1, k_2) \in \mathcal{C} \times R \times R$,

$$F_\Psi(\Psi^0, \lambda^0)(h, k_1, k_2) = (\xi, \xi_1, \xi_2) \tag{4.24}$$

has a solution for arbitrary $(\xi, \xi_1, \xi_2) \in \mathcal{C} \times R \times R$. In explicit form the equations (4.24) are

$$h(z) - C_A(z, A^0)h - P_m(z)k_1 = \xi(z) \tag{4.25}$$

$$- C_A(x_0, A^0)h - P_m(x_0)k_1 = \xi_1, \tag{4.26}$$

$$- C_{zA}(x_0, A^0)h - P_m'(x_0)k_1 + [g_{zz}(x_0) - C_{zz}(x_0, A^0) - I_{zz}^0(x_0)]k_2 = \xi_2. \tag{4.27}$$

Since by assumption $P_m(x_0) \neq 0$, we may use the second of these equations to eliminate k_1 from the first equation. Thus, we obtain

$$h(z) - \{C_A(z, A^0) - [P_m(z)/P_m(x_0)]C_A(x_0, A^0)\}h = \xi(z) - \xi_1 P_m(z)/P_m(x_0). \tag{4.28}$$

Since $C_A(z;A^0)$ is a completely continuous operator on $\mathcal{C}$, (4.28) is a regular Fredholm equation. We assume that (i) the homogeneous form of (4.28) has no nontrivial solution, and (ii) the coefficient of k_2 in (4.27) is not zero [i.e., the zero of $D(z)$ is of first order]. Under these assumptions, Eq. (4.24) has a unique solution for any right side. It then follows from the implicit function theorem that Eq. (4.22) has a unique solution in $\mathcal{C} \times R \times R$ in a neighborhood of λ^0, such that $\Psi(\lambda^0) = \Psi^0$. The absorptive part $A(\lambda)$, the zero location $\mu(\lambda)$, and the dependent elasticity $\chi(\lambda)$ all vary continuously with the independent elasticity λ in a neighborhood of λ^0.

It is straightforward to generalize the implicit function arguments to the cases in which (1) $D(z)$ has more than one zero in T, (2) $D(z)$ has complex zeros, (3) $D(z)$ has zeros of higher order, and (4) D and σ have a zero at the same point.

5. MAPPING OF THE SCATTERING AMPLITUDE

In the previous sections we have considered the unitarity equation as a mapping of the space $\mathcal{C}$ of real-analytic functions into itself. Here we will briefly examine the unitarity equation on the Banach space $\mathfrak{A} = \mathcal{C} \times \mathcal{C}$ of functions $f(z)$ with the properties: (a) $f(z)$ is analytic inside the domain T, (b)

$$\|f\| = \sup_{z\in T} |f(z)|. \qquad (5.1)$$

The function $f(z)$ can be decomposed into its dispersive and absorptive parts, $D(z)$ and $A(z)$, which are real-analytic functions, as follows:

$$D(z) = [f(z) + f^*(z^*)]/2, \qquad (5.2)$$

$$A(z) = [f(z) - f^*(z^*)]/2i, \qquad (5.3)$$

$$f(z) = D(z) + iA(z). \qquad (5.4)$$

Note: the elements of $\mathfrak{A}$ are the ordered pairs (D, A). Even though f need not be real-analytic, the scalars for the Banach space $\mathfrak{A}$ are real numbers.

In this section we express the unitarity equation as a mapping of the scattering amplitude $f(z)$; here we shall allow zeros of the cross section $\sigma(z)$ as well as zeros of $D(z)$ in T. In the spirit of Sec. 3 our treatment will involve an integral equation of the third kind. The unitarity equation for the scattering amplitude is equivalent to

$$F(f,I) \equiv f^2(z) - \sigma(z) - 2if(z)[B(z;f) + I(z)] = 0, \qquad (5.5)$$

where $B(z;f)$ is the integral operator of (4.2) and $I(z)$ is given in (2.5). The operator F maps the space $\mathfrak{A}$ into itself. As in the previous sections, we will show that a given solution of (5.5), $f_0(z) \in \mathfrak{A}$, is a member of a continuum of solutions, corresponding to the same cross section $\sigma(z)$, and produced by variation of $I(z)$.

As before, the existence of the continuum is established via the implicit function theorem. The significant premise for the theorem is that an appropriate Fréchet derivative have an inverse. The Fréchet derivative of F with respect to f evaluated at f_0 is

$$F_f(f_0)h = 2[f_0(z) - i(B(z,f_0) + I^0(z)]h - 2if_0(z)B_f(z,f_0)h, \qquad (5.6)$$

where $B_f(z,f)h$ is expressed in terms of the dispersive and absorptive parts of h (h_D and h_A, respectively), as

$$B_f(z,f_0)h = \frac{1}{2\pi}\int_{-1}^{1} du \int_0^{2\pi} dw\ [A_0(y)h_A(u) + D_0(y)h_D(u)]. \qquad (5.7)$$

$B_f(z,f_0)$ is a completely continuous operator on the space $\mathfrak{A}$ for $f_0 \in \mathfrak{A}$. At a solution of (5.5),

$$D_0(z) = f_0(z) - i[B(z,f_0) + I^0(z)]. \qquad (5.8)$$

The operator $F_f(f_0)$ has an inverse on $\mathfrak{A}$ unless $D_0(z)$ vanishes at some point $z_0 \in T$, in the latter case $F_f(f_0)$ is an integral operator of the third kind. When $D_0(z)$ or $\sigma_0(z)$ does have a zero in T, however, there exists an inverse of an appropriate extension of $F_f(f_0)$ on an augmented space.

For simplicity, we describe the case in which $D_0(z)$ has only one real, simple zero in $\overline{T}$, at an interior point x_0, such that $\sigma(x_0) \neq 0$. One may easily extend the argument to treat all of the cases of Sec. 4. As in Sec. 3, we take on inelasticity parameter $I_m = \chi$ [chosen such that $P_m(x_0) \neq 0$] to be a dependent variable. We are led to consider the following Fréchet derivative operator, which maps (h,k) in $\mathfrak{A} \times R$ into ξ in $\mathfrak{A}$:

$$F_f(f_0)h + F_\chi(f_0)k = \xi. \qquad (5.9)$$

This operator does have an inverse which maps $\mathfrak{A}$ into $\mathfrak{A} \times R$, under reasonable restrictions which are analogous to those of Sec. 3. The implicit function theorem assures the existence of a solution $f(I) \in \mathfrak{A}$ of (5.5), such that $f(I^0) = f^0$. This solution varies continuously with respect to I, when the variation of one particular inelasticity, I_m, is constrained.

The considerations of this section allow us to take the unitarity integral $B(z,f)$ over the physical region, even if $\sigma(z)$ has zeros in T. However, the elements of the Banach space $\mathfrak{A}$ are pairs of real-analytic functions and third-kind integral equations are encountered.

ACKNOWLEDGMENTS

We benefited greatly from conversations with Professor David Atkinson, which were made possible through a travel grant from NATO. We also wish to thank Professor Atkinson, Professor Mahoux, and Professor Yndurâin for sending a preliminary copy of Ref. 4.

*Research supported in part by the National Science Foundation.

[1] D. Atkinson, P. W. Johnson, and R. L. Warnock, Commun. Math. Phys. **28**, 133 (1972).

[2] In the Cutkosky–Deo approach to phase-shift analysis [Phys. Rev. **174**, 1859 (1968)], one does have all partial waves, but the amplitudes are still expressed in terms of a finite number of parameters. Again, chi-squared has finite number of minima, and the continuum ambiguity does not appear.

[3] A. Martin, *Scattering Theory: Unitarity, Analyticity, and Crossing*, Lecture Notes in Physics, Vol. 3 (Springer-Verlag, Heidelberg, 1969).

[4] D. Atkinson, G. Mahoux, and F. J. Yndurâin, "Construction of a Unitary Analytic Scattering Amplitude. I-Scalar Particles," Saclay preprint, September, 1972; submitted to Nucl. Phys.

[5] A. Gersten, Nucl. Phys. B **12**, 537 (1969).

[6] E. Hille, *Analytic Function Theory, Vols. 1 and 2* (Ginn, Boston, 1959).

[7] For u and w within the region of integration in (2.3), it follows from (2.4) that $z \in T$ implies $\zeta \in T$. This is easily proved by representing points in the complex plane by unifocal elliptical coordinates (z_1, θ): $z = z_1\cos\theta + i(z_1^2 - 1)^{1/2}\sin\theta$

[8] The constant M approaches infinity as z_0'' approaches 1.

[9] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon, Oxford, 1964).

[10] Ref. 6, Vol. 2, Theorem 15.2.3, p. 246.

[11] G. R. Bart and R. L. Warnock, "Linear Integral Equations of the Third Kind," Argonne-IIT preprint, November, 1971; to be published in the SIAM J. Math. Anal.

[12] M. L. Goldberger and K. M. Watson, *Collision Theory* (Academic, New York, 1964), p. 605.

Scattering from a random rough surface: Diagram methods for elastic media

John A. DeSanto

Naval Research Laboratory, Washington, D. C. 20375

(Received 27 March 1973)

In a previous paper [J. Math. Phys. **13**, 1903 (1972)], Feynman diagram methods were used to construct the Green's function and its first two moments for scalar wave scattering from a random rough surface with a Neumann boundary condition. This paper extends these formal diagram methods to the calculation of the Green's function and its first two moments for an elastic half-space bounded by a random rough surface considered as a free boundary. The surface height is a single valued bounded function with Gaussian statistics. The Dyson and Bethe–Salpeter equations, for the mean and second moment respectively, of the Green's function, are derived. Some simplifications of these integral equations and some examples are presented.

1. INTRODUCTION

A method to calculate the Green's function and its first two statistical moments for the scalar wave equation in a half space bounded by a random rough surface was previously presented.[1] The boundary was considered to be a hard or Neumann boundary, and the surface height to be a single valued bounded function with Gaussian statistics and zero mean. The Green's function was applied to a scattering problem for plane wave incidence on the surface. The method used to discuss the Green's function was in terms of Feynman-like diagrams analogous to those used in random volume scattering problems.[2] Here we extend these formal diagram methods to the calculation of the Green's function and its first two moments for an isotropic elastic half space bounded by a random rough surface. The surface is treated as a free boundary, i.e., one having zero stress. As in Ref. 1, the method can be generalized to calculate higher order moments of the Green's function, but we only consider the first two.

The method is straightforward. The deterministic surface Green's function is obtained by formulating an integral equation in Sec. 2. The Born term and kernel of the integral equation are expressed in terms of the infinite space elastic Green's function which is known.[3] To formulate the integral equation requires a boundary condition and we choose a zero stress (free) boundary for simplicity. Previously, Case and Hazeltine[4] calculated the elastic half space Green's function for a flat surface and applied their result to the problem of elastic radiation from a small source in the earth's interior. Also, Karal and Keller[5] have considered elastic wave propagation as a random volume scattering problem. Their aim was to calculate effective wave numbers. Our results in this section can be considered as a rough surface generalization of the Case–Hazeltine results, and reduce to the latter in the flat surface limit.

Fourier methods and a gauge condition argument are used in Sec. 3 to define an integral equation in momentum- or k-space for the surface Green's function and an auxiliary function related to it which is used in the actual calculations. The kernel of the integral equation can be factored into propagator, vertex and interaction components and k-space diagram rules are associated with each component. It is shown how to express the field Green's function in terms of the surface Green's function, and a general reduction method is derived which expresses the outgoing scattered field in terms of the incident field. This is also used in the discussion of the mutual coherence function in Sec. 4.

The main discussion of the paper, that of a random rough surface, is presented in Sec. 4. The statistical arguments are the same as those for the scalar case in Ref. 1, and the diagram rules are accordingly extended. A coupled set of Dyson type integral equations is obtained for the mean of the Green's function. Its solution represents the coherent part of the source field. A coupled set of Bethe–Salpeter type integral equations is obtained for the mean of the square of the Green's function and it is shown how this quantity can be related to the intensity. Simplifications of both sets of integral equations result from the translational invariance of the statistical problem, and lowest order examples of coherent and incoherent diagrams are presented.

Finally, in the Appendix, we calculate a kernel function which is used in the derivation of the integral equation in Sec. 2.

2. SURFACE GREEN'S FUNCTION

The problem is to calculate the Green's function $\Gamma_{in}(\mathbf{x},\mathbf{x}')$ which satisfies the inhomogeneous elastic differential equation

$$[\Delta^*\Gamma(\mathbf{x},\mathbf{x}')]_{in} + k_0^2\Gamma_{in}(\mathbf{x},\mathbf{x}') = -\delta_{in}\delta(\mathbf{x},\mathbf{x}') \qquad (2.1)$$

in the region V indicated in Fig. 1. The Δ^* operator is defined by

$$[\Delta^*\Gamma(\mathbf{x},\mathbf{x}')]_{in} = \mu\partial_m\partial_m\Gamma_{in}(\mathbf{x},\mathbf{x}') + (\lambda+\mu)\partial_i\partial_m\Gamma_{mn}(\mathbf{x},\mathbf{x}'). \qquad (2.2)$$

The region V is an isotropic elastic half space characterized by the two elastic constants λ, the Lamé modulus, and μ, the shear modulus.[3] V is geometrically specified by $z \geq h(x_\perp)$ where $h(x_\perp)$ is a random variable and $x_\perp$ the transverse component of a 3-vector $\mathbf{x} = (x_\perp, z)$. The abbreviation $\partial_i = \partial/\partial x_i$ is used, the subscripts $i, m, n = 1, 2, 3$, the summation convention is assumed, k_0 is the free space wave number, δ_{in} the kronecker delta, and $\delta(\mathbf{x})$ the three-dimensional Dirac delta function. In addition, Γ_{in} satisfies a zero stress (free) boundary condition when $z = h(x_\perp)$. This is specified later.

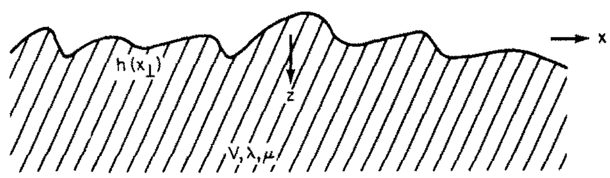

FIG. 1 Random rough surface $z = h(x_\perp)$ bounding an isotropic elastic half space V specified by Lamé constants λ and μ.

Copyright © 1973 by the American Institute of Physics

The infinite space Green's functions $\Gamma^{0\pm}_{in}(\mathbf{x},\mathbf{x}') = \Gamma^{0\pm}_{in}(\mathbf{x}-\mathbf{x}')$ satisfy the same equation

$$[\Delta^*\Gamma^0(\mathbf{x},\mathbf{x}')]_{ij} + k_0^2\Gamma^0_{ij}(\mathbf{x},\mathbf{x}') = -\delta_{ij}\delta(\mathbf{x}-\mathbf{x}') \quad (2.3)$$

but with outgoing (+) or incoming (−) wave boundary conditions as $|\mathbf{x}-\mathbf{x}'| \to \infty$. We have dropped the + and − superscripts for simplicity. They will be used later as necessary. The explicit representation of $\Gamma^0_{ij}(\mathbf{x})$ is

$$\Gamma^0_{ij}(\mathbf{x}) = \mu^{-1}\delta_{ij}G_0^t(\mathbf{x}) + k_0^{-2}\partial_i\partial_j[G_0^t(\mathbf{x}) - G_0^l(\mathbf{x})], \quad (2.4)$$

where $G_0^{t,l}$ are the scalar infinite space Green's functions

$$G_0^{t,l}(\mathbf{x}) = \exp\{ik_{t,l}|\mathbf{x}|\}/|\mathbf{x}| \quad (2.5)$$

and the transverse (t) and longitudinal (l) wave numbers are given by

$$k_t^2 = k_0^2/\mu, \quad k_l^2 = k_0^2/(\lambda + 2\mu). \quad (2.6)$$

Forming the quantity

$$\Gamma^0_{ij}(\mathbf{x}',\mathbf{x})[\Delta^*\Gamma(\mathbf{x},\mathbf{x}'')]_{in} - [\Delta^*\Gamma^0(\mathbf{x}',\mathbf{x})]_{ij}\Gamma_{in}(\mathbf{x},\mathbf{x}''),$$

it is possible to write the identity

$$\begin{aligned}\Gamma_{jn}(\mathbf{x},\mathbf{x}'')\delta(\mathbf{x}'-\mathbf{x}) - \Gamma^0_{jn}(\mathbf{x}',\mathbf{x})\delta(\mathbf{x}-\mathbf{x}'') &= \partial_l\{\Gamma^0_{ij}(\mathbf{x}',\mathbf{x})[T\Gamma(\mathbf{x},\mathbf{x}'')]_{lin} \\ &\quad - [T\Gamma^0(\mathbf{x}',\mathbf{x})]_{lij}\Gamma_{in}(\mathbf{x},\mathbf{x}'')\},\end{aligned} \quad (2.7)$$

where the traction operator T is defined as

$$\begin{aligned}[T\Gamma(\mathbf{x},\mathbf{x}')]_{lin} &= \mu\delta_{lm}\{\partial_m\Gamma_{in}(\mathbf{x},\mathbf{x}') + \partial_i\Gamma_{mn}(\mathbf{x},\mathbf{x}')\} \\ &\quad + \lambda\delta_{il}\partial_m\Gamma_{mn}(\mathbf{x},\mathbf{x}').\end{aligned} \quad (2.8)$$

Multiplying (2.8) by the unit step function $\theta(z' - h(x'_\perp))$

$$\theta(z) = \begin{cases} 1, & z > 0 \\ 0, & z < 0, \end{cases}$$

integrating the result over all space, doing a partial integration on the resulting integral term, and using the free boundary condition

$$[\delta_{l3} - \partial_{l\perp}h(x_\perp)][T\Gamma(\mathbf{x}_s,\mathbf{x}'')]_{lin} = 0, \quad (2.9)$$

where $\mathbf{x}_s$ is a 3-vector evaluated on the surface $\mathbf{x}_s = (x_\perp, h(x_\perp))$, yields the result

$$\begin{aligned}\Gamma^D_{jn}(\mathbf{x}',\mathbf{x}'') &= \Gamma^0_{jn}(\mathbf{x}',\mathbf{x}'')\theta(z'' - h(x''_\perp)) \\ &\quad - \int d^2x_\perp N_l(x_\perp)[T'\Gamma^0(\mathbf{x}',\mathbf{x}_s)]_{lij} \\ &\quad \times \Gamma_{in}(\mathbf{x}_s,\mathbf{x}''),\end{aligned} \quad (2.10)$$

where T' means differentiation on the primed coordinate, the normal (into V) defined by $N_l(x_\perp) = \delta_{l3} - \partial_{l\perp}h(x_\perp)$ as in (2.9), the integration over the full surface, and the discontinuous Green's function Γ^D_{jn} defined by

$$\Gamma^D_{jn}(\mathbf{x}',\mathbf{x}'') = \Gamma_{jn}(\mathbf{x}',\mathbf{x}'')\theta(z' - h(x'_\perp)). \quad (2.11)$$

It is possible to write the kernel in the integral term in (2.10) as (see Appendix)

$$\begin{aligned}N_l(x_\perp)[T'\Gamma^0(\mathbf{x}'-\mathbf{x}_s)]_{lij} &= -\tfrac{1}{2}\delta(x'_\perp - x_\perp)\epsilon(z' - h(x'_\perp))\{\delta_{ij} - \delta_{i3}\partial_{j\perp}h(x_\perp) \\ &\quad - \Lambda\delta_{j3}\partial_{i\perp}h(x_\perp)\} - R_{ji}(\mathbf{x}',\mathbf{x}_s),\end{aligned} \quad (\text{A5})$$

where $\Lambda = \lambda/(\lambda+\mu)$, $\epsilon(z) = \theta(z) - \theta(-z)$, and R_{ji} is defined as

$$R_{ji}(\mathbf{x}',\mathbf{x}_s) = (2\pi)^{-3}\int d^3k e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}_s)}R_{jim}(\mathbf{k})N_m(x_\perp)$$

with (A6, 7)

$$\begin{aligned}iR_{jim}(\mathbf{k}) &= G_0^t(k)\left\{\delta_{ij}\left[k_{m\perp} + \delta_{m3}P\left(\frac{k_t^2 - k_\perp^2}{k_3}\right)\right]\right. \\ &\quad \left. + \delta_{jm}\left[k_{i\perp} + \delta_{i3}P\left(\frac{k_t^2 - k_\perp^2}{k_3}\right)\right]\right\} \\ &\quad + G_0^l(k)\Lambda\delta_{im}\left[k_{j\perp} + \delta_{j3}P\left(\frac{k_l^2 - k_\perp^2}{k_3}\right)\right] \\ &\quad - 2[G_0^t(k) - G_0^l(k)]k_ik_jk_m/k_t^2 \\ &\quad + \frac{2\delta_{i3}\delta_{j3}\delta_{m3}}{k_t^2}P\left(\frac{k_t^2 - k_l^2}{k_3}\right),\end{aligned} \quad (\text{A8})$$

where the symbol P stands for the Cauchy principle value distribution.

Substituting (A5) into (2.9) and letting $\mathbf{x}' \to \mathbf{x}'_s$ through positive z' values yields the result

$$\begin{aligned}\Gamma^s_{jn}(\mathbf{x}'_{s'},\mathbf{x}'') &= \Gamma^0_{jn}(\mathbf{x}'_{s'},\mathbf{x}'') + \int d^2x_\perp R_{ji}(\mathbf{x}'_{s'},\mathbf{x}_s) \\ &\quad \times U_{ip}(x_\perp)\Gamma^s_{pn}(\mathbf{x}_s,\mathbf{x}''),\end{aligned} \quad (2.12)$$

where we have defined the surface Green's function Γ^s_{jn} as

$$\Gamma^s_{jn}(\mathbf{x}'_s,\mathbf{x}'') = Q_{jm}(x'_\perp)\Gamma_{mn}(\mathbf{x}'_s,\mathbf{x}''), \quad (2.13)$$

$$Q_{jm}(x_\perp) = \tfrac{1}{2}\{\delta_{jm} + \delta_{m3}\partial_{j\perp}h(x_\perp) + \Lambda\delta_{j3}\partial_{m\perp}h(x_\perp)\} \quad (2.14)$$

with the inverse

$$\Gamma_{mn}(\mathbf{x}'_s,\mathbf{x}'') = U_{mp}(x'_\perp)\Gamma^s_{pn}(\mathbf{x}'_{s'},\mathbf{x}''), \quad (2.15)$$

$$\begin{aligned}U_{ip}(x_\perp) &= 2(\delta_{ip} - \delta_{i3}\delta_{p3}) \\ &\quad + 2\{1 - \Lambda[\partial_\perp h(x_\perp)]^2\}^{-1}N_i(x_\perp) \\ &\quad \times\{\Lambda N_p(x_\perp) + (1-\Lambda)\delta_{p3}\}.\end{aligned} \quad (2.16)$$

Equation (2.12) is the integral equation for the surface Green's function which we deal with in the next section.

The flat surface limit of (2.12) is found by setting $h = 0$ in (2.12), noting that $N_m(x_\perp) = \delta_{m3}$, $U_{mp} = 2\delta_{mp}$, and that in (A6) for R_{ji} it is possible to do the k_3 integration to yield

$$\begin{aligned}R_{ji}(\mathbf{x}'_{s'},\mathbf{x}_s) &= R_{ji}(x'_\perp, x_\perp) \\ &\equiv (2\pi)^{-2}\int d^2k_\perp e^{ik_\perp\cdot(x'_\perp - x_\perp)}R_{ji}(k_\perp),\end{aligned}$$

$$R_{ji}(k_\perp) = \tfrac{1}{2}k_{i\perp}\delta_{j3}\frac{K_l - \Lambda K_t}{K_t(K_l + K_t)} + \tfrac{1}{2}k_{j\perp}\delta_{i3}\frac{\Lambda K_t - K_l}{K_l(K_l + K_t)},$$

where $K^2_{l,t} = k^2_{l,t} - k_\perp^2$. The integral equation resulting from (2.12) can be Fourier transformed and agrees with the result of Case and Hazeltine.[4] This is readily apparent if we note that $R_{ji}(k_\perp) = -H^0_{ji}(\vec{k})$ in the latter's notation [Eq. (2.28) in Ref. 4].

3. DIAGRAM AND REDUCTION METHODS

In order to discuss a diagram notation, it is convenient to express $\Gamma^s_{jn}(\mathbf{x}'_{s'},\mathbf{x}'')$ as a Fourier transform. Introduce the following Fourier transforms in (2.12):

$$\Gamma^s_{jn}(\mathbf{x}'_{s'},\mathbf{x}'') = (2\pi)^{-6} \iint d^3k' d^3k'' e^{i\mathbf{k}'\cdot\mathbf{x}'_s}\Gamma^s_{jn}(\mathbf{k}',\mathbf{k}'')e^{-i\mathbf{k}''\cdot\mathbf{x}''}, \tag{3.1}$$

$$\Gamma^0_{jn}(\mathbf{x}'_s,\mathbf{x}'') = (2\pi)^{-6} \iint d^3k' d^3k'' e^{i\mathbf{k}'\cdot\mathbf{x}'_s} \times [(2\pi)^3\delta(\mathbf{k}''-\mathbf{k}')\Gamma^0_{jn}(\mathbf{k}')]e^{-i\mathbf{k}''\cdot\mathbf{x}''}, \tag{3.2}$$

$$R_{ji}(\mathbf{x}'_s,\mathbf{x}'') = (2\pi)^{-3} \int d^3k e^{i\mathbf{k}\cdot(\mathbf{x}'_s-\mathbf{x}'')}R_{ji}(\mathbf{k},\mathbf{x}_s) \tag{A6}$$

with [combining (A7) and (A10)]

$$R_{ji}(\mathbf{k},\mathbf{x}_s) = -iP_n(k)R_{njim}(\mathbf{k})[\delta_{m3}-\partial_{m\perp}h(x_\perp)],$$

where the functions P_n and R_{njim} are defined by (A11) and (A12). Setting the resulting integrand equal to zero using the same gauge transformation arguments used in Ref. 1, there results the k-space integral equation

$$\Gamma^s_{jn}(\mathbf{k}',\mathbf{k}'') = (2\pi)^3\delta(\mathbf{k}'-\mathbf{k}'')\Gamma^0_{jn}(\mathbf{k}') + \int d^3k L_{jp}(\mathbf{k}',\mathbf{k})\Gamma^s_{pn}(\mathbf{k},\mathbf{k}''), \tag{3.3}$$

where the kernel L_{jp} is defined as

$$L_{jp}(\mathbf{k}',\mathbf{k}) = \frac{-i}{(2\pi)^3}P_n(k')R_{njim}(\mathbf{k}') \times \int d^2x_\perp e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_s}[\delta_{m3}-\partial_{m\perp}h(x_\perp)]U_{ip}(x_\perp). \tag{3.4}$$

Using (2.16) for U_{ip}, integrating (3.4) by parts and dropping any surface terms (see Ref. 1), enables us to write

$$L_{jp}(\mathbf{k}',\mathbf{k}) = P_n(k')V_{njp}(\mathbf{k}',\mathbf{k})A(\mathbf{k}'-\mathbf{k}), \tag{3.5}$$

where $V_{njp}(\mathbf{k}',\mathbf{k})$ and $A(\mathbf{k})$ are given by

$$V_{njp}(\mathbf{k}',\mathbf{k}) = \frac{-2i}{(2\pi)^3}R_{njim}(\mathbf{k}')\frac{(\mathbf{k}'-\mathbf{k})_m}{k'_3-k_3} \times \left(\delta_{ip}-\delta_{i3}\delta_{p3}+(\mathbf{k}'-\mathbf{k})_i \times \frac{\Lambda(\mathbf{k}'-\mathbf{k})_p+(1-\Lambda)(k'_3-k_3)\delta_{p3}}{(k'_3-k_3)^2-\Lambda(k'_\perp-k_\perp)^2}\right), \tag{3.6}$$

$$A(\mathbf{k}) = \int d^2x_\perp e^{-ik_\perp\cdot x_\perp}e^{-ik_3h(x_\perp)}. \tag{3.7}$$

Thus (3.5) has been factored into propagator (P), vertex (V), and interaction (A) terms. Equation (3.7) is the same interaction term which appeared in Ref. 1.

It is convenient to define away the delta function term in (3.3) by introducing the 3-index function G^s_{njm} via

$$\Gamma^s_{jl}(\mathbf{k}',\mathbf{k}'') = (2\pi)^3\delta(\mathbf{k}'-\mathbf{k}'')\Gamma^0_{jl}(\mathbf{k}') + (2\pi)^3P_n(k')G^s_{njm}(\mathbf{k}',\mathbf{k}'')\Gamma^0_{ml}(\mathbf{k}''). \tag{3.8}$$

Substituting (3.5) and (3.8) into (3.3) yields the integral equation

$$G^s_{mjn}(\mathbf{k}',\mathbf{k}'') = V_{mjn}(\mathbf{k}',\mathbf{k}'')A(\mathbf{k}'-\mathbf{k}'') + \int d^3k V_{mjp}(\mathbf{k}',\mathbf{k})A(\mathbf{k}'-\mathbf{k})P_l(k)G^s_{lpn}(\mathbf{k},\mathbf{k}''). \tag{3.9}$$

This is the integral equation we deal with. A diagram interpretation can be given to each term in the equation similar to that presented in Ref. 1. It is illustrated in Fig. 2. When elementary diagrams are pieced together we must integrate over internal momenta k and repeated indices, of course, must be summed.

Once the surface Green's function is known, the field or discontinuous Green's function Γ^D_{jn} can be found. The procedure is to use the Fourier transform

$$\Gamma^D_{jn}(\mathbf{x}',\mathbf{x}'') = (2\pi)^{-6} \iint d^3k' d^3k'' e^{i\mathbf{k}'\cdot\mathbf{x}'}\Gamma^D_{jn}(\mathbf{k}',\mathbf{k}'')e^{-i\mathbf{k}''\cdot\mathbf{x}''} \tag{3.10}$$

with (2.15), (3.1), and (3.2) in (2.10). Since $z', z'' > h$, the kernel of (2.10) has no singularities and can be differentiated directly. It can easily be seen that, using (2.8) and direct differentiation, it is possible to write

$$N_l(x_\perp)[T'\Gamma^0(\mathbf{x}',\mathbf{x}_s)]_{lij} = -R^0_{ji}(\mathbf{x}',\mathbf{x}_s), \tag{3.11}$$

$$R^0_{ji}(\mathbf{x}',\mathbf{x}_s) = (2\pi)^{-3} \int d^3k e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}_s)}R^0_{jim}(\mathbf{k})N_m(x_\perp), \tag{3.12}$$

$$\begin{aligned} iR^0_{jim}(\mathbf{k}) &= \mu[k_m\Gamma^0_{ij}(\mathbf{k})+k_i\Gamma^0_{jm}(\mathbf{k})]+\Lambda\delta_{im}k_jG^l_0(k) \\ &= G^t_0(k)[\delta_{ij}k_m+\delta_{jm}k_i-2k_ik_jk_m/k_t^2] \\ &\quad + G^l_0(k)(\Lambda\delta_{im}k_j+2k_ik_jk_m/k_t^2). \end{aligned} \tag{3.13}$$

These latter three equations are presented as an analog of the presentation of the kernel in Sec. 2. The notation is similar except for the superscript "0" on the R functions. This is to indicate that in addition to the fact that (3.13), e.g., can be derived directly, it also follows from (A10) through (A12) by putting the principle value terms "on-shell," viz. by setting $k_t^2 = k_3^2 + k_\perp^2$ and $k_l^2 = k_3^2 + k_\perp^2$ in the principle value terms. Using the results of this discussion, (2.10) becomes the integral relation

$$\Gamma^D_{jn}(\mathbf{k}',\mathbf{k}'') = (2\pi)^3\delta(\mathbf{k}'-\mathbf{k}'')\Gamma^0_{jn}(\mathbf{k}') + P_n(k')\int d^3k V^0_{njp}(\mathbf{k}',\mathbf{k})A(\mathbf{k}'-\mathbf{k})\Gamma^s_{pn}(\mathbf{k},\mathbf{k}''). \tag{3.14}$$

Note that this is similar to the integral equation (3.3) [with (3.5)] for $\Gamma^s_{jn}(\mathbf{k}',\mathbf{k}'')$. The difference is that we have gone on-shell, i.e., have replaced V_{njp} with V^0_{njp}. We easily get the result

$$\Gamma^s_{jn}(\mathbf{k}',\mathbf{k}'')\big|_{\text{on-shell}} = \Gamma^D_{jn}(\mathbf{k}',\mathbf{k}'') \tag{3.15}$$

and, in general, using (3.3), (3.5), and (3.14) we get

$$\Gamma^D_{jn}(\mathbf{k}',\mathbf{k}'') = \Gamma^s_{jn}(\mathbf{k}',\mathbf{k}'') + P_l(k')\int d^3k[V^0_{ljp}(\mathbf{k}',\mathbf{k}) - V_{ljp}(\mathbf{k}',\mathbf{k})]A(\mathbf{k}'-\mathbf{k})\Gamma^s_{pn}(\mathbf{k},\mathbf{k}''). \tag{3.16}$$

Equation (3.15) follows by putting (3.16) on-shell.

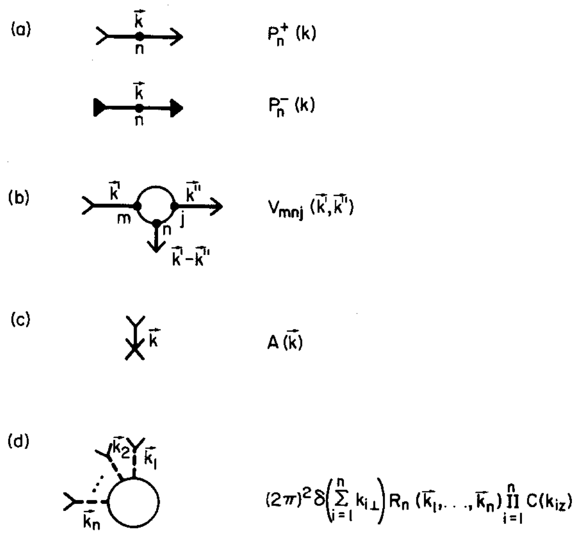

FIG. 2 Diagram rules for the terms in the integral equations for the elastic surface Green's function G^s_{rmn}, its mean, and the mean of its square. They are formally similar to those in the scalar case (Ref. 1). Parts (a), (b), and (c) were introduced in the deterministic discussion and (d) in the statistical part.

Using (3.14) it is possible to define a 3-index Green's function G^D_{npu} via

$$\Gamma^D_{pl}(\mathbf{k}',\mathbf{k}'') = (2\pi)^3\delta(\mathbf{k}'-\mathbf{k}'')\Gamma^0_{pl}(\mathbf{k}') + (2\pi)^3 P_n(k')G^D_{npu}(\mathbf{k}',\mathbf{k}'')\Gamma^0_{ul}(\mathbf{k}''). \quad (3.17)$$

Substituting (3.17) and (3.8) into (3.14) yields the relation

$$G^D_{rmu}(\mathbf{k}',\mathbf{k}'') = V^0_{rmu}(\mathbf{k}',\mathbf{k}'')A(\mathbf{k}'-\mathbf{k}'') + \int d^3k V^0_{rmp}(\mathbf{k}',\mathbf{k})A(\mathbf{k}'-\mathbf{k})P_l(k)G^s_{lpu}(\mathbf{k},\mathbf{k}'') \quad (3.18)$$

which expresses G^D in terms of G^s. The coordinate space representation of (3.17) is

$$\Gamma^{D+}_{pl}(\mathbf{x}',\mathbf{x}'') = \Gamma^0_{pl}(|\mathbf{x}'-\mathbf{x}''|) + (2\pi)^3 \times \iint d^3x_1 d^3x_2 P^+_n(|\mathbf{x}'-\mathbf{x}_1|) \times G^{D+}_{npu}(\mathbf{x}_1,\mathbf{x}_2)\Gamma^0_{ul}(|\mathbf{x}_2-\mathbf{x}''|). \quad (3.19)$$

Hence the outgoing (0) scattered field $\psi^{(0)}_p(\mathbf{x}')$ can thus be expressed in terms of the incident (i) field $\psi^{(i)}_u(\mathbf{x})$ as

$$\psi^{(0)}_p(\mathbf{x}') = (2\pi)^3 \iint d^3x_1 d^3x_2 P^+_n(|\mathbf{x}'-\mathbf{x}_1|) \times G^{D+}_{npu}(\mathbf{x}_1,\mathbf{x}_2)\psi^{(i)}_u(\mathbf{x}_2), \quad (3.20)$$

where we've dropped the Born term in going from (3.19) to (3.20) because our interest is only in the scattered field. Noting that $G^D_{3pu}=0$ since $V^0_{3pu}=0$, (3.20) can be written

$$\psi^{(0)}_p(\mathbf{x}') = (2\pi)^3 \iint d^3x_1 d^3x_2 [P^+_1(|\mathbf{x}'-\mathbf{x}_1|)G^{D+}_{1pu}(\mathbf{x}_1,\mathbf{x}_2) + P^+_2(|\mathbf{x}'-\mathbf{x}_1|)G^{D+}_{2pu}(\mathbf{x}_1,\mathbf{x}_2)]\,\psi^{(i)}_u(\mathbf{x}_2). \quad (3.21)$$

It is possible to decompose the outgoing scattered field into four components

$$\psi^{(0)}_p(x) = \sum_{\alpha,\beta}\psi^{(0)\alpha,\beta}_p(\mathbf{x}), \qquad \alpha,\beta = t,l, \quad (3.22)$$

where t stands for transverse (shear or S-wave) and l for longitudinal (compressional or P wave). The incident field breaks down into two components, viz.

$$\psi^{(i)}_p(x) = \sum_\gamma \psi^{(i)\gamma}_p(\mathbf{x}), \gamma = t,l, \quad (3.23)$$

corresponding to a transverse or longitudinal incident field. The two superscripts on the outgoing field components, e.g., $\psi^{(0)\alpha,\beta}_p$, indicate the αth outgoing field in terms of the βth incident field.

The expansion of each of the above fields in terms of plane wave fields $\phi^{(0)\alpha,\beta}_p$ and $\phi^{(i)\gamma}$ are given by

$$\psi^{(0)\alpha,\beta}_p(\mathbf{x}) = \int d^2k_\perp e^{i\mathbf{k}\cdot\mathbf{x}}\phi^{(0)\alpha,\beta}_p(k_\perp), \qquad k_z = K_\alpha, \quad (3.24)$$

$$\psi^{(i)\gamma}_p(\mathbf{x}) = \int d^2k'_\perp e^{i\mathbf{k}'\cdot\mathbf{x}}\phi^{(i)\gamma}_p(k'_\perp), \qquad k'_z = -K_\gamma. \quad (3.25)$$

Here the specifications $k_z = K_\alpha$ and $k'_z = -K_\gamma$ are taken to ensure that $\psi^{(0)}$ and $\psi^{(i)}$ are outgoing and incident waves, respectively. If we write the two-dimensional representation

$$P^+_{1,2}(\mathbf{x}) = G^{t,l+}_0(\mathbf{x}) = \pi i(2\pi)^{-3}\int d^2k_\perp \times e^{ik_\perp\cdot x_\perp}e^{i|z|K_{t,l}}K^{-1}_{t,l} \quad (3.26)$$

for the propagators, substitute Eqs. (3.22) through (3.26) into (3.21) in the far field limit, and equate appropriate terms, the result is

$$\phi^{(0)\alpha,\beta}_p(k_\perp) = \int d^2k'_\perp T^{\alpha,\beta+}_{pu}(k_\perp,k'_\perp)\phi^{(i)\beta}_u(k'_\perp), \quad (3.27)$$

where, in general (i.e., for both boundary values)

$$T^{\alpha,\beta\pm}_{pu}(k_\perp,k'_\perp) = \frac{\pi i}{k_z}\iint d^3x_1 d^3x_2 e^{-i\mathbf{k}\cdot\mathbf{x}_1}G^{D\pm}_{\alpha pu}(\mathbf{x}_1,\mathbf{x}_2)e^{i\mathbf{k}'\cdot\mathbf{x}_2} = \frac{\pi i}{k_z}G^{s\pm}_{\alpha pu}(\mathbf{k},\mathbf{k}') \qquad k_z = \pm K_\alpha, \quad k'_z = \mp K'_\beta. \quad (3.28)$$

Note that since we are on-shell in (3.28) we are able to replace the Fourier transform of G^D with that of G^s. We have also used the notation that

$$G^{s\pm}_{\alpha pu} = \begin{cases} G^{s\pm}_{1pu}, & \alpha = t \\ G^{s\pm}_{2pu}, & \alpha = l. \end{cases}$$

Hence using (3.22), (3.24), and (3.27), we can write

$$\psi^{(0)}_p(\mathbf{x}) = \sum_{\alpha,\beta}\iint d^2k_\perp d^2k'_\perp e^{i\mathbf{k}\cdot\mathbf{x}}T^{\alpha,\beta+}_{pu}(k_\perp,k'_\perp)\phi^{(i)\beta}_u(k'_\perp) \quad (3.29)$$

with the on-shell conditions of (3.28).

A similar analysis can be carried out for the complex conjugate fields and the result is

$$\psi^{(0)*}_p(\mathbf{x}) = \sum_{\alpha,\beta}\iint d^2k_\perp d^2k'_\perp e^{-ik_\perp\cdot x_\perp}e^{-izK_\alpha} \times T^{\alpha,\beta-}_{pu}(-k_\perp,-k'_\perp)\phi^{(i)\beta*}_u(k'_\perp). \quad (3.30)$$

The field expansions (3.29) and (3.30) will be used in Sec. 4 in the mutual coherence function and intensity.

4. RANDOM ROUGH SURFACE

Now treat the surface as a centered Gaussian distributed random variable and calculate the first two of the statistical moments of $\Gamma^s_{pl}(\mathbf{k}',\mathbf{k}'')$. From (3.8) these are reduced to calculating the moments of $G^s_{rmn}(\mathbf{k}',\mathbf{k}'')$. This is done via (3.9) whose Born expansion shows that to calculate the mean of G^s, $\langle G^s\rangle$, it is necessary to know how to deal with functionals of the surface having the general form $\langle\Pi^n_{i=1}A(\mathbf{k}_i)\rangle$. These were discussed in Ref. 1 where it was shown that the nth order products could be cluster decomposed as[6]

$$\left\langle\prod_{i=1}^n A(\mathbf{k}_i)\right\rangle = \sum_{j\,\mathrm{perm}}\sum_{M=1}^n\sum_{\{m_i\}_M=n}\prod_{i=1}^n A_{m_i}(\{\mathbf{k}_j\}_{m_i}) \quad (4.1)$$

with $\sum_{\{m_i\}_M}$ equal to the sum over all unordered M-element sets $\{m_i\}_M$ such that $\sum^M_{i=1}m_i = n$ and $\sum_{j\,\mathrm{perm}}$ equal to the sum over all different labelings j of the unordered m_i element sets $\{\mathbf{k}_j\}_{m_i}$ with $j = 1,2,\ldots,n$. It was also shown in Ref. 1 that, with the two-point correlation function

$$\Gamma(x_\perp - x'_\perp) = \langle h(x_\perp)h(x'_\perp)\rangle, \quad (4.2)$$

it was possible to write

$$\left\langle\prod_{i=1}^n A(\mathbf{k}_i)\right\rangle = \exp\left(-\tfrac{1}{2}\Gamma(0)\sum_{m=1}^n k^2_{mz}\right). \times (2\pi)^2\delta\left(\sum_{m=1}^n k_{m\perp}\right)I_n(\mathbf{k}_1,\mathbf{k}_2,\ldots,\mathbf{k}_n), \quad (4.3)$$

where the I_n integral is defined as

$$I_n(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \int\cdots\int d\rho_{1\perp}\cdots d\rho_{(n-1)_\perp} \times \exp\left\{-i\sum_{m=1}^{n-1}\rho_{m\perp}\cdot k_{m\perp} - \sum_{i<j}^{n-1}\Gamma(\rho_{i\perp}-\rho_{j\perp}) \times k_{iz}k_{jz} - \sum_{i=1}^{n-1}\Gamma(\rho_{i\perp})k_{iz}k_{nz}\right\}. \quad (4.4)$$

Further it was also shown that each cluster function A_m could be written as

$$A_m(\mathbf{k}_1,\ldots,\mathbf{k}_m) = (2\pi)^2\delta\left(\sum_{i=1}^{m} k_{i\perp}\right)\prod_{i=1}^{m} C(k_{iz})R_m(\mathbf{k}_1,\ldots,\mathbf{k}_m), \tag{4.5}$$

where the $C(k_z)$ are the characteristic functions

$$C(k_z) = \langle \exp\{-ik_z h(x_\perp)\}\rangle = \exp[-1/2\,\Gamma(0)k_z^2] \tag{4.6}$$

which arise from calculating moments of the A functions, viz. $(R_1 \equiv 1)$

$$A_1(\mathbf{k}) = \langle A(\mathbf{k})\rangle = \int d^2x_\perp e^{-ik_\perp\cdot x_\perp}C(k_z) = (2\pi)^2\delta(k_\perp)C(k_z). \tag{4.7}$$

Using (4.1), (4.3), and (4.5) it is easily seen that the R_m functions can be written in terms of I_m and $R_{m'}$, $m' < m$, or in terms of I_n for $n \le m$. Examples were presented in Ref. 1. As is obvious from our formulation, the statistical problem here is the same as in Ref. 1.

Using the partial summation technique discussed in Ref. 1, in problems in random media propagation,[2] and field theory,[7] the mean of $G^s_{\gamma mn}$ can be written as

$$\langle G^{s\pm}_{\gamma mu}(\mathbf{k}',\mathbf{k}'')\rangle = M_{\gamma mu}(\mathbf{k}',\mathbf{k}'') \int d^3k M_{\gamma mn}(\mathbf{k}',\mathbf{k})P^{\pm}_l(k)\langle G^{s\pm}_{lnu}(\mathbf{k},\mathbf{k}'')\rangle. \tag{4.8}$$

The additional diagram notation in the statistical case and the function $M_{\gamma mn}(\mathbf{k}',\mathbf{k}'')$ are shown in Fig. 3. $M_{\gamma mn}$ is the sum of connected diagrams and is the analog of the "mass operator" in random media propagation.[2]

It is possible to factor a transverse delta function out of $M_{\gamma mn}$,

$$M_{\gamma mn}(\mathbf{k}',\mathbf{k}) = \delta(k'_\perp - k_\perp)\mathfrak{M}_{\gamma mn}(k'_z, k_z), \tag{4.9}$$

and hence also out of $\langle G^{s\pm}_{\gamma mn}\rangle$,

$$\langle G^{s\pm}_{\gamma mn}(\mathbf{k}',\mathbf{k}'')\rangle = \delta(k'_\perp - k''_\perp)g^{\pm}_{\gamma mn}(k'_z, k''_z), \tag{4.10}$$

where we have suppressed the transverse momentum dependence of $\mathfrak{M}$ and g. Substituting (4.9) and (4.10) into (4.8) yields the reduced Dyson equation

$$g^{\pm}_{\gamma mn}(k'_z,k''_z) = \mathfrak{M}_{\gamma mn}(k'_z,k''_z) + \int dk_z \mathfrak{M}_{\gamma mp}(k'_z,k_z)P^{\pm}_l(k_z) \times g^{\pm}_{lpn}(k_z,k''_z). \tag{4.11}$$

The full Dyson equation for $\langle\Gamma^{s\pm}_{pl}\rangle$ can be found from (3.8) and (4.8). It is given by

$$\langle\Gamma^{s\pm}_{pl}(\mathbf{k}',\mathbf{k}'')\rangle = (2\pi)^3\delta(\mathbf{k}'-\mathbf{k}'')\Gamma^{0\pm}_l(\mathbf{k}') + P^{\pm}_\gamma(k')\int d^3k M_{\gamma pn}(\mathbf{k}',\mathbf{k})\langle\Gamma^{s\pm}_{nl}(\mathbf{k},\mathbf{k}'')\rangle. \tag{4.12}$$

Equation (4.11) is exact but unwieldy since the Born term and the kernel $\mathfrak{M}$ involve an infinite series of terms. The simplest approximation of this equation is to choose for $\mathfrak{M}$ the first term in the diagram expansion for M in Fig. 3. The approximation neglects correlation effects and is in a sense an average surface approximation. $M_{\gamma mn}$ is given by

$$M_{\gamma mn}(\mathbf{k}',\mathbf{k}) \cong M^{(1)}_{\gamma mn}(\mathbf{k}',\mathbf{k}) = (2\pi)^2\delta(k'_\perp - k_\perp)C(k'_z - k_z)V_{\gamma mn}(\mathbf{k}',\mathbf{k}). \tag{4.13}$$

Substituting (4.13) and (4.9) in (4.11) yields the integral equation

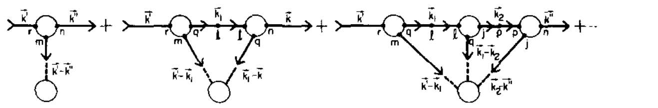

FIG. 3 The connected diagram sum equal to the mass operator $M_{\gamma mn}(\mathbf{k}',\mathbf{k}'')$ in the Dyson equation.

$$g^{\pm(1)}_{\gamma mn}(k'_z,k''_z) = (\pi i)^{-1}C(k'_z - k''_z)R_{\gamma mn3}(k'_z) + (\pi i)^{-1}R_{\gamma mp3}(k'_z)\int dk_z \times C(k'_c - k_z)P^{\pm}_q(k_z)g^{\pm(1)}_{qpn}(k_z,k''_z), \tag{4.14}$$

where

$$P^{\pm}_q(k_z) = \begin{cases} (k_z^2 + k'^2_\perp - k_t^2)^{-1}, & q=1\\ (k_z^2 + k'^2_\perp - k_l^2)^{-1}, & q=2\\ 1, & q=3.\end{cases}$$

Equation (4.14) is the analog of Eq. (42) in the scalar case.[1] This completes the discussion of the first moment.

Next, consider the second moment. The mutual coherence function C_{mn} is defined as[8]

$$C_{mn}(\mathbf{x},\mathbf{x}') = \langle\psi^{(0)}_m(\mathbf{x})\psi^{(0)*}_n(\mathbf{x}')\rangle. \tag{4.15}$$

Using the outgoing scattered fields given by (3.29) and (3.30), this can be written as $(k_z = K_\alpha, k'_z = K'_\gamma)$

$$C_{mn}(\mathbf{x},\mathbf{x}') = \sum_{\alpha,\beta,\gamma,\delta}\iint d^2k_\perp d^2k'_\perp e^{i[\mathbf{k}\cdot\mathbf{x}-\mathbf{k}'\cdot\mathbf{x}']} \times\langle\phi^{(0)\alpha,\beta}_m(k_\perp)\phi^{(0)\gamma,\delta*}_n(k'_\perp)\rangle, \tag{4.16}$$

where

$$\langle\phi^{(0)\alpha,\beta}_m(k_\perp)\phi^{(0)\gamma,\delta*}_n(k'_\perp)\rangle = \iint d^2k''_\perp d^2k'''_\perp\langle T^{\alpha,\beta+}_{mp}(k_\perp,k''_\perp)T^{\gamma,\delta-}_{nq}(-k'_\perp,-k'''_\perp)\rangle \times\phi^{(i)\beta}_p(k''_\perp)\phi^{(i)\delta*}_q(k'''_\perp). \tag{4.17}$$

Note that the products $\langle T^+T^-\rangle$ can be written in terms of the second moment of the G^s functions via (3.28), viz.

$$\langle T^{\alpha,\beta+}_{mp}(k_\perp,k''_\perp)T^{\gamma,\delta-}_{nq}(-k'_\perp,-k'''_\perp)\rangle = \frac{\pi^2}{K_\alpha K'_\gamma}\langle G^{s+}_{\alpha mp}(\mathbf{k},\mathbf{k}'')G^{s-}_{\gamma nq}(-\mathbf{k}',-\mathbf{k}''')\rangle \tag{4.18}$$

with the on-shell restrictions

$$k_z = +K_\alpha, \quad k'_z = +K'_\gamma$$
$$k''_z = -K''_\beta, \quad k'''_z = -K'''_\delta.$$

Again using partial summation techniques and diagram properties, it is possible to write an integral equation for $\langle G^{s+}\cdots G^{s-}\cdots\rangle$ similar to Eq. (38) in Ref. 1. It is given by

$$\begin{aligned}\langle G^{s+}_{\alpha mp}(\mathbf{k},\mathbf{k}_1)G^{s-}_{\gamma nq}(\mathbf{k}',\mathbf{k}'_1)\rangle &= \langle G^{s+}_{\alpha mp}(\mathbf{k},\mathbf{k}_1)\rangle\langle G^{s-}_{\gamma nq}(\mathbf{k}',\mathbf{k}'_1)\rangle K^{\alpha mp}_{\gamma nq}(\mathbf{k},\mathbf{k}'|\mathbf{k}_1,\mathbf{k}'_1)\\ &+ \int d^3k_2 K^{\alpha mu}_{\gamma nq}(\mathbf{k},\mathbf{k}'|\mathbf{k}_2,\mathbf{k}'_1)P^+_\gamma(k_2)\langle G^{s+}_{\gamma up}(\mathbf{k}_2,\mathbf{k}_1)\rangle\\ &+ \int d^3k'_2 K^{\alpha mp}_{\gamma nu}(\mathbf{k},\mathbf{k}'|\mathbf{k}_1,\mathbf{k}'_2)P^-_\gamma(k'_2)\langle G^{s-}_{\gamma uq}(\mathbf{k}'_2,\mathbf{k}'_1)\rangle\\ &+ \iint d^3k_2 d^3k'_2 K^{\alpha mu}_{\gamma nv}(\mathbf{k},\mathbf{k}'|\mathbf{k}_2,\mathbf{k}'_2)P^+_\gamma(k_2)P^-_s(k'_2)\\ &\times\langle G^{s+}_{\gamma up}(\mathbf{k}_2,\mathbf{k}_1)G^{s-}_{svq}(\mathbf{k}'_2,\mathbf{k}'_1)\rangle.\end{aligned} \tag{4.19}$$

This is called the reduced Bethe-Salpeter equation. The function $K_{\gamma pq}^{\alpha mn}$ corresponds to the intensity operator of random volume scattering theory,[2] and is equal to the sum of connected incoherent diagrams shown in Fig. 4. (We use the convention that the superscripts of K are to be summed over in the usual way. These are the only superscripts so summed.) Using (4.19) and (3.8) the full Bethe-Salpeter equation can be written as

$$\begin{aligned}\langle \Gamma_{mn}^{s+}(\mathbf{k},\mathbf{k}_1)\Gamma_{pq}^{s-}(\mathbf{k}',\mathbf{k}_1')\rangle &= \langle \Gamma_{mn}^{s+}(\mathbf{k},\mathbf{k}_1)\rangle\langle \Gamma_{pq}^{s-}(\mathbf{k}',\mathbf{k}_1')\rangle \\ &+ P_r^{+}(k)P_j^{-}(k') \iint d^3k_2 d^3k_2' K_{j\,pv}^{r\,mu}(\mathbf{k},\mathbf{k}'|\mathbf{k}_2,\mathbf{k}_2') \\ &\times \langle \Gamma_{un}^{s+}(\mathbf{k}_2,\mathbf{k}_1)\Gamma_{vq}^{s-}(\mathbf{k}_2',\mathbf{k}_1')\rangle. \qquad (4.20)\end{aligned}$$

It is possible to reduce the integrals in (4.19) and (4.20) by using the translational invariance of the function K to define a new function $\bar{K}$ via

$$\begin{aligned}K_{\gamma pq}^{\alpha mn}(\mathbf{k},\mathbf{k}'|\mathbf{k}_1,\mathbf{k}_1') &= \delta(k_\perp - k_{1\perp} - k_\perp' + k_{1\perp}') \\ &\times \bar{K}_{\gamma pq}^{\alpha mn}(\mathbf{k}+\mathbf{k}',\mathbf{k}-\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1') \qquad (4.21)\end{aligned}$$

The second moments of the Green's functions Γ and G can be similarly factored. Additional simplifications occur by using the on-shell conditions of (4.18), but the general expressions are cumbersome and not very illuminating, and we do not write them. Instead we turn to some simple examples.

For an incident plane wave

$$\phi_p^{(i)\beta}(k_\perp) = \delta(k_\perp - k_{i\perp})\chi_p^\beta, \qquad (4.22)$$

where χ_p^β is a factor which carries the information that $k_{iz} = -K_t(\beta = t)$ or $k_{iz} = -K_l(\beta = l)$ depending on whether the plane wave is put on the transverse or longitudinal shell, respectively (incident stress or longitudinal wave, respectively). Combining (4.16), (4.17), and (4.22), we can write

$$\begin{aligned}&\sum_{\substack{\alpha\beta\\ \gamma\delta}} \langle \phi_m^{(0)\alpha,\beta}(k_\perp)\phi_n^{(0)\gamma,\delta *}(k_\perp')\rangle \\ &= \sum_{\substack{\alpha\beta\\ \gamma\delta}} \langle T_{mp}^{\alpha,\beta+}(k_\perp,k_{i\perp})T_{nq}^{\gamma,\delta-}(-k_\perp',-k_{i\perp})\rangle \chi_p^\beta \chi_q^{\delta *}.\end{aligned}$$

Multiplying this result by δ_{mn} and summing yields a definition of intensity I:

$$\begin{aligned}\delta(k_\perp' - k_\perp)I(k_\perp,k_{i\perp}) &= \sum_{\substack{\alpha\beta\\ \gamma\delta}} \langle T_{mp}^{\alpha,\beta+}(k_\perp,k_{i\perp}) \\ &\times T_{mq}^{\gamma,\delta-}(-k_\perp',-k_{i\perp})\rangle \chi_p^\beta\chi_q^{\delta *}, \qquad (4.23)\end{aligned}$$

where $I(k_\perp, k_{i\perp})$ is the intensity scattered in the $k_\perp$ direction due to an incident plane wave in the $k_{i\perp}$ direction. This becomes using (4.18)

$$\begin{aligned}\delta(k_\perp' - k_\perp)I(k_\perp,k_{i\perp}) &= \sum_{\substack{\alpha\beta\\ \gamma\delta}} \frac{\pi^2}{K_\alpha K_\gamma}\langle G_{\alpha mp}^{s+}(\mathbf{k},\mathbf{k}_i) \\ &\times G_{\gamma mq}^{s-}(-\mathbf{k}',-\mathbf{k}_i)\rangle \chi_p^\beta\chi_q^{\delta *}. \qquad (4.24)\end{aligned}$$

An example of the use of this equation is the calculation of the lowest order coherent contribution to the intensity. It is indicated by the diagram in Fig. 5a. Choosing for simplicity the incident field to have a particular on-shell behavior (either one), the β and δ sums in (4.24) can be done $\left(\sum_\beta \chi_p^\beta = \chi_p\right)$ and the result is

$$\begin{aligned}I(k_\perp,k_{i\perp}) = \delta(k_\perp - k_{i\perp})|A_m C(K_t - k_{iz}) \\ + B_m C(K_l - k_{iz})|^2 \qquad (4.25)\end{aligned}$$

with $k_{iz} = -K_t$ or $-K_l$ depending on the choice of incident field and where A_m and B_m are defined by

$$\begin{aligned}K_t A_m = K_t\chi_m + \chi_p(k_{p\perp} + \delta_{p3}K_t) \\ \times \{\delta_{m3}(1 - 2K_t^2/k_t^2) - 2K_t k_{m\perp}/k_t\}, \qquad (4.26)\end{aligned}$$

$$K_l B_m = (k_{m\perp} + \delta_{m3}K_l)\{\Lambda\chi_3 + 2K_l\chi_p(k_{p\perp} + \delta_{p3}K_l)/k_t^2\}. \qquad (4.27)$$

The intensity can be normalized to $\Gamma(0) = 0$ ($C(k_z) = 1$), and the delta function indicates specular scattering.

A second example is the calculation of the lowest order incoherent intensity arising from the diagram in Fig. 5b. The result is

$$\begin{aligned}I(k_\perp,k_{i\perp}) &= \sum_{\alpha,\gamma} \frac{4\pi^4}{K_\alpha K_\gamma} C(K_\alpha - k_{iz})C(K_\gamma - k_{iz}) \\ &\times V_{\alpha mp}^0(k_\perp, K_\alpha|k_{i\perp},k_{iz})V_{\gamma mq}^0(-k_\perp,-K_\gamma|-k_{i\perp},-k_{iz}) \\ &\times \chi_p\chi_q^* R_2(k_\perp - k_{i\perp}, K_\alpha - k_{iz}|-K_\gamma + k_{iz}), \qquad (4.28)\end{aligned}$$

where $\alpha,\gamma = t, l$ and $k_{iz} = -K_t$ or $-K_l$. The terms $V_{\alpha mp}^0$ are defined as the on-shell version of (3.6) with the obvious expanded notation

$$V_{\alpha mp}^0(\mathbf{k}',\mathbf{k}) = V_{\alpha mp}^0(k_\perp', k_z'|k_\perp,k_z) \qquad (4.29)$$

to indicate the on-shell parameters. The function R_2 was defined in Appendix B of Ref. 1 (and misprinted). It is given by

$$R_2(\mathbf{k}_1,\mathbf{k}_2) = \int d^2y_\perp e^{-ik_{1\perp}\cdot y_\perp}\{\exp[-\Gamma(y_\perp)k_{1z}k_{2z}] - 1\}. \qquad (4.30)$$

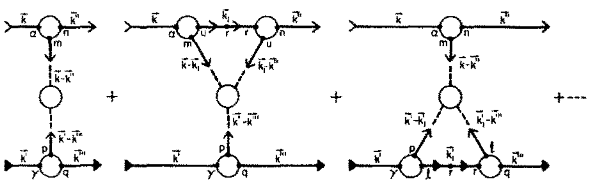

FIG. 4 The connected diagram sum equal to the intensity operator $K_{\gamma pq}^{\alpha mn}(\mathbf{k},\mathbf{k}'|\mathbf{k}'',\mathbf{k}''')$ in the Bethe-Salpeter equation.

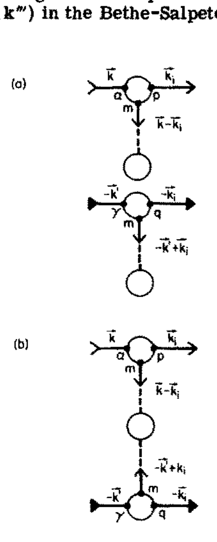

FIG. 5 Lowest order coherent (a) and incoherent (b) contributions to the intensity. Note the sum over m.

Note that it is independent of $k_{2\perp}$ so that the notational expansion can be written

$$R_2(\mathbf{k}_1,\mathbf{k}_2) = R_2(\mathbf{k}_1, k_{2z}) = R_2(k_{1\perp}, k_{1z} \mid k_{2z}). \quad (4.31)$$

5. SUMMARY AND CONCLUSIONS

The matrix Green's function and its first two moments for the isotropic elastic half space bounded by a random rough free boundary have been presented as solutions of integral equations and as series expansions in terms of diagrams. The results can be thought of as corrections to average (flat) surface theories due to the surface randomness. They can be used, e.g., in seismic calculations and as approximations in finite elastic media bounded by random walls. As pointed out in Ref. 1, there are many computational difficulties involved in solving the Dyson and Bethe-Salpeter equations and these difficulties are compounded in this paper by the fact that the given equations are really coupled sets of equations. The main advantage of this presentation is in the systematic approach to higher order corrections. Approximation schemes generated using feedback from experimental results would be necessary to actually use the theory.

Finally we take the opportunity of this paper (see Ref. 9) to correct some misprints in Ref. 1.

ACKNOWLEDGMENTS

I am grateful to John Hayes for a conversation on diagrams, to John McCoy for several remarks about elasticity, and to Frances Vandivier for her careful typing.

APPENDIX. CALCULATION OF THE KERNEL

We wish to calculate the kernel function in (2.10) which is defined by

$$\begin{aligned} N_l(x_\perp)[T'\Gamma^0(\mathbf{x}',\mathbf{x}_s)]_{lij} &= \mu\,\delta_{lm}\{\partial'_m\Gamma^0_{ij}(\mathbf{x}',\mathbf{x}_s) + \partial'_i\Gamma^0_{mj}(\mathbf{x}',\mathbf{x}_s)\} \\ &\quad + \lambda\delta_{li}\partial'_m\Gamma^0_{mj}(\mathbf{x}',\mathbf{x}_s) \end{aligned} \quad (A1)$$

where we have used (2.8) for the traction operator. Using (2.4) we have

$$\begin{aligned} \Gamma^0_{ij}(\mathbf{x}',\mathbf{x}_s) &= \mu^{-1}\delta_{ij}G^t_0(\mathbf{x}',\mathbf{x}_s) \\ &\quad + k_0^{-2}\partial'_i\partial'_j\{G^t_0(\mathbf{x}',\mathbf{x}_s) - G^l_0(\mathbf{x}',\mathbf{x}_s)\} \end{aligned} \quad (A2)$$

and thus to find (A1) we must calculate terms like $\partial'_j G^{t,l}_0(\mathbf{x}',\mathbf{x}_s)$ and $\partial'_m\partial'_i\partial'_j\{G^t_0(\mathbf{x}',\mathbf{x}_s) - G^l_0(\mathbf{x}',\mathbf{x}_s)\}$. The former terms were discussed in Appendix of Ref. 1 where it was shown that

$$\begin{aligned} \partial'_j G^{t,l}_0(\mathbf{x}',\mathbf{x}_s) &= \frac{i}{(2\pi)^3}\int d^3k\,G^{t,l}_0(k) \\ &\quad\times\left[k_{j\perp} + \delta_{j3}P\left(\frac{K^2_{t,l}}{k_3}\right)\right] e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}_s)} \\ &\quad - \tfrac{1}{2}\delta_{j3}\epsilon(z'-h(x_\perp))\delta(x'_\perp - x_\perp), \end{aligned} \quad (A3)$$

where P stands for the Cauchy principle value and $K^2_{t,l} = k^2_{t,l} - k^2_\perp$. The point of the calculation is to subtract off the terms which are singular when we pass to the surface limit in constructing the integral equation (2.12). A similar straightforward and lengthy calculation can be performed on the latter term and, when combined with (A3), the result can be written as [note that $\partial_m\Gamma^0_{mj} = (\lambda+2\mu)^{-1}\partial_j G^l_0$]

$$\begin{aligned} \partial'_m\Gamma^0_{ij}(\mathbf{x}',\mathbf{x}_s) &= -\tfrac{1}{2}\delta(x'_\perp - x_\perp)\delta(z'-h(x_\perp)) \\ &\quad\times[\mu^{-1}\delta_{ij}\delta_{m3} + k_0^{-2}(k_l^2 - k_t^2)\delta_{i3}\delta_{j3}\delta_{m3}] \\ &\quad + \frac{i}{(2\pi)^3}\int d^3k\,e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}_s)}\Bigg\{\frac{\delta_{ij}}{\mu}G^t_0(k) \\ &\quad\times\left[k_{m\perp} + \delta_{m3}P\left(\frac{K_t^2}{k_3}\right)\right] - k_0^{-2} \\ &\quad\times[G^t_0(k) - G^l_0(k)][k_ik_jk_m - k_3k^2\delta_{i3}\delta_{j3}\delta_{m3}] \\ &\quad - k_0^{-2}\delta_{i3}\delta_{j3}\delta_{m3}[k_t^2K_t^2G^t_0(k) \\ &\quad - k_l^2K_l^2G^l_0(k)]P(1/k_3)\Bigg\}. \end{aligned} \quad (A4)$$

Substituting (A4) in (A1) yields after another lengthy calculation

$$\begin{aligned} N_l(x_\perp)&[T'\Gamma^0(\mathbf{x}',\mathbf{x}_s)]_{lij} \\ &= -\tfrac{1}{2}\delta(x'_\perp - x_\perp)\epsilon(z'-h(x_\perp)) \\ &\quad\times\{\delta_{ij} - \delta_{i3}\partial_{j\perp}h(x_\perp) - \Lambda\delta_{j3}\partial_{i\perp}h(x_\perp)\} - R_{ji}(\mathbf{x}',\mathbf{x}_s), \end{aligned} \quad (A5)$$

where

$$R_{ji}(\mathbf{x}',\mathbf{x}_s) = (2\pi)^{-3}\int d^3k\,e^{i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}_s)}R_{ji}(\mathbf{k},\mathbf{x}_s) \quad (A6)$$

and

$$R_{ji}(\mathbf{k},\mathbf{x}_s) \equiv R_{jim}(\mathbf{k})N_m(x_\perp), \quad (A7)$$

$$\begin{aligned} iR_{jim}(\mathbf{k}) &= G^t_0(k)\delta_{ij}\left[k_{m\perp} + \delta_{m3}P\left(\frac{K_t^2}{k_3}\right)\right] \\ &\quad + G^t_0(k)\delta_{jm}\left[k_{i\perp} + \delta_{i3}P\left(\frac{K_t^2}{k_3}\right)\right] \\ &\quad - 2k_ik_jk_m[G^t_0(k) - G^l_0(k)]/k_t^2 \\ &\quad + \Lambda\delta_{im}\left[k_{j\perp} + \delta_{j3}P\left(\frac{K_l^2}{k_3}\right)\right]G^l_0(k) \\ &\quad + 2\,\frac{k_t^2 - k_l^2}{k_t^2}P\left(\frac{1}{k_3}\right)\delta_{i3}\delta_{j3}\delta_{m3}. \end{aligned} \quad (A8)$$

Note that, on the energy shell (see Sec. 3) where $K_t^2 = k_3^2$ and $K_l^2 = k_3^2$ we have that

$$iR^0_{jim}(\mathbf{k}) = \mu[k_m\Gamma^0_{ij}(\mathbf{k}) + k_i\Gamma^0_{jm}(\mathbf{k})] + \Lambda\delta_{im}k_jG^l_0(k). \quad (A9)$$

This is the result of direct differentiation without any singularities being present. Equation (A8) can be further rewritten as

$$iR_{jim}(\mathbf{k}) \equiv P_n(k)R_{njim}(\mathbf{k}), \quad (A10)$$

where we have introduced the vector propagator $P_n(k)$

$$P_n(k) = \begin{cases} G^t_0(k), & n=1 \\ G^l_0(k), & n=2 \\ 1, & n=3 \end{cases} \quad (A11)$$

and where

$$R_{1jim}(\mathbf{k}) = \delta_{ij}\left[k_{m\perp} + \delta_{m3}P\left(\frac{K_t^2}{k_3}\right)\right] - 2\,\frac{k_ik_jk_m}{k_t^2}$$

$$+\ \delta_{jm}\left[k_{i\perp} + \delta_{i3}P\left(\frac{K_t^2}{k_3}\right)\right], \quad \text{(A12a)}$$

$$R_{2jim}(\mathbf{k}) = \Lambda\delta_{im}\left[k_{j\perp} + \delta_{j3}P\left(\frac{K_l^2}{k_3}\right)\right] + 2\,\frac{k_ik_jk_m}{k_t^2}, \quad \text{(A12b)}$$

$$R_{3jim}(\mathbf{k}) = 2\delta_{i3}\delta_{j3}\delta_{m3}\left(\frac{k_t^2 - k_l^2}{k_t^2}\right)P\left(\frac{1}{k_3}\right). \quad \text{(A12c)}$$

On-shell we have that

$$R^0_{1jim}(\mathbf{k}) = k_m\delta_{ij} + k_i\delta_{jm} - 2k_ik_jk_m/k_t^2\,, \quad \text{(A13a)}$$

$$R^0_{2jim}(\mathbf{k}) = \Lambda\delta_{im}k_j + 2k_ik_jk_m/k_t^2, \quad \text{(A13b)}$$

$$R^0_{3jim}(\mathbf{k}) = 0. \quad \text{(A13c)}$$

[1]G. G. Zipfel, Jr. and J. A. DeSanto, J. Math. Phys. **13**, 1903 (1972).
[2]U. Frisch, "Wave Propagation in Random Media," in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
[3]V. D. Kupradze, *Potential Methods in the Theory of Elasticity* (Davey, New York, 1965).
[4]K. M. Case and R. M. Hazeltine, J. Math. Phys. **11**, 2546 (1970).
[5]F. C. Karal and J. B. Keller, J. Math. Phys. **5**, 537 (1964).
[6]K. Huang, *Statistical Mechanics* (Wiley, New York, 1963), Chap, 14.
[7]P. Roman, *Introduction to Quantum Field Theory* (Wiley, New York, 1969).
[8]M. J. Beran, "Propagation of Radiation in Random Media" (Lecture Notes, University of Pennsylvania, 1971).
[9]There are three misprints in Ref. 1 which are corrected here. In Eq. (10) of Ref. 1 the Green's function appearing next to the delta function is $G_0^\pm$ rather than $G_D^\pm$. In the kernel of Eq. (42) the principle value term is $(k''_z)^{-1}$ rather than $(k'')^{-1}$. Finally in the last few lines of Appendix B, the function $R_2(\mathbf{k}_1, \mathbf{k}_2)$ is slightly incorrect, and the correct version is presented in Eq. (4.30) of this paper.

Existence of generalized translation operators from the Agranovitch–Marchenko transformation (Jost solutions)

Marcel Coz

Department of Physics and Astronomy, University of Kentucky, Lexington, Kentucky 40506

Christiane Coudray

*Institut de Physique Nucleaire, Groupe I, 91 Orsay, France**

(Received 25 September 1972; final revised manuscript received 16 November 1972)

The A and M transformation for finding an integral equation for the kernel of a generalized translation operator is adapted to the s-wave regular solution. Its extension to higher l-values is then considered for Jost solutions. The integral equations for the G.T.O. kernels are similar to the s wave one, with the difference that the Riemann function for the l-wave harmonic partial differential equation has to be introduced. As a consequence the condition $\int_{(x+y)/2}^{\infty} s^l |V(s)| ds < \infty, (x+y)>0$, must be satisfied for the l wave, if one wants a continuous kernel.

1. INTRODUCTION

Only problems related to solutions defined as plane waves when r goes to infinity (Jost solutions) are treated in this paper. The nonrelativistic inverse problem solution, the only kind of inverse problem considered in this paper, is related to the existence of an integral representation of a solution for a Schrödinger equation E_1 (with interaction V_1) in terms of a solution for a second Schrödinger equation E_2 (with interaction V_2). The two solutions are specified by the same kind of boundary condition. In what follows, this integral representation is referred to as a generalized translation. The operator for the integral representation is a generalized translation operator (G.T.O.).[1]

Let us consider the case when the boundary conditions are defined by the behavior of the solutions at infinity. The integral representation is

$$\psi_1(x) = \psi_2(x) + \int_x^{\infty} K_{12}(x,y)\psi_2(y)dy. \tag{1}$$

Formal differentiations under the integral sign of Eq. (1), together with integrations by parts, lead to a hyperbolic partial differential equation for the G.T.O. kernel K_{12}. If Coulomb forces are disregarded, as will be done in this paper, one obtains

$$\left(\frac{\partial^2}{\partial x^2} + k^2 - \frac{l(l+1)}{x^2} - V_1(x)\right) K_{12}(x,y) = \left(\frac{\partial^2}{\partial y^2} + k^2 - \frac{l(l+1)}{y^2} - V_2(y)\right) K_{12}(x,y), \tag{2a}$$

$$\lim_{y\to\infty} K_{12}(x,y) = 0, \tag{2b}$$

$$2\frac{d}{dx} K_{12}(x,x) = V_2(x) - V_1(x), \tag{2c}$$

$$K_{12}(x,y) = 0, \quad x > y. \tag{2d}$$

Solving Eq. (2) means specifying the conditions the potentials V_1, V_2 must satisfy to secure the existence of K_{12}. The standard method consists in defining new variables related to the characteristics of Eq. (2):

$$\xi = x + y, \quad \eta = x - y,$$

$$K_{12}(\xi,\eta) = \int_{\eta}^{\xi} d\alpha \int_{\xi}^{\infty} d\beta \Big(l(l+1)[(\alpha+\beta)^{-2} - (\alpha-\beta)^{-2}] + \tfrac{1}{4}[V_1(\alpha+\beta) - V_2(\alpha-\beta)]\Big) K_{12}(\alpha,\beta) + \tfrac{1}{2}\int_{\xi}^{\infty} d\beta [V_2(\beta) - V_1(\beta)]. \tag{3}$$

The solution of Eq. (3) is obtained by iteration. This standard method has a severe shortcoming since the double integral contains a kernel which is singular when l is greater than zero. Although this inconvenience can be remedied using the Riemann method, this way of proceeding is not considered here. Rather, we will apply a transformation due to Agranovitch and Marchenko[2] and applied by them to the s-wave irregular solution known as the Jost solution.[3]

We will refer to the method as the A and M transformation. The A and M transformation offers a compact form for the resolvent of the Schrödinger equation as we shall see in Sec. 3.

We have studied the p-wave case in detail and shown how the extension to higher l wave has to be constructed, and in so doing have noticed that requirements specifying the behavior of the potentials depend on the value of l.

Section 2 is concerned with the A and M transformation and its use for the s-wave regular solution.[4] An extension of the irregular solution when the variable extends from $+\infty$ to $-\infty$ is given so that problems at fixed energy[5] can be studied; this allows a simple proof of Loeffel's theorem and an extension of it for the case where the reference potential is Coulombian.

Section 3 deals with the p-wave problem, the extension to higher waves, as well as the l-dependence requirement.

2. THE A AND M TRANSFORMATION

The A and M transformation is extensively developed in Ref. 2, so it is unnecessary to recall details given there. The case of the irregular solution considered by the previous authors, as well as its variant for the regular solution, is only summarized. An extension to the case where the variable runs from $+\infty$ to $-\infty$ is given for later use.

A. Irregular solutions

A and M Theorem: If $\sigma_1(x) < \infty, x > 0$, then the Schrödinger equation

$$\left[\frac{d^2}{dx^2} + k^2 - V(x)\right] f(k,x) = 0, \quad \lim_{x\to\infty} f(k,x) = e^{ikx}$$

has a solution

$$f(k,x) = e^{ikx} + \int_x^{\infty} K(x,y) e^{iky} dy,$$

Copyright © 1973 by the American Institute of Physics

where the function $K(x,y)$ satisfies

$$|K(x,y)| \leq \tfrac{1}{2}\sigma_0[(x+y)/2]\exp[\sigma_1(x)]. \quad (2.1)$$

In this theorem $\sigma_i = \int_0^\infty t^i |V(t)|dt$.

Outline of Proof: The integral equation for $f(k,x)$

$$f(k,x) = e^{ikx} - \frac{1}{k}\int_x^\infty \sin k(x-s)V(s)f(k,s)ds, \quad (2.2)$$

and its possible representation

$$f(k,x) = e^{ikx} + \int_x^\infty K(x,s)\,e^{iks}\,ds, \quad (2.3)$$

are considered together.

Following Ref. 2, one obtains

$$K(x,y) = \tfrac{1}{2}\int_{(x+y)/2}^\infty V(s)ds + \tfrac{1}{2}\int_x^{(x+y)/2} V(s)ds \times \int_{y+x-s}^{y+s-x} K(s,u)du + \tfrac{1}{2}\int_{(x+y)/2}^\infty V(s)ds\int_s^{y+s-x} K(s,u)du, \quad (2.4)$$

from which the bound (2.1) is derived.

From Eq. (2.4), as well as from Eq. (2.3), one can obtain the partial differential equation

$$\left(\frac{\partial^2}{\partial x^2} - V(x)\right)K(x,y) = \frac{\partial^2}{\partial y^2}K(x,y),$$

$$\lim_{x+y\to\infty} K(x,y) = \lim_{x+y\to\infty}\frac{\partial}{\partial y}K(x,y) = 0,$$

$$K(x,x) = \tfrac{1}{2}\int_x^\infty V(s)ds. \quad (2.5)$$

When the conditions of the theorem are satisfied, a G.T.O. exists. Extension to two potentials V_1 (reference potential), V_2 and two Schrödinger equations is immediate. Conditions are

$$\int_x^\infty t^i|V_j(t)|dt < \infty, \quad i = 0,1, \quad j = 1,2.$$

The bound for the kernel K_{12} is obtained from those of K_{10}, K_{02}.

In order to prepare Sec. 2C, we extend the equations

$$f(k,s) = e^{ikx} - \int_x^\infty \sin k(x-s)\,\frac{1}{k}\,V(s)f(k,s)ds,$$

$$f(k,x) = e^{ikx} + \int_x^\infty K(x,s)\,e^{iks}\,ds$$

to negative values of x with the restriction $s \geq x$.

For any positive x, one has

$$\int_{-x}^\infty |t-x|\ |V(t)|dt = \int_{-x}^x |t-x|\ |V(t)|dt + \int_{-x}^\infty |t-x|\ |V(t)|dt \leq \int_{-x}^x x|V(t)|dt + \int_x^\infty t|V(t)|dt.$$

For the existence of the G.T.O., one must now require the two inequalities

$$\int_x^\infty t|V(t)|dt < \infty, \quad \int_{-x}^x |V(t)|dt < \infty. \quad (2.6)$$

B. Regular solution

In Ref. 2 (p. 18) a bound is given for the regular solution

$$\psi(k,x) = \frac{1}{k}\sin kx + \int_0^x \frac{1}{k}\sin k(x-s)V(s)\psi(k,s)ds, \quad (2.7)$$

$$|\psi(k,x)e^{-ikx}| \leq x\exp\int_0^x t|V(t)|dt. \quad (2.8)$$

The existence of an integral representation

$$\psi(k,x) = \frac{1}{k}\sin kx + \int_0^x K(x,y)\frac{1}{k}\sin ky\,dy \quad (2.9)$$

can be obtained from (2.8). However, our interest lies in the integral equation satisfied by the kernel $K(x,y)$ of Eq. (2.9) so we proceed otherwise.

Comparison of (2.7) and (2.9) gives

$$\int_0^x K(x,s)\frac{1}{k}\sin ks\,ds = \int_0^x \frac{1}{k}\sin k(x-s)V(s)\frac{1}{k}\sin ks\,ds + \int_0^x \frac{1}{k}\sin k(x-s)V(s)ds\int_0^s K(s,u)\frac{1}{k}\sin ku\,du = J_1 + J_2, \quad (2.10)$$

$$J_1 = \tfrac{1}{2}\int_0^x \frac{\sin kt}{k}dt\int_{(x-t)/2}^{(x+t)/2} V(s)ds, \quad (2.11)$$

$$J_2 = \tfrac{1}{2}\Bigg(\int_0^x \frac{\sin kt}{k}dt\int_0^{x-t} V(s)ds\int_{x-s-t}^{t-x+s} K(s,u)du + \int_0^x V(s)ds\int_{t+x-s}^s K(s,u)du - \int_{(x+t)/2}^x V(s)ds \times \int_{t+x-s}^s K(s,u)du - \int_0^{(x-t)/2} V(s)ds\int_{x-s-t}^s K(s,u)du\Bigg). \quad (2.12)$$

An inverse sine transform provides the integral equation for $K(s,y)$:

$$K(x,y) = \tfrac{1}{2}\int_{(x-y)/2}^{(x+y)/2} V(s)ds + \tfrac{1}{2}\int_{x-y}^x V(s)ds\int_{y-x+s}^s \times K(s,u)du + \tfrac{1}{2}\int_{(x-y)/2}^{x-y} V(s)ds\int_{x-s-y}^s K(s,u)du - \tfrac{1}{2}\int_{(x+y)/2}^x V(s)ds\int_{y+x-s}^s K(s,u)du, \quad \text{for } x \geq y. \quad (2.13)$$

When $y = x$, Eq. (2.11) reduces to

$$K(x,x) = \tfrac{1}{2}\int_0^x V(s)ds.$$

We define

$$\int_0^x s^i|V(s)|ds = \sigma_i(x)$$

and solve for $K(x,y)$ by iteration. The following bounds are then obtained:

$$|K_0(x,y)| \leq \tfrac{1}{2}[\sigma_0((x+y)/2) - \sigma_0((x-y)/2)] \leq \tfrac{1}{2}\sigma_0(x),$$

$$K_n(x,y) \leq \tfrac{1}{2}\sigma_0(x)\frac{1}{n!}[8\sigma_1(x)]^n,$$

that is,

$$|K(x,y)| \leq \tfrac{1}{2}\sigma_0(x)\exp 8\sigma_1(x). \quad (2.14)$$

Theorem: If

$$\sigma_0(x) = \int_0^x |V(t)|dt < \infty,$$

a G.T.O. for regular solutions exists. Its kernel is a continuous function.

The theorem can be extended to secure the existence of a G.T.O. between two Schrödinger equations with potentials V_1, V_2 satisfying the condition of the latter theorem.

Insertion of estimate (2.14) into Eq. (2.13) shows that in the limit when y goes to zero, $K(x,y)$ vanishes.

C. Problems at fixed energy

This subsection is an application of subsection 2A. To apply the G.T.O. theory[1] to fixed energy problems,[5] one must consider the differential operators

$$r^2\left(\frac{\partial^2}{\partial r^2} - \frac{l(l+1)}{r^2} - V(r) + k^2\right),$$
$$r^2\left(\frac{\partial^2}{\partial r^2} - \frac{l(l+1)}{r^2} - V_0(r) + k^2\right) \tag{2.15}$$

and the integral representation

$$\psi_l(k,r) = \psi_l^0(k,r) + \int_0^r K(r,s)s^{-2}\psi_l^0(k,s)ds. \tag{2.16}$$

We let $V_0 = 0$ in (2.15) for reasons of simplicity. The kernel $K(r,s)$ satisfies a partial differential equation that we will not specify. Instead, we set $kr = e^{-u}$, $ks = e^{-v}$, $A(u,v) = e^{-1/2(u+v)} = K(u,v)$ and derive the equation for $A(u,v)$

$$\left(\frac{\partial^2}{\partial u^2} + e^{-2u}[1 - V(e^{-u})]\right) A(u,v) = \left(\frac{\partial^2}{\partial v^2} + e^{-2v}\right) A(u,v), \tag{2.17}$$

$$\lim_{v\to\infty} A(u,v) = \lim_{v\to\infty}\frac{\partial}{\partial v} A(u,v) = 0,$$
$$\frac{d}{du} A(u,u) = -\tfrac{1}{2}e^{-2u}V(e^{-u}).$$

Equations (2.17) and (2.5) with two potentials are identical when one defines

$$V_1 = e^{-2u}[1 - V(e^{-u})], \quad V_2 = e^{-2u} \tag{2.18}$$

and allows the variable to extend from $+\infty$ to $-\infty$.

The following conditions are, therefore, required for the existence of $K(r,s)$. For $r < 1$

$$\int_0^r |1 - V(t)|t|\ln t|^i dt < \infty, \quad i = 0, 1,$$

and for $r > 1$

$$\int_0^1 |1 - V(t)|t|\ln t|^i dt < \infty, \quad i = 0, 1,$$
$$\int_1^x t|1 - V(t)|dt < \infty \tag{2.19}$$

According to (2.19) the potential $V(t)$ must be less singular at the origin than the centrifugal barrier. Condition (2.19) is identical to the one obtained by Loeffel[6] who uses properties of the Laplace transform and obtains

$$\int_0^1 t^{1-2\epsilon}|1 - V(t)|dt < \infty. \tag{2.20}$$

It is similar to the one obtained by Kelemen[7] although less restrictive, since there is no bound on the extension of V for complex t:

$$|V(x + iy)| = |V(\mathrm{Re}^{i\theta})| \le Mr^{2\gamma-2}(1 + |\theta|^k), \tag{2.21}$$

with $\gamma > 0$, $k < 1/2$.

When $V_0 = \eta/r$, as in Ref. 1, one defines

$$V_1 = e^{-2u}[1 - V(e^{-u})] + \eta e^{-u}, \quad V_2 = e^{-2u} + \eta e^{-u}.$$

There is no modification on the condition that V must satisfy for the existence of a G.T.O.

3. EXTENSION OF THE A AND M TRANSFORMATION TO HIGHER PARTIAL WAVES

Both the A and M[2] or the Gel'fand–Levitan[4] results are restricted to potentials with a regularity property, regularity at infinity for A and M, regularity at the origin for G.L. One may want to know how the results can be extended when either the centrifugal barrier or the Coulomb repulsion are included in the reference potential. In this paper, attention is limited to the centrifugal barrier. In addition, we limit ourselves to solutions defined by their behavior at infinity; the case of solutions defined by a property of regularity at the origin can be treated without any difficulty. The p-wave case occupies parts A and B of this section, higher waves in part C.

A. Integral equation (p wave)

We use the following definitions:

$$h(kx) = e^{ikx}[1 - (1/ikx)],$$

$$G_0(k,x,s) = -\frac{\sin k(s-x)}{k} + \frac{1}{ksx}\frac{d}{dx}\frac{\sin k(s-x)}{k}.$$

The integral equation for the irregular solution is

$$f(k,x) = h(kx) - \int_x^\infty G_0(k,x,s)V(s)f(k,s)ds. \tag{3.1}$$

Equation (3.1) can be solved by iteration if[3]

$$\int_0^\infty s|V(s)|ds < \infty. \tag{3.2}$$

In other words, one can express $f(k,x)$ as an infinite series or as

$$f(k,x) = h(kx) - \int_x^\infty G(k,x,s)V(s)h(ks)ds, \tag{3.3}$$

Where $G(k,x,s)$ is the resolvent for the Schrödinger equation with potential $V(x)$.

Equation (3.3) suggests the search for a compact form

$$f(k,x) = h(kx) + \int_x^\infty K(x,y)h(ky)dy. \tag{3.4}$$

If $K(x,y)$ exists, it must satisfy a partial harmonic differential equation[8] with the boundary conditions

$$\lim_{y\to\infty} K(x,y) = 0,$$
$$\frac{d}{dx}K(x,x) = -\tfrac{1}{2}V(x). \tag{3.5}$$

Comparison of eqs. (3.1) and (3.4) gives

$$\int_x^\infty K(x,s)e^{iks}[1 - (1/iks)]ds = \int_x^\infty V(s)G_0(k,x,s)ds \times e^{iks}[1 - (1/iks)] + \int_x^\infty V(s)ds \times \int_s^\infty G_0(k,s,u)K(s,u)du\, e^{iku}[1 - (1/iku)]. \tag{3.6}$$

Both sides of Eq. (3.6) are multiplied by the differential operator D,

$$D = \frac{1}{ik}\frac{d}{dk}k,$$

and by the integral operator,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iky}dk.$$

By so doing, one gets

$$\text{l.h.s.} = K(x,y)y, \qquad \text{r.h.s.} = J_1 + J_2.$$

The definition of $G_0(k,x,s)$ and the integration over k yield

$$J_1 = \tfrac{1}{2}\int_{(x+y)/2}^{\infty} V(s)\,ds\,A(s,s), \tag{3.7}$$

$$\begin{aligned}J_2 = \tfrac{1}{2}\int_x^{(x+y)/2} V(s)\,ds \int_{x+y-s}^{\infty} K(s,u)A(s,u)\,du\\ + \tfrac{1}{2}\int_{(x+y)/2}^{\infty} V(s)ds\int_s^{\infty} K(s,u)A(s,u)du\\ - \tfrac{1}{2}\int_x^{\infty} V(s)\,ds \int_{x+y-s}^{\infty} K(s,u)B(s,u)\,du,\end{aligned} \tag{3.8}$$

$$A(s,u) = B(s,u) \equiv R(u,s)y,$$

$$R = 1 + (1/8usxy)(u+s-x-y)(s-u+y-x) \times[(u+y)^2-(s+x)^2]. \tag{3.9}$$

From $R(u,s)$ one gets $R(s,s)$. The integral equation for $K(x,y)$ becomes

$$\begin{aligned}K(x,y) = \tfrac{1}{2}\int_{(x+y)/2}^{\infty} V(s)R(s,s)\,ds\\ + \tfrac{1}{2}\int_x^{(x+y)/2} V(s)ds\int_{x+y-s}^{y+s-x} K(s,u)R(u,s)du\\ + \tfrac{1}{2}\int_{(x+y)/2}^{\infty} V(s)ds\int_s^{y+s-x} K(s,u)R(u,s)du.\end{aligned} \tag{3.10}$$

Equation (3.10) reduces to Eq. (2.4) when

$$R(u,s) = R(s,s) \equiv 1.$$

B. Existence theorem

We proceed by associating bounds with $K_0, K_1, \cdots, K_n$:

$$\begin{aligned}|K_0| \le \tfrac{1}{2}\int_{(x+y)/2}^{\infty}|V(s)|ds + \tfrac{1}{2}\int_{(x+y)/2}^{\infty}|V(s)|ds(1/8s^2x)\\ \times(2s-x-y)(y-x)[(s+y)^2-(s+x)^2]\\ \le \tfrac{1}{2}\sigma_0[(x+y)/2] + (18/16x)\int_{(x+y)/2}^{\infty} s|V(s)|ds.\end{aligned} \tag{3.11}$$

A fortiori one has

$$|K_0| \le (3/x)\sigma_1[(x+y)/2]. \tag{3.12}$$

We consider the R-function in the domains D_1, D_2 defined by the double integrals

$$D_1: \int_{(x+y)/2}^{\infty} ds \int_s^{y+s-x} du, \qquad D_2: \int_x^{(x+y)/2} ds\int_{y+x-s}^{y+s-x} du.$$

In the domain D_1

$$|R(u,s)| \le 1 + (4s/x), \tag{3.13}$$

while in the domain D_2,

$$|R(u,s)| \le 1 + (8s/x). \tag{3.14}$$

We choose

$$|R| \le 1 + (8s/x). \tag{3.15}$$

With this estimate the following bounds are obtained:

$$\begin{aligned}|K_1| &\le (3/x)\sigma_1[(x+y)/2]5\sigma_1(x),\\ |K_2| &\le (3/x)\sigma_1[(x+y)/2](5\sigma_1(x))^2\frac{1}{2},\\ |K_n| &\le (3/x)\sigma_1[(x+y)/2](5\sigma_1(x)^n(1/n!),\\ |K| &\le (3/x)\sigma_1[(x+y)/2]\exp 5\sigma_1(x).\end{aligned} \tag{3.16}$$

Theorem: If $\sigma_0(x)$, $\sigma_1(x)$ exist, a G.T.O. can be defined for an $l = 1$ wave. Its kernel is bounded by the estimate (3.16). For other properties relative to the p-wave, see Ref. 9.

C. Extension to higher waves

Although estimates (2.1) and (3.16) require the same asymptotic moments from the potential, their structures are different. It is, therefore, interesting to consider the $l = 2$ wave case.

We obtained the function $R(u,s;x,y)$ by

$$\frac{1}{y}R(u,s;y,x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iky}\frac{1}{ik}\frac{d}{dk}[G_0(k,s,x)h(ku)k]dk. \tag{3.17}$$

Integration by parts of Eq. (3.20)

$$R(u,s;y,x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(-ky)G_0(k,s,x)h(ku)dk, \tag{3.18}$$

which is precisely the definition of the Riemann function defined for the harmonic equation by Chaundy[10] and Copson[11]:

$$\left(\frac{d}{dy^2}-\frac{2}{y^2}\right)R(y,x) = \left(\frac{d^2}{dx^2}-\frac{2}{x^2}\right)R(y,x).$$

Defining $\alpha = x+y$, $\beta = x-y$,

$$x_1 = -\frac{(\alpha-A)(\beta-B)}{(\alpha+\beta)(A+B)}, \qquad x_2 = \frac{(\alpha-A)(\beta-B)}{(\alpha-\beta)(A-B)}$$

one gets

$$R(\alpha,\beta;A,B) = P_1(1-2x_1-2x_2+2x_1x_2), \tag{3.19}$$

or in our notations

$$P_1\left(1+\frac{1}{8usxy}(u+s-x-y)(s-u+y-x) \times[(u+y)^2-(s+x)^2]\right) \tag{3.20}$$

Instead of using the operator D, we may choose to write Eq. (3.10) directly by introducing the proper Riemann function.

We follow this method for higher partial waves. For $l = 2$, for instance, we have

$$\begin{aligned}K(x,y) = \tfrac{1}{2}\int_{(x+y)/2}^{\infty} V(s)ds\,R_2(s,s)\\ + \tfrac{1}{2}\int_{(x+y)/2}^{\infty} V(s)ds\int_s^{y+s-x} K(s,u)duR_2(u,s)\\ + \tfrac{1}{2}\int_x^{(x+y)/2} V(s)ds\int_{y-s+x}^{y+s-x} K(s,u)duR_2(u,s).\end{aligned} \tag{3.21}$$

Writing (3.20) as $P_1(X)$

$$X = 1 + \frac{1}{8usxy}(u+s-x-y) \times(s-u+y-x)[(u+y)^2-(s+x)^2],$$

$$R_2(u,s;x,y) = P_2(X) = \tfrac{1}{2}(3X^2-1). \tag{3.22}$$

We have now

$$|K_0| \le \tfrac{1}{2}\int_{(x+y)/2}^{\infty} |V(s)|\,ds\,R_2(s,s).$$

The recursion relation for the Legendre polynomials

$$P_{l+1}(X) = [(2l+1)/(l+1)]XP_l(X) - (l/(l+1))P_{l-1}(X)$$

is used, as well as the estimates

$$X(s,s) \le 1 + (4s/x), \qquad X(u,s) \le 1 + (8s/x), \tag{3.23}$$

to obtain a bound for $K_0(x,y)$:

$$|K_0(x,y)| \le \tfrac{1}{2}\int_{(x+y)/2}^{\infty} |V(s)|\,ds[2 + (12s/x) + (24s^2/x^2)]. \tag{3.24}$$

Since $s > x$,

$$|K_0(x,y)| \leqslant (20/x^2)\sigma_2\left[\frac{(x+y)}{2}\right], \tag{3.25}$$

$$|K_1(x,y)| \leqslant 20\sigma_2\left(\frac{(x+y)}{2}\right)\int_x^{\infty} s\,|V(s)|\,ds \times\left(2 + \frac{24s}{x} + \frac{96x^2}{x^2}\right)\frac{1}{2s^2} \leqslant (20/x^2)\sigma_2((x+y)/2)61\sigma_1(x).$$

The same method as before gives

$$|K(x,y)| \leqslant (20/x^2)\sigma_2\left[\frac{(x+y)}{2}\right]\exp 61\sigma_1(x) \tag{3.26}$$

A condition on the existence of a moment of order two is included in the upper bound for $K(x,y)(l=2)$. It must be clear that the recursion relation between the Legendre polynomials introduces a condition on the moment of order l for an l-wave.

However, the conditions obtained are sufficient conditions and exceptions may exist of potentials not satisfying the theorems and susceptible of furnishing a continuous kernel. In addition to this remark, there may exist kernels which are simply "generalized functions."[12]

Nevertheless, it is interesting to compare our results to others obtained by Faddeev in a related are.[13] Faddeev's theorem is stated as follows: if $S(k)$ is the S-function for the Schrödinger operator, $L^{(l)}$, with the potential $g(x)$, it is also the S-function for the Schrödinger operator, $L^{(m)}$, where the corresponding potential $g^{(m)}(x)$ behaves like $g(x)$ as $x \to 0$ and as $x \to \infty$. The class of potentials considered in Faddeev's theorem are such that

$$\int_0^{\infty} x|g(x)|\,dx < \infty.$$

Comparison with Faddeev's theorem shows there is room for solutions of Abranovitch–Marchenko equations which are not continuous functions, and demonstrates, in addition, the interest of problems related to solutions defined by their behavior when r approaches zero.

*Laboratoire associé au C.N.R.S.

[1]B. M. Levitan, *Generalized Translation Operators* (IPST, Jerusalem, 1961); C. Coudray and M. Coz, Ann. Phys. (N.Y.) **61**, 488 (1970).

[2]L. S. Agranovitch and V. A. Marchenko, *The Inverse Problem of Scattering Theory* (Gordon and Breach, New York, 1963); see also Ref. 1a, p. 93.

[3]R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951); R. G. Newton, J. Math. Phys. **1**, 319 (1960); *Scattering of Waves and Particles* (McGraw-Hill, New York, 1966).

[4]N. Dunford and J. T. Schwartz, *Linear Operators* (Wiley, New York, 1963), Vol. II, Chap. XIII, p. 1622; I. M. Gel'fand and B. M. Levitan, Am. Math. Soc. Transl. Ser. 2 1 (1960).

[5]T. Regge, Nuovo Cimento **14**, 1251 (1950).

[6]J. J. Loeffel, Ann. Inst. Henri Poincare **8** (4), 339 (1968).

[7]P. J. Kelemen, J. Math. Phys. **12**, 1142 (1971).

[8]G. Darboux, *Lecons sur la théorie des surfaces* (Gauthier-Villars, Paris, 1896), Vol. II.

[9]H. Cornille, J. Math. Phys. **8**, 2268 (1967); H. Cornille and P. Rubinstein, J. Math. Phys. **9**, 1501 (1968).

[10]T. W. Chaundy, Quart. J. Math. Oxford **10**, 310 (1939).

[11]E. T. Copson, Arch. Ration. Mech. Anal. **1**, 324 (1958).

[12]L. Schwartz, *Théories des distributions* (Hermann, Paris, 1966), new ed., Vols, 1,2.

[13]L. D. Faddeev, J. Math. Phys. **4**, 98 (1963). Faddeev considers only the reductions of an l-problem to an (l-1)-problem. The same subject is treated in Ref. 2, p. 153. See also M. Crumm, Quart. J. Math. Oxford **6** (2), 121 (1955); M. G. Krein, Dokl. Akad. Nauk SSSR **113** (5), 970 (1957). Another author has demonstrated interest in this reduction: A. D. Corbella, J. Math. Phys. **12**, 9 (1971).

Application of Kraichnan's direct interaction approximation to kinematic dynamo theory. II. Incompressible, helical velocity turbulence and a pair of coupled, singular, nonlinear integral equations

I. Lerche

Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, Illinois
(Received 9 April 1973)

Using Kraichnan's direct interaction approximation, we set up the equations governing the normal modes of the ensemble average magnetic field under incompressible, nonmirror symmetric velocity turbulence. We show that (i) the Green's stress tensor enjoys equipartition of its symmetric and antisymmetric parts at the normal mode frequencies of the ensemble average field, (ii) for static velocity turbulence, including helicity, there are no growing modes, (iii) the commonly used first order smoothing theory approximation is invalid when compared to the Kraichnan equations, for the Kraichnan equations do not satisfy Hammerstein's theorem while first order smoothing theory requires the satisfaction of Hammerstein's theorem, (iv) if there is to be any growth of the ensemble average magnetic field it must come from time dependent velocity turbulence, and when the velocity turbulence is time dependent we have so far been unable to solve the Kraichnan equations. We have done these calculations for two reasons. First to illustrate, by exact solution, the manner in which the normal modes of the ensemble average magnetic field depend on the helicity and Reynolds number of the turbulent velocity field. Second to show that approximate treatments of the hydromagnetic equation (like first order smoothing theory), rather than exact solution, are liable to give rise to substantial error in view of the fact that the Kraichnan equations do not satisfy Hammerstein's theorem.

I. INTRODUCTION

In a previous paper Ref. 1, hereinafter referred to as L1, we reported calculations of kinematic dynamo activity using Kraichnan's[2] direct interaction approximation (DIA). The calculations were done for an infinite homogeneous medium which possessed an incompressible, isotropic, mirror-symmetric, homogeneous, stationary, turbulent velocity field. We pointed out in L1 that under Kraichnan's DIA the true turbulence problem is replaced by a model, or models, which lead *without approximation* to *closed* expressions for the ensemble average Green's function and magnetic field. Further it has been shown (Kraichnan,[2]; Frisch[3]) that the model turbulence problems represent physically realizable situations so that one is guaranteed that the results make physical sense. [One is *not* guaranteed *a priori* that the model turbulence problem results are representative of the ensemble system that nature provides, but Frisch has shown that the model problem results are indeed approximate solutions of the true turbulence problem.]

We also pointed out that the Kraichnan equations are valid for arbitrary values of the parameters involved (in particular the magnetic Reynolds number), unlike approximation methods that have been used to date (Krause and Roberts,[4] Lerche,[5] Krause[6]) whose validity has *not* been established in general, but whose invalidity *has* been established in particular cases (Kraichnan, Lerche and Parker,[7]).

Altogether then, Kraichnan's DIA provides a powerful method of investigating the average properties of turbulent systems for arbitrarily large values of any parameters involved.

In this paper we intend to extend the work reported in L1 to a more general form of velocity turbulence than the incompressible, isotropic velocity turbulence used in L1.

The motivation behind this extension is two-fold. First for incompressible, mirror-symmetric, isotropic and *static* velocity turbulence we showed that the normal modes of the ensemble average magnetic field were degenerative under a large class of turbulent velocity fields (see L1 for further details). The question arises: Can helical, velocity turbulence produce normal modes which grow? And, if so, what are the criteria for growth (e.g., particular dependences of the helical turbulence on frequency, or levels of the helical term exceeding critical values, etc.)?

Second, there has arisen a debate (Krause and Roberts, Lerche[8]) concerning the way to handle various turbulence problems. We proved in L1 that the so-called first order smoothing theory (Krause and Roberts) is not a uniformly convergent expansion of the Kraichnan DIA equations and Kraichnan has also remarked on this point. Accordingly, the DIA equations have to be solved exactly, for any approximate solution will give rise to a substantial error according to Hammerstein's[9] theorem. The question arises: Is the same sort of nonuniformity present when the velocity turbulence includes a helical contribution?

For these, and other, reasons we believe that the problem investigated here is of more than academic interest.

Kraichnan has given the general method of obtaining the model DIA equations from the true turbulence problem, and we refer the interested reader to his elegant, and excellent, paper for an appreciation of the details of the method.

In the interest of brevity we shall quote here only those results which are pertinent to our particular problem.

II. BASIC EQUATIONS

Consider an infinite medium possessing a constant resistivity η devoid of any large scale velocity field but possessing a turbulent velocity field $\mathbf{v}(\mathbf{x}, t)$ which has zero mean. The behavior of a magnetic field, $\mathbf{b}(\mathbf{x}, t)$, in such a medium is governed by the induction equation

 Copyright © 1973 by the American Institute of Physics

$$\frac{\partial b_i}{\partial t} - \eta \nabla^2 b_i = \epsilon_{ijk}\epsilon_{klm}\frac{\partial}{\partial x_j}(v_l b_m), \tag{1}$$

with $\partial b_i/\partial x_i = 0$.

Under Kraichnan's direct interaction approximation the equations describing the evolution of the ensemble average magnetic field, $\mathbf{B}(\mathbf{x}, t)$, and the ensemble average Green's tensor, G, are

$$\frac{\partial B_i}{\partial t} - \eta\nabla^2 B_i = \epsilon_{ijk}\epsilon_{klm}\epsilon_{aJK}\epsilon_{KLM}\frac{\partial}{\partial x_j}\int_{-\infty}^{t} dt' d^3\mathbf{x}' \times \left(G_{ma}(\mathbf{x}, t|\mathbf{x}', t')\frac{\partial}{\partial x'_J}[U_{lL}(\mathbf{x}, t|\mathbf{x}', t')B_M(\mathbf{x}', t')]\right) \tag{2}$$

and

$$\left(\frac{\partial}{\partial t} - \eta\nabla^2\right)G_{iu}(\mathbf{x}, t|\mathbf{x}', t') = \delta_{iu}\delta(\mathbf{x}-\mathbf{x}')\delta(t-t') + \epsilon_{ijk}\epsilon_{klm}\epsilon_{abK}\epsilon_{KLM}\frac{\partial}{\partial x_j}\int_{t'}^{t} dt'' d^3\mathbf{x}'' \times \left(G_{ma}(\mathbf{x}, t|\mathbf{x}'', t'')\frac{\partial}{\partial x''_b}[U_{lL}(\mathbf{x}, t|\mathbf{x}'', t'')G_{Mu}(\mathbf{x}'', t''|\mathbf{x}', t')]\right), \tag{3}$$

where $U_{lL}(\mathbf{x}, t|\mathbf{x}', t') = \langle v_l(\mathbf{x}, t)v_L(\mathbf{x}', t')\rangle$,

with $G_{ij}(\mathbf{x}, t|\mathbf{x}', t') = 0$ in $t < t'$;

since only forward-going (in time) Green's functions are physically permissible.

For homogeneous, stationary velocity turbulence (and we shall restrict our attention to just this form of turbulence for the remainder of the paper), we have

$$U_{lL}(\mathbf{x}, t|\mathbf{x}', t') = U_{lL}(\mathbf{x}-\mathbf{x}', t-t'). \tag{4}$$

Then by inspection of Eq. (3) we see that the Green's tensor must be homogeneous and stationary, so that

$$G_{ij}(\mathbf{x}, t|\mathbf{x}', t') = G_{ij}(\mathbf{x}-\mathbf{x}', t-t'). \tag{5}$$

It then follows that with

$$B_i(\mathbf{x}, t) = \int d^3\mathbf{k}d\omega B_i(\mathbf{k}, \omega)\exp[i(\mathbf{k}.\mathbf{x} - \omega t)]. \tag{6}$$

and:

$$[U_{ij}(\mathbf{r}, \tau), G_{ij}(\mathbf{r}, \tau)] = \int d^3\mathbf{k}d\omega[U_{ij}(\mathbf{k}, \omega), G_{ij}(\mathbf{k}, \omega)] \times \exp[i(\mathbf{k}.\mathbf{r} - \omega\tau)], \tag{7}$$

we obtain from Eqs. (2) and (3) the expressions

$$(\eta k^2 - i\omega)B_i(\mathbf{k}, \omega) = -(2\pi)^4\int d^3\mathbf{K}d\Omega[\![k_j K_b B_b(\mathbf{k}, \omega) \times [-U_{ia}(\mathbf{K}, \Omega)G_{ja}(\mathbf{k}-\mathbf{K}, \omega-\Omega) + U_{ja}(\mathbf{K}, \Omega)G_{ia}(\mathbf{k}-\mathbf{K}, \omega-\Omega)] + k_j k_b B_a(\mathbf{k}, \omega)[U_{jb}(\mathbf{K}, \Omega)G_{ia}(\mathbf{k}-\mathbf{K}, \omega-\Omega) - U_{ib}(\mathbf{K}, \Omega)G_{ja}(\mathbf{k}-\mathbf{K}, \omega-\Omega)]]\!] \tag{8}$$

and

$$(\eta k^2 - i\omega)G_{iu}(\mathbf{k}, \omega) = (2\pi)^{-4}\delta_{iu} - (2\pi)^4\int d^3\mathbf{K}d\Omega k_j \times (k_b - \mathbf{K}_b)[\![U_{ia}(\mathbf{K}, \Omega)G_{ja}(\mathbf{k}-\mathbf{K}, \omega-\Omega)G_{bu}(\mathbf{k}, \omega) + U_{jb}(\mathbf{K}, \Omega)G_{ja}(\mathbf{k}-\mathbf{K}, \omega-\Omega)G_{au}(\mathbf{k}, \omega) - U_{ib}(\mathbf{K}, \Omega)G_{ja}(\mathbf{k}-\mathbf{K}, \omega-\Omega)G_{au}(\mathbf{k}, \omega) - U_{jb}(\mathbf{K}, \Omega)G_{ia}(\mathbf{k}-\mathbf{K}, \omega-\Omega)G_{bu}(\mathbf{k}, \omega)]\!], \tag{9}$$

where use has been made of $k_i B_i(\mathbf{k}, \omega) = 0$ to eliminate some of the terms in Eq. (8).

Note that Eq. (8) is linear in **B**. So a solution to it exists if, and only if, a dispersion relation is satisfied. Our task is to obtain that dispersion relation and see if it possesses any growing modes. If so we then have regenerative kinematic dynamo action under kinematic velocity turbulence. In order to obtain the dispersion relation from Eq. (8) we must do two things: first, we must specify the tensor form of $U_{ij}(\mathbf{k}, \omega)$; second, we must then solve Eq. (9) *exactly* for $G_{iu}(\mathbf{k}, \omega)$. Armed with this information we can then substitute for U_{ij} and G_{iu} in Eq. (8) to obtain the dispersion relation.

III. INCOMPRESSIBLE, HELICAL VELOCITY TURBULENCE

In an infinite medium with no preferred axis the most general form of incompressible velocity turbulence is given through (Batchelor[10])

$$U_{ij}(\mathbf{k}, \omega) = E(k, \omega)(\delta_{ij} - k_i k_j k^{-2}) + i\epsilon_{ij\mu}k_\mu H(k, \omega), \tag{10}$$

so that $k_i U_{ij} = k_j U_{ij} = 0$.

Further, $E(k, \omega) \geq 0$ for all real k and ω by Cramér's[11] theorem; and

$$-E(k, \omega) \leq kH(k, \omega) \leq E(k, \omega), \tag{11}$$

for all real k and ω, again by Cramér's theorem.

We now define the two basic integrals

$$I_{iajb} \equiv \int d^3K d\Omega U_{ia}(\mathbf{K}, \Omega)G_{jb}(\mathbf{k}-\mathbf{K}, \omega-\Omega) \tag{12a}$$

and

$$J_{iajb\lambda} \equiv \int d^3\mathbf{K}d\Omega K_\lambda U_{ia}(\mathbf{K}, \Omega)G_{jb}(\mathbf{k}-\mathbf{K}, \omega-\Omega). \tag{12b}$$

In terms of these integrals, Eqs. (8) and (9) become

$$B_i(\mathbf{k}, \omega)\{\delta_{ia}(\eta k^2 - i\omega) + (2\pi)^{-4}[k_j(J_{jbiba} - J_{ibjba}) + k_j k_b(I_{jbia} - I_{ibja})\} = 0 \tag{13}$$

and

$$G_{iu}(\mathbf{k}, \omega)(\eta k^2 - i\omega) = (2\pi)^{-4}\delta_{iu} - (2\pi)^4 k_j \times [G_{bu}(\mathbf{k}, \omega)k_b(I_{iaja} - I_{jaia}) + G_{au}(\mathbf{k}, \omega)k_b(I_{jbia} - I_{ibja}) - G_{bu}(\mathbf{k}, \omega)(J_{iajab} - J_{jaiab})], \tag{14}$$

where we have used the incompressibility condition $k_i U_{ij} = k_j U_{ij} = 0$ to eliminate several of the terms in Eqs. (8) and (9).

Since the velocity turbulence has no preferred axis, it follows that G_{iu} also has no preferred axis and that G_{iu} must take on the general form

$$G_{iu}(\mathbf{k}, \omega) = R(k, \omega)\delta_{iu} + S(k, \omega)k_i k_u + i\epsilon_{iu\beta}k_\beta F(k, \omega), \tag{15}$$

where R, S, and F are (so far) unknown scalar functions of arguments k, ω.

Upon substituting Eqs. (10) and (15) into Eq. (14) it is a simple, but tedious, matter to show that R, S, and F must satisfy the equations

$$R(k, \omega)[\eta k^2 - i\omega + (2\pi)^4\mathfrak{M}(\mathbf{k}, \omega)] = (2\pi)^{-4} - (2\pi)^4 F(k, \omega)\mathfrak{N}(\mathbf{k}, \omega), \tag{16a}$$

$$S(k, \omega)(\eta k^2 - i\omega) = (2\pi)^{-4}k^{-2}[R(k, \omega)\mathfrak{M}(k, \omega) - F(k, \omega)\mathfrak{N}(k, \omega)], \tag{16b}$$

$$F(k, \omega)[\eta k^2 - i\omega + (2\pi)^4\mathfrak{M}(k, \omega)] = -(2\pi)^4 k^{-2}R(k, \omega)\mathfrak{N}(k, \omega), \tag{16c}$$

where

$$\mathfrak{M}(k,\omega) = \int d^3K d\Omega (k^2 - (\mathbf{k}.\mathbf{K})^2 K^{-2}) \times E(K,\Omega) R(|\mathbf{k}-\mathbf{K}|, \omega-\Omega), \tag{17a}$$

$$\mathfrak{N}(k,\omega) = \int d^3K d\Omega (k^2 - (\mathbf{k}.\mathbf{K})^2 K^{-2}) \times [E(K,\Omega)(k^2+K^2)F(|\mathbf{k}-\mathbf{K}|, \omega-\Omega) + H(K,\Omega)K^2 R(|\mathbf{k}-\mathbf{K}|, \omega-\Omega)]. \tag{17b}$$

It is then a simple, but tedious, matter using (15) with (16a), (16b), and (16c) to show that the dispersion relation obtained from Eq. (8) is given by

$$\eta k^2 - i\omega + (2\pi)^4 \mathfrak{M}(k,\omega) = \pm (2\pi)^4 k^{-1} \mathfrak{N}(k,\omega). \tag{18}$$

So the immediate task before us is to solve equations (16) and then to use the results in Eq. (18) to obtain the complex (in general) frequencies, ω, as functions of the real wave number k at which Eq. (8) is satisfied.

It is clear by inspection of Eq. (16) that S is a "passive" quantity in the equations in the sense that, once R and F are known from Eqs. (16a, 16c, 17), S follows directly; but no knowledge of S is required in order to solve Eqs. (16a), (16c), (17). Further, only R and F are necessary in order to obtain the dispersion relation (18). Accordingly, we can concentrate our attention on Eqs. (16a), (16c), (17), (18) and ignore Eq. (16b) for the moment.

Before delving into the detailed method of solution to Eqs. (16a), (16c), (17) there are a few properties of the equations that can be utilized.

First note that when the dispersion relation (18) is satisfied at some complex ω this implies [from equation (16c)] that

$$kF(k,\omega) = \pm R(k,\omega). \tag{19}$$

But from Eq. (16a) we have

$$\eta k^2 - i\omega + (2\pi)^4 \mathfrak{M}(k,\omega) = \frac{(2\pi)^{-4}}{R} - (2\pi)^4 F(k,\omega)\mathfrak{N}/R, \tag{20}$$

and then Eqs. (16a), (16c), and (18) can only be compatible if, when Eq. (18) is satisfied, we have

$$|R| = \infty = |F|, \quad \text{with } Fk/R = \pm 1. \tag{21}$$

In other words, the dispersion relation (18) for the normal modes of the magnetic field is given by the solution to Eqs. (16a), (16c) together with the constraint (21).

In order to capitalize on the constraint (21), it is useful at this point to write

$$g(k,\omega) = (2\pi)^{-4}/R(k,\omega), \quad f(k,\omega) = (2\pi)^{-4}/F(k,\omega),$$

so that Eqs. (16a) and (16c) become

$$\eta k^2 - i\omega + \mathfrak{M}(k,\omega) = g(k,\omega) - \mathfrak{N}(k,\omega) g(k,\omega)/f(k,\omega) \tag{22}$$

and

$$g(k,\omega)[\eta k^2 - i\omega + \mathfrak{M}(k,\omega)] = -k^{-2} f(k,\omega)\mathfrak{N}(k,\omega), \tag{23}$$

with

$$\mathfrak{M}(k,\omega) = \int d^3K d\Omega E(K,\Omega)[k^2 - (\mathbf{k}.\mathbf{K})^2/K^2] \times / g(|\mathbf{k}-\mathbf{K}|, \omega-\Omega), \tag{24}$$

$$\mathfrak{N}(k,\omega) = \int d^3K d\Omega [k^2 - (\mathbf{k}.\mathbf{K})^2/K^2] \times [E(K,\Omega)(k^2+K^2)/f(|\mathbf{k}-\mathbf{K}|, \omega-\Omega) + K^2 H(K,\Omega)/g(|\mathbf{k}-\mathbf{K}|, \omega-\Omega)]. \tag{25}$$

Then the dispersion relation (18) takes on the constraint form

$$g = 0 = f, \quad gk/f = \pm 1. \tag{26}$$

Before proceeding further with Eqs. (22)-(26) it is convenient to simplify the equations by introducing different variables.

First, note that

$$\mathfrak{M}(k,\omega) = k^2 \int_0^\infty \kappa^4 d\kappa \frac{S_1(k,\kappa,\Omega)}{g(\kappa,\omega-\Omega)} d\Omega, \tag{27a}$$

where

$$S_1(k,\kappa,\Omega) = 2\pi \int_1^{+1} d\mu (1-\mu^2) \times E((k^2+\kappa^2+2k\kappa\mu)^{1/2}, \Omega)(k^2+\kappa^2+2k\kappa\mu)^{-1}, \tag{27b}$$

and that

$$\mathfrak{N}(k,\omega) = k^2 \int_0^\infty \kappa^4 d\kappa d\Omega \left(\frac{S_2(k,\kappa,\Omega)}{f(\kappa,\omega-\Omega)} + \frac{S_3(k,\kappa,\Omega)}{g(\kappa,\omega-\Omega)}\right), \tag{28a}$$

where

$$S_2(k,\kappa,\Omega) = 2\pi \int_{-1}^{+1} d\mu (1-\mu^2)(2k^2+\kappa^2+2k\kappa\mu) \times (k^2+\kappa^2+2k\kappa\mu)^{-1} E((k^2+\kappa^2+2k\kappa\mu)^{1/2}, \Omega) \tag{28b}$$

and

$$S_3(k,\kappa,\Omega) = 2\pi \int_{-1}^{+1} d\mu (1-\mu^2) H((k^2+\kappa^2+2k\kappa\mu)^{1/2}, \Omega) \tag{28c}$$

Then in Eqs. (22)-(25) write

$$g(k,\omega) = \eta k^2 \Phi(k,\omega), \quad f(k,\omega) = \eta k^2 \Psi(k,\omega)$$

to obtain

$$1 - i\omega/(\eta k^2) + \eta^{-2} \int_0^\infty \kappa^2 \frac{S_1(k,\kappa,\Omega) d\kappa d\Omega}{\Phi(\kappa,\omega-\Omega)} = \Phi(k,\omega) - \eta^{-2}\Phi(k,\omega)\Psi(k,\omega)^{-1} \int_0^\infty \kappa^2 d\kappa d\Omega \times [S_2(k,\kappa,\Omega)/\Psi(\kappa,\omega-\Omega) + S_3(k,\kappa,\Omega)/\Phi(\kappa,\omega-\Omega)] \tag{29}$$

and

$$\Phi(k,\omega)[1 - i\omega/(\eta k^2) + \eta^{-2} \int_0^\infty \kappa^2 d\kappa d\Omega S_1(k,\kappa,\Omega)/\Phi(\kappa,\omega-\Omega)] = -(\eta k)^{-2}\Psi(k,\omega) \int_0^\infty \kappa^2 d\kappa d\Omega [S_2(k,\kappa,\Omega)/\Psi(\kappa,\omega-\Omega) + S_3(k,\kappa,\Omega)/\Phi(\kappa,\omega-\Omega)]. \tag{30}$$

The dispersion relation (18) takes on the constraint form

$$\Phi = 0 = \Psi, \quad \Phi k/\Psi = \pm 1. \tag{31}$$

Consider then the situation in which the velocity turbulence is static:

$$E(k,\omega) = E(k)\delta(\omega), \quad E(k) \geq 0, \tag{32}$$

so that, by Cramér's theorem,

$$H(k,\omega) = H(k)\delta(\omega), \tag{33}$$

with

$$-E(k) \leq kH(k) \leq E(k).$$

Then Eqs. (29) and (30) can be written

$$1 - i\omega(\eta k^2)^{-1} + \eta^{-2} \int_0^\infty \kappa^2 J_1(k,\kappa)\Phi(\kappa,\omega)^{-1} d\kappa = \Phi(k,\omega) - \Phi(k,\omega)\eta^{-2}\Psi(k,\omega)^{-1} \int_0^\infty \kappa^2 d\kappa \times [J_2(k,\kappa)\Psi(\kappa,\omega)^{-1} + J_3(k,\kappa)\Phi(\kappa,\omega)^{-1}] \tag{34}$$

and

$$\Phi(k,\omega)[1 - i\omega(\eta k^2)^{-1} + \eta^{-2}\int_0^\infty \kappa^2 J_1(k,\kappa)\Phi(\kappa,\omega)^{-1}d\kappa]$$
$$= -(k\eta)^{-2}\Psi(k,\omega)\int_0^\infty \kappa^2 d\kappa[J_2(k,\kappa)\Psi(\kappa,\omega)^{-1} + J_3(k,\kappa)\Phi(\kappa,\omega)^{-1}], \qquad (35)$$

with

$$J_1(k,\kappa) = 2\pi\int_{-1}^{+1} d\mu(1-\mu^2)E(|\mathbf{k}+\boldsymbol{\kappa}|)|\mathbf{k}+\boldsymbol{\kappa}|^{-2}, \qquad (36a)$$

$$J_2(k,\kappa) = 2\pi\int_{-1}^{+1} d\mu(1-\mu^2)(k^2+|\mathbf{k}+\boldsymbol{\kappa}|^2) \times E(|\mathbf{k}+\boldsymbol{\kappa}|)|\mathbf{k}+\boldsymbol{\kappa}|^{-2}, \qquad (36b)$$

$$J_3(k,\kappa) = 2\pi\int_{-1}^{+1} d\mu(1-\mu^2)H(|\mathbf{k}+\boldsymbol{\kappa}|). \qquad (36c)$$

Now let $E(k)$ be characterized by a scale length L (the correlation length) and an "intensity" L^3v^2, and let $H(k)$ be characterized by the same scale length and "intensity" L^4v^2. Then write $k \to kL$, $\omega \to \eta L^{-2}\omega$ so that in dimensionless form (with $\Phi \to \Phi$, $\Psi \to \Psi/L$) Eqs. (34) and (35) become

$$1 - i\omega k^{-2} + R^2\int_0^\infty \kappa^2 d\kappa J_1(k,\kappa)\Phi(\kappa,\omega)^{-1}$$
$$= \Phi(k,\omega) - R^2\Phi(k,\omega)\Psi(k,\omega)^{-1}\int_0^\infty \kappa^2 d\kappa \times [J_2(k,\kappa)\Psi(\kappa,\omega)^{-1} + J_3(k,\kappa)\Phi(\kappa,\omega)^{-1}] \qquad (36')$$

and

$$\Phi(k,\omega)[1 - i\omega k^{-2} + R^2\int_0^\infty \kappa^2 d\kappa J_1(k,\kappa)\Phi(\kappa,\omega)^{-1}]$$
$$= -\Psi(k,\omega)(R/k)^2\int_0^\infty \kappa^2 d\kappa[J_2(k,\kappa)\Psi(\kappa,\omega)^{-1} + J_3(k,\kappa)\Phi(\kappa,\omega)^{-1}], \qquad (37)$$

with $R = Lv/\eta$ and where the dimensionless dispersion relation is

$$\Phi = 0 = \Psi, \quad \Phi k/\Psi = \pm 1, \qquad (38)$$

with all quantities measured in units of the scale length L and the r.m.s. turbulent velocity v.

The immediate task before us is to solve Eqs. (36) and (37) for Φ and Ψ, and to then look for a common zero of Φ and Ψ as functions of the wave number k, such that $\Phi k/\Psi = \pm 1$ at the common zero of Φ and Ψ. This then determines the frequency as a function of k, at which normal modes of the ensemble average magnetic field exist. We must then see if any of the modes [which vary as $\exp(-i\omega t\eta L^{-2})$] exist with $\mathrm{Im}\,\omega > 0$. If so the magnetic field grows in time and we then have regenerate kinematic dynamo activity.[12]

IV. CONSTRUCTION OF THE DISPERSION RELATION

As in Paper 1, suppose that Φ and Ψ share a simple common zero at $k = m$ and a second simple common zero at $k = M(\neq m)$. Suppose further that Φ and Ψ have no other zeros and that m and M are real. Note that $J_\alpha(k,\kappa) = J_\alpha(-k,\kappa) = J_\alpha(k,-\kappa)$ ($\alpha = 1,2,3$) so that both Φ and Ψ are symmetric under the interchange $k \to -k$. We shall also suppose that the velocity turbulence spectrum [defined through $E(k)$ and $H(k)$] is such that the $J_\alpha(k,\kappa)$ are analytic functions in the finite domains of the complex k and κ planes with essential singularities on the circle at infinity. [We demonstrate in Appendix A that it is possible to choose $E(k)$ and $H(k)$ so that the $J_\alpha(k,\kappa)$ are indeed analytic.]

Then as $M \to \infty$ (m finite) it can be shown (see Appendix A and also Paper 1) that

$$\int_0^\infty \kappa^2 \frac{J_\alpha(k,\kappa)}{\Phi(\kappa,\omega)} d\kappa = i\pi m^2 \frac{J_\alpha(k,m)}{\Phi'(m,\omega)}, \quad (\alpha = 1,2,3), \qquad (39)$$

and

$$\int_0^\infty \kappa^2 \frac{J_2(k,\kappa)}{\Psi(\kappa,\omega)} d\kappa = i\pi m^2 \frac{J_2(k,m)}{\Psi'(m,\omega)}, \qquad (40)$$

where

$$\Phi'(m,\omega) \equiv \partial\Phi(k,\omega)/\partial k|_{k=m}, \quad \Psi'(m,\omega) \equiv \partial\Psi(k,\omega)/\partial k|_{k=m}.$$

Equations (36) and (37) then give

$$1 - i\omega k^{-2} + i\pi R^2 m^2 J_1(k,m)(\Phi')^{-1}$$
$$= \Phi(k,\omega) - i\pi R^2 m^2\Phi(k,\omega)\Psi(k,\omega)^{-1} \times [J_2(k,m)(\Psi')^{-1} + J_3(k,m)(\Phi')^{-1}] \qquad (41)$$

and

$$\Phi(k,\omega)[1 - i\omega k^{-2} + i\pi R^2 m^2 J_1(k,m)(\Phi')^{-1}]$$
$$= -i\pi R^2 m^2\Psi(k,\omega)k^{-2}[J_2(k,m)(\Psi')^{-1} + J_3(k,m)(\Phi')^{-1}]. \qquad (42)$$

Equations (41) and (42) can readily be solved as follows. First set $k = m$ in both Eqs. (41) and (42) to obtain

$$1 - im^{-2}\omega + i\pi R^2 m^2 J_1(m,m)(\Phi')^{-1}$$
$$= -i\pi R^2 m\Lambda[J_2(m,m)(\Psi')^{-1} + J_3(m,m)(\Phi')^{-1}], \qquad (43)$$

where $\Lambda = \pm 1$ as follows from Eq. (38).

Now differentiate Eqs. (41) and (42) with respect to k and then set $k = m$ to obtain

$$\Psi'[1 - i\omega m^{-2} + i\pi R^2 m^2 J_1(m,m)(\Phi')^{-1}$$
$$= -i\pi R^2 m^2\Phi'[J_2(m,m)(\Psi')^{-1} + J_3(m,m)(\Phi')^{-1}] \qquad (44a)$$

and

$$\Phi'[1 - i\omega m^{-2} + i\pi R^2 m^2 J_1(m,m)(\Phi')^{-1}]$$
$$= -i\pi R^2\Psi'[J_2(m,m)(\Psi')^{-1} + J_3(m,m)(\Phi')^{-1}]. \qquad (44b)$$

From Eqs. (44a) and (44b) we obtain either

Type 1

$$\Psi' = \Lambda m\Phi', \qquad (45a)$$

$$\Phi' = -i\pi R^2(1 - i\omega m^{-2})^{-1}[m^2 J_1(m,m) + J_2(m,m) + \Lambda m J_3(m,m)], \qquad (45b)$$

or

Type 2

$$\Psi' = -\Phi' J_2(m,m)/J_3(m,m), \qquad (46a)$$

and

$$\Phi' = -i\pi R^2 m^2 J_1(m,m)(1 - i\omega m^{-2})^{-1}. \qquad (46b)$$

But we can also compute Φ' in a different manner. Eliminate $\Psi(k,\omega)$ from Eq. (41) by using Eq. (42) to obtain

$$D_0(k,\omega)^2 = D_0(k,\omega)\Phi(k,\omega) + (-i\pi R^2 m^2 k^{-1})^2 \times [J_2(k,m)(\Psi')^{-1} + J_3(k,m)(\Phi')^{-1}]^2, \qquad (47)$$

where

$$D_0(k,\omega) = 1 - i\omega k^{-2} + i\pi R^2 m^2 J_1(k,m)(\Phi')^{-1}. \qquad (48)$$

Now differentiate Eq. (47) with respect to k to obtain

$$2D_0(k,\omega)D_0'(k,\omega) = D_0'(k,\omega)\Phi(k,\omega) + D_0(k,\omega)\Phi'(k,\omega) + (-i\pi R^2m^2)^2[[-2k^{-3}[J_2(k,m)/\Phi' + J_3(k,m)/\Phi']^2 + 2k^{-2}[J_2(k,m)/\Phi' + J_3(k,m)/\Phi'] \times [J_2'(k,m)/\Phi' + J_3'(k,m)/\Phi']]]. \qquad (48')$$

Consider then the application of the Type 1 and Type 2 relations to Eq. (48).

Type 1 derivatives

Here $D_0(m,\omega) \neq 0$ and $J_2(m,m)/\Psi' + J_3(m,m)/\Phi' \neq 0$, so when Eq. (45a) is substituted into Eq. (48) we obtain

$$\begin{aligned}2[1 - i\omega m^{-2} &+ i\pi R^2m^2J_1(m,m)/\Phi'] \times [2i\omega m^{-3} + i\pi R^2m^2J_1'(m,m)/\Phi'] \\ &= \Phi'[1 - i\omega m^{-2} + i\pi R^2m^2J_1(m,m)/\Phi'] \\ &\quad - 2m^{-3}(-i\pi R^2m^2/\Phi')^2[J_2(m,m)/\Lambda m + J_3(m,m)]^2 \\ &\quad + 2m^{-2}(-i\pi R^2m^2/\Phi')^2[J_2(m,m)/(\Lambda m) + J_3(m,m)] \\ &\quad \times [J_2'(m,m)/(\Lambda m) + J_3'(m,m)] \qquad (48)\end{aligned}$$

where

$$J_\alpha'(m,m) \equiv (\partial/\partial k)J_\alpha(k,m)|_{k=m} \equiv J_\alpha'$$

We now substitute for Φ' from Eq. (45b) to obtain the quadratic equation

$$y^2[2m^2J_1 + J_2 + \Lambda mJ_3 + m^3J_1' + mJ_2' + \Lambda m^2J_3'] - 2y + (i\pi/2)R^2m = 0, \qquad (50)$$

with

$$1 - i\omega m^{-2} = (m^2J_1 + J_2 + \Lambda mJ_3)y. \qquad (51)$$

For a given Λ Eq. (50) admits of the two solutions

$$fy = 1 \pm (1 - \tfrac{1}{2}i\pi R^2mf)^{1/2}, \qquad (52)$$

where

$$f = 2m^2J_1 + J_2 + m^3J_1' + mJ_2' + \Lambda m(J_3 + mJ_3'). \qquad (53)$$

From Appendix B we have that

$$\begin{aligned}I_0 &\equiv m^2J_1 + J_2 + \Lambda mJ_3 \\ &= \pi\int_{-1}^{+1} d\mu(1-\mu)E(2^{1/2}m(1+\mu)^{1/2}) \\ &\quad \times [4 + 2\mu + 2^{1/2}\Lambda(1+\mu)^{1/2}\alpha(2^{1/2}m(1+\mu)^{1/2})], \qquad (54)\end{aligned}$$

where $\alpha(x) = xH(x)/E(x)$ with $|\alpha| \leq 1$ Cramér's theorem.

Since $E(k) > 0$ it follows by inspection of Eq. (54) that $I_0 > 0$. Accordingly, the signature of the real and imaginary parts of $1 - i\omega m^{-2}$ is given by the signature of the real and imaginary parts of y.

Also from Appendix B we have that

$$\begin{aligned}f &\equiv \pi\int_{-1}^{+1} d\mu E(2^{1/2}m(1+\mu)^{1/2}) \\ &\quad \times [3 + 5\mu + 4\mu^2 + 2^{3/2}\Lambda\mu(1+\mu)^{1/2} \\ &\quad \times \alpha(2^{1/2}m(1+\mu)^{1/2})]. \qquad (55)\end{aligned}$$

We shall return in a moment to the Type 1 dispersion relation (52). It is opportune at this point to consider the Type 2 dispersion relation.

Type 2 derivatives

Here $D_0(m,\omega) = 0$ and $J_2/\Psi' + J_3/\Phi' = 0$ so that if we substitute Eqs. (46) into Eq. (49) (with $k = m$), we obtain the identity statement $0 = 0$.

However, if we differentiate Eq. (48) once more with respect to k and then set $k = m$, we obtain

$$[D_0'(m,\omega)]^2 = \Phi'D_0'(m,\omega) + (-i\pi R^2m)^2(J_2'/\Psi' + J_3'/\Phi')^2. \qquad (56)$$

Now when Eqs. (46) are used in Eq. (56) to eliminate Φ', we obtain the cubic equation

$$u^3[h^2 - (mJ_1J_2)^{-2}(J_2J_3' - J_3J_2')^2] - 4m^{-1}hu^2 + u(4m^{-2} - i\pi R^2m^2J_1h) + 2i\pi R^2mJ_1 = 0, \qquad (57)$$

where $u = 1 - i\omega/m^2$, $h = 2m^{-1} + J_1'/J_1$.

Equation (57) has three roots and hence gives three possible frequencies.

Altogether, then, allowing for the fact that $\Lambda = \pm 1$, Eqs. (52) and (57) indicate the existence of *seven* possible modes of propagation. We must select from this profusion of modes those that are physically permissible.

V. PHYSICALLY ACCEPTABLE MODES

There are selection processes which winnow out the physically unacceptable modes of Eqs. (52) and (57). First, as $R \to 0$ only those modes which reduce to $i\omega = m^2$ are acceptable, for $R \to 0$ corresponds to the absence of any turbulent velocity field in which case the modes must be just the free-decay modes of a resistive medium.

It might be thought that if the helicity in the velocity field is zero the modes should reduce to the isotropic modes calculated elsewhere (Paper 1). This is, however, not completely correct. To emphasize the point, consider the relative contributions to $G_{iu}(k,\omega)$ [Eq. (15)] at dispersion [i.e., when the dispersion relation Eq. (18) is satisfied]. Then at dispersion G_{iu} is proportional to

$$R(\delta_{iu} - k_ik_uk^{-2} \pm i\epsilon_{iu\beta}k_\beta k^{-1}), \qquad (58)$$

no matter how the velocity turbulence is chosen [i.e., Eq. (58) is *generally* true and not just for the static turbulence for which we have been able to obtain specific representations of the dispersion relation]. Thus at dispersion

$$G_{iu}G_{iu}^* = |R|^2(2+2), \qquad (59)$$

where the first (second) factor 2 [in Eq. (59) arises from the symmetric (antisymmetric] part of G_{iu}. In other words there is *equipartition* of the Green's tensor stress for each Fourier mode at dispersion. And this result is independent of the level of the helical velocity turbulence—provided it is nonzero.

The case of completely isotropic turbulence is a singular limit of the Kraichnan DIA equations in this respect.[13] We shall return to this point again (Sec. VII). For the moment we content ourselves by noting that the reduction of the dispersion relations to the isotropic results in the absence of helicity *cannot* be used to single out physically acceptable modes.

If we apply the selection rule that, as $R \to 0$, $i\omega \to m^2$, then out of the profusion of modes we provisionally select three:

(A) From Eq. (51) we have the two modes

$$1 - i\omega m^{-2} = f^{-1}(m^2 J_1 + J_2 + \Lambda m J_3) \times [1 - (1 - \tfrac{1}{2} i\pi R^2 m f)^{1/2}], \quad (60)$$

where

$$f = 2m^2 J_1 + J_2 + m^3 J_3' + m J_2' + \Lambda m (J_3 + m J_3'), \quad (61)$$

and $\Lambda = \pm 1$.

(B) From Eq. (57) we have the single mode

$$1 - i\omega m^{-2} = 2(mh)^{-1} y, \quad (62)$$

where

$$y = \gamma + q - \tfrac{1}{2}\gamma q^{-1}(1 - is - 2\gamma), \quad (63a)$$

$$\gamma = \tfrac{2}{3}(1 - \epsilon^2)^{-1}, \qquad s = \tfrac{1}{4}\pi R^2 m^4 J_1 h, \quad (63b)$$

$$\epsilon = (m J_1 J_2 h)^{-1}(J_2 J_3' - J_3 J_2'), \quad (63c)$$

and

$$2^{1/3} q = -\left[\!\left[\tfrac{3}{2}\gamma[\gamma(1 - \tfrac{4}{3}\gamma) + is(1-\gamma)] - \gamma\{\tfrac{9}{4}[\gamma(1 - \tfrac{4}{3}\gamma) + is(1-\gamma)]^2 + \tfrac{1}{2}\gamma(1 - 2\gamma - is)^3\}^{1/2}\right]\!\right]^{1/3}. \quad (63d)$$

An alternative way of writing the cubic is

$$y^3(1 - \epsilon^2) - 2y^2 + y(1 - is) + is = 0. \quad (63e)$$

The cube root in Eq. (63d) is to be chosen so that

$$q = -\tfrac{1}{12} - \tfrac{1}{4}(1 + 4is)^{1/2} + \tfrac{1}{2}6^{-1/2} \times [-(1 + 2is) + (1 + 4is)^{1/2}]^{1/2}, \qquad \epsilon \to 0, \quad (64a)$$

$$q = \tfrac{1}{2}[-\gamma + (2\gamma)^{1/2}(1 - \tfrac{3}{2}\gamma)^{1/2}], \qquad s \to 0, \quad \epsilon \text{ finite} \quad (64b)$$

and

$$q = -\tfrac{1}{3}, \qquad s \to 0, \qquad \epsilon \to 0. \quad (64c)$$

In order to obtain a growing mode in Eq. (62) a *necessary* requirement is that $\text{Re}y > 0$, i.e.,

$$l \equiv \gamma + \text{Re}q - \tfrac{1}{2}\gamma |q|^{-2}[(1 - 2\gamma)\,\text{Re}q - s\,\text{Im}q] > 0. \quad (65)$$

In order to test whether it is possible to have $l > 0$, it would appear that we must solve Eq. (63d) for the real and imaginary parts of q. However, we shall demonstrate that this is not so, and that there is a simple way to see if $l > 0$.

Type A Modes

In order to obtain unstable modes from Eq. (60) we require that f [written out explicitly in Eq. (55)] should change sign; for we have already shown that $m^2 J_1 + J_2 + \Lambda m J_3 > 0$ for any finite m [Eq. (54)] and the signature of the real and imaginary parts of $1 - i\omega m^{-2}$ is therefore proportional to the signature of f.

In order that f change sign it is necessary, but not sufficient, that

$$q' \equiv 3 + 5\mu + 4\mu^2 + 2^{3/2}\Lambda\mu\alpha(2^{1/2}m(1+\mu)^{1/2})(1+\mu)^{1/2}, \quad (66)$$

change sign from positive to negative in $-1 \leqslant \mu \leqslant 1$, with $-1 \leqslant \alpha \leqslant 1$, $\Lambda = \pm 1$.

If we can show that $\beta > 0$, where

$$\beta \equiv 3 + 5\mu + 4\mu^2 - 2^{3/2}|\mu|(1+\mu)^{1/2} \quad (67)$$

in $-1 \leqslant \mu \leqslant 1$, then q' is intrinsically positive, and, therefore, f is positive; and then the Type A modes are oscillatory but decaying.

To demonstrate that $\beta > 0$ is relatively easy. For suppose $\beta < 0$ in some domain of μ encompassed by $-1 \leqslant \mu \leqslant 1$.

Then

$$3 + 5\mu + 4\mu^2 < 2^{3/2}|\mu|(1+\mu)^{1/2}. \quad (68)$$

Now $3 + 5\mu + 4\mu^2 > 0$ in $0 \leqslant |\mu| \leqslant 1$.

Therefore, by squaring both sides of inequality (68) we obtain the requirement that $\beta < 0$ as

$$16\mu^4 + 32\mu^3 + 41\mu^2 + 30\mu + 9 < 0. \quad (69)$$

Now in $1 \geqslant \mu \geqslant 0$ the left-hand side of (69) is intrinsically positive violating the inequality. So if $\beta < 0$ anywhere, it must occur in $0 \geqslant \mu \geqslant -1$.

To demonstrate that this is impossible set $\mu = -x$. Then for $\beta < 0$ in $0 \leqslant x \leqslant 1$, we require

$$16x^4 + 41k^2 + 9 - 2x(15 + 16x^2) < 0. \quad (70)$$

But the left-hand side of inequality (70) can be written

$$16x^2(x-1)^2 + 25(x - \tfrac{3}{5})^2$$

which is intrinsically positive. Accordingly, β is positive, therefore $f > 0$ and the Type A modes are degenerative.

Type B Modes

In order to obtain an unstable mode from Eq. (62) we must check that inequality (65) is satisfied for some value of R, m and the helicity (other than zero) for the root of Eq. (63a) which reduces to $y = 0$ on $R = 0$. It would appear at first sight that in order to do this we require specific functional forms of $E(k)$ and $H(k)$ in order to compute s and γ [Eq. (63b)]. And then the root of Eq. (62) would be model dependent. This in turn would lead to controversy, for suppose no unstable mode were found, it could then be argued that this is an artifact of the particular forms of $E(k)$ and $H(k)$ chosen.

To circumvent such difficulties it is opportune here to use the asymptotic ($m \to 0$, $m \to \infty$) nature of the J_α functions in order to illustrate how the problem can be solved in general, *irrespective* of the functional form of the turbulent velocity field-defined through $E(k)$ and $H(k)$.

As $m \to 0$ we see from Appendix B that $\epsilon(m \to 0) \propto m^r \to 0$ for $r \neq 0$; $\epsilon(m \to \infty) \propto m \to \infty$ provided the helicity $H(k)$ is finite. Thus for any finite value of the helicity satisfying Cramér's theorem there exists a wave number m_* at which $\epsilon^2 = 1$.

Thus as m varies between zero and infinity, γ varies in $\infty \geqslant \gamma \geqslant \tfrac{2}{3}$, $-\infty \leqslant \gamma \leqslant 0$. Likewise, s defined by Eq. (63b), varies in $\infty \geqslant s \geqslant 0$ as m and the Reynolds number R are varied.

Altogether then it is both sufficient and necessary to inspect the roots of the cubic Eq. (63e) as functions of s and γ in the above ranges.

The root which reduces to

$$2y = 1 - (1 + 4is)^{1/2}$$

as $\gamma \to \frac{2}{3}$ is the only physically allowable Type B mode, for as $s \to 0, y$ must tend to zero.

Consider then the roots of the cubic (63e). If we are to obtain a regenerative mode, the physically acceptable root of Eq. (63e) (y_1, say) must have $\mathrm{Re}y_1 > 0$ for some range of γ and s. Now on $\gamma = \frac{2}{3}$, $\mathrm{Re}y_1 < 0$. Therefore, if $\mathrm{Re}y_1$ changes sign there must exist a set of γ, s values on which $\mathrm{Re}y_1 = 0$.

It is simple to show that one of the roots is purely imaginary ($Y = -is/6$ on $\gamma = -s^2/6$. The other two roots are never purely imaginary (except on $s = 0$ where one of them is finite and the other is zero). Further, in $\gamma > -s^2/6$ the root Y is degenerative ($\mathrm{Re}Y < 0$), while in $\gamma < -s^2/6$ it is regenerative [$\mathrm{Re}Y > 0$). The only question to settle is whether the Y root is, in fact, the same as the root y_1. If so we then have regenerative (i.e., a growing mode) kinematic dynamo action.

In order to determine whether the mode Y "matches" smoothly onto the mode y_1, consider the roots of the cubic (63e) in the limit $\gamma \to -\infty$, s finite. Then

$$y_a = 3\gamma + 0(s), \quad \mathrm{Re}y_a < 0, \tag{71a}$$

$$4y_b = 1 - is - (1 - s^2 + 6is)^{1/2} + 0(1/\gamma), \quad \mathrm{Re}y_b < 0, \tag{71b}$$

$$4y_c = 1 - is + (1 - s^2 + 6is)^{1/2} + 0(1/\gamma), \quad \mathrm{Re}y_c > 0. \tag{71c}$$

Now as $s \to 0$ the physically acceptable mode in $\gamma \to -\infty$ is $y_b \equiv y_1$. But $\mathrm{Re}y_b < 0$ in $\gamma < -s^2/6$, whereas $\mathrm{Re}Y > 0$ in $\gamma < -s^2/6$. Accordingly, the physically acceptable mode of Eq. (63e) always has $\mathrm{Re}y_1 < 0$ for all allowable values of γ and s. Thus the Type B mode is degenerative.

Altogether then for static velocity turbulence which gives rise to analytic $J_\alpha(k, \kappa)$ over the finite domains of the complex k and κ planes there are no growing modes. *All modes are degenerative. This result includes helical, as well as isotropic, mirror symmetric, incompressible velocity turbulence.*

As we shall demonstrate (Sec. VI) these *exact* results are in sharp contrast to the results obtained under the assumptions of first order smoothing theory (FOST), which is, therefore, one way to illustrate the incorrectness of the FOST approximation in problems of this nature.

VI. THE *DIA* EQUATIONS, FIRST ORDER SMOOTHING AND HAMMERSTEIN'S THEOREM

Let us consider once again Eqs. (36) and (37). Write

$$\Psi(k, \omega) = k\Phi(k, \omega)Y(k, \omega) \tag{72}$$

to obtain

$$\begin{aligned}\Phi(k, \omega)k &= [1 - i\omega k^{-2} + R^2 \int_0^\infty d\kappa\kappa^2 J_1(k, \kappa)/\Phi(\kappa, \omega)]k \\ &\quad + R^2 Y(k, \omega)^{-1} \int_0^\infty \kappa^2(\Phi(\kappa, \omega))^{-1} \\ &\quad \times (J_3(k, \kappa) + J_2(\kappa, \omega)/\kappa Y(\kappa, \omega))d\kappa\end{aligned} \tag{73}$$

and

$$\begin{aligned}&k[1 - i\omega k^{-2} + R^2 \int_0^\infty \kappa^2(J_1(k, \kappa)/\Phi(\kappa, \omega))d\kappa] \\ &= - R^2 Y(k, \omega) \int_0^\infty \kappa^2(\Phi(\kappa, \omega))^{-1} \\ &\quad \times [J_3(k, \kappa) + J_2(k, \omega)/\kappa Y(\kappa, \omega)]d\kappa.\end{aligned} \tag{74}$$

The dispersion relation for the normal modes of the ensemble average magnetic field is given through

$$\Phi(k, \omega) = 0 \quad \text{and} \quad Y = \pm 1. \tag{75}$$

The first order smoothing theory (FOST) result can be obtained from Eqs. (73) and (74) as follows. In the absence of helicity $Y = \infty$ and Φ is finite [see L1 or Eq. (16)]. So for finite helicity $Y \alpha \delta^{-1}$, where δ measures the "strength" of the helical component of the velocity turbulence. Then ignore the terms involving J_2/Y compared to J_3 in Eqs. (73) and (74) since J_2/Y is proportional to $\delta\epsilon$ while J_3 is proportional to ϵ. Here ϵ measures the "strength" of the mirror symmetric component of the velocity turbulence. [The reason for this approximation is that we will expand Φ later in powers of ϵ and this will require $\epsilon \ll 1$ in order to be valid at all.]

In the remaining terms in Eq. (73) set $Y = \pm 1$ and, on the right-hand side of Eq. (73), set $\Phi = 1 - i\omega k^{-2}$ while on the left-hand side set $\Phi = 0$. Upon so doing we obtain

$$\begin{aligned}1 - i\omega k^{-2} = - R^2 k \int_0^\infty \kappa^2(1 - i\omega\kappa^{-2})^{-1} \\ \times [J_1(k, \kappa) \pm J_3(k, \kappa)/k]d\kappa,\end{aligned} \tag{76}$$

which is *precisely* the first order smoothing result obtained elsewhere (Krause and Roberts, Lerche, Ref. 8).

But Eq. (76) is not correct. This can be shown in several ways. The simplest, for our purposes, is to suppose that we can set $Y = \pm 1$ in Eq. (73) to obtain an equation for Φ of the form

$$\Phi = a(k, \omega) + \int_0^\infty \frac{b(k, \kappa)}{\Phi(\kappa, \omega)} d\kappa, \tag{77}$$

where a and b are known.

Equation (77) is in Hammerstein's normal form (Hammerstein, 1930)[14] and we can apply Hammerstein's theorem to it. This theorem says that there exists a uniformly convergent approximation to the nonlinear equation provided that, for all Φ

$$\left| \int^\Phi f(\kappa, u)du \right| \leq \tfrac{1}{2}C_1 |\Phi|^2 + C_2 |\Phi|, \tag{78}$$

where C_1 and C_2 are positive constants.

In our case $f \propto 1/\Phi$ and since $|\Phi|[C_1|\Phi| + 2C_2]$ is *not* greater than $|\ln\Phi|$ for all Φ (we note that we are interested in the zeros of Φ), there does not exist a uniformly convergent approximation to Eq. (77). In other words, first order smoothing theory is inaccurate (see e.g. Lerche and Parker). Kraichnan has demonstrated this inaccuracy of first order smoothing theory under a much wider class of conditions than we are concerned with here. For our purposes the above argument is sufficient to illustrate the nonuniformity of convergence of the FOST results.

Further, if we consider the behavior of Eq. (76) it indicates the presence of *growing* (i.e., $\mathrm{Im}\omega > 0$) modes in the long wavelength $k \to 0$ limit (Krause, Lerche, Krause and Roberts). But the exact calculations (Sec. V) demonstrate that there are no growing modes—all modes decay. This illustrates directly that one cannot treat violations of Hammerstein's theorem lightly.

VII. DISCUSSION AND CONCLUSION

In this paper we have set up the general Kraichnan equations applicable to the turbulent kinematic dynamo problem under an homogeneous, stationary turbulent velocity field which is incompressible, isotropic, but not mirror symmetric.

We demonstrated that when the helicity of the velocity field is nonzero then the Green's stress tensor enjoys equipartition between its symmetric and antisymmetric parts at the normal modes of the average magnetic field. For zero helicity this is not true, for then there is no antisymmetric part to the Green's stress tensor. In some sense, then, the zero-helicity limit is a singular limit of the nonlinear Kraichnan integral equations. We do not yet completely understand the physics underlying this singular behavior. Mathematically, it arises because the degree (and number) of the nonlinear equations is reduced in the absence of helicity, and under such conditions some kind of singular behavior often occurs.

We solved the Kraichnan equations *exactly* when the velocity turbulence was homogeneous and static, and we obtained the dispersion relation describing the normal modes of the magnetic field. We demonstrated that *all* of the physically acceptable modes of the dispersion relation gave rise to degenerative (i.e., decaying modes) kinematic dynamo activity. There are *no* growing modes. These results are in sharp contrast to the first order smoothing theory approximation, which gives rise to growing modes in the presence of helical velocity turbulence (Krause and Roberts).

The resolution of this difference in results was shown to be the inapplicability of the first order smoothing theory approximation to the Kraichnan equations, for the Kraichnan equations do not satisfy Hammerstein's theorem while FOST requires satisfaction of Hammerstein's theorem in order to be valid at all. This point has been stressed in more general circumstances than we are concerned with here by Kraichnan.

In summary then, what we have shown is that the presence of helicity in the velocity turbulence is *not* sufficient to guarantee growing modes, contrary to the results of approximate treatments based on FOST—which is an invalid approximation (Kraichnan; Lerche and Parker).

If there is to be any growth of the average magnetic field it cannot come from *static* velocity turbulence. Therefore growing magnetic fields *may* occur only when the velocity turbulence is time dependent. Unfortunately, the Kraichnan equations are then sufficiently complex that, so far, we have been unable to solve them exactly. And an exact solution is necessary if we are to place any confidence in the solution. We are still working on this problem and any further progreess will be reported in due course.

ACKNOWLEDGMENTS

We are grateful to Professor E.N. Parker and Dr. B.-C. Low for several valuable discussions. The work reported here was supported by the United States Air Force, Air Force Systems Command under contract F-19628-72-C-0056.

APPENDIX A

Here we demonstrate two points. First that it is possible to choose the velocity turbulence spectra $E(k)$ and $H(k)$ (subject to Cramér's theorem) so that the $J_\alpha(k,\kappa)$ are analytic in the finite domains of the complex k and κ planes. Secondly that Eqs. (39) and (40) are valid when the $J_\alpha(k,\kappa)$ are analytic in the finite domain.

A. The velocity turbulence spectra and analyticity of $J_\alpha(K,\kappa)$

In order to illustrate the point choose

$$kH(k) = E(k)\nu k(k^2+1)^{-1}, \tag{A1}$$

with $2^{1/2}\nu < 3$, so that

$$-E(k) \leqslant kH(k) \leqslant E(k), \quad \text{all } k.$$

Further choose, for example,

$$E(k) = k^4(k^2+1)\exp(-k^2). \tag{A2}$$

Then it is a simple matter to demonstrate that *all* the $J_\alpha(k,\kappa)$ are analytic in the finite domains of the complex k and κ planes with essential singularities of the form $\exp[-(k\pm\kappa)^2]$ on the circles at infinity.

The choice (A1) and (A2) is by no means unique. It is chosen merely to demonstrate that it is possible to find a class of $H(k)$ and $E(k)$ which give rise to analytic $J_\alpha(k,\kappa)$.

B. The integrals (39) and (40)

In order to illustrate the general method of obtaining the results (39) and (40) consider the integral

$$I = \int_{-\infty}^{\infty} \kappa^2 \frac{J(k,\kappa)}{D(\kappa)}\, d\kappa, \tag{A3}$$

where $\pm(m+i\epsilon)$ and $\pm(M+i\epsilon)$ $(M>m)$ are zeros of $D(\kappa) \equiv 1 - i\omega\kappa^{-2} + \Lambda_1 J(k,m) + \Lambda_2 J(\kappa,M)$ occurring at $\kappa = \pm(m+i\epsilon)$, $\kappa = \pm(M+i\epsilon)$ with $\epsilon \to +0$. Further, Λ_1 and Λ_2 are functions only of m, M, and ω.

Upon performing a contour integration in complex κ space around a semicircle in the upper half complex κ plane we see that the integral I converges on the semicircle at infinity if $k<M$. [We have in mind that both $E(k)$ and $H(k)$ will vary as (powers of k) x $\exp(-k^2)$ at large k.]

And then

$$I = 2\pi i \left(m^2 \frac{J(k,m)}{D(\kappa=m)'} + M^2 \frac{J(k,M)}{D(\kappa=M)'} \right), \quad k<M, \tag{A4}$$

where

$$D(\kappa=\beta)' = \partial D(\kappa)/\partial\kappa\,|_{\kappa=\beta}. \tag{A5}$$

Now let $M\to\infty$, so that

$$I(M\to\infty) \propto J(k,m), \quad \text{all finite } k.$$

Then if we assume that both Φ and Ψ can be written in the form

$$\Phi \text{ or } \Psi = 1 - i\omega\kappa^{-2} + \sum_{i=1}^{3} [\Lambda_i J_i(k,m) + \beta_i J_i(k,M)], \tag{A6}$$

we obtain Eqs. (39) and (40) as $M\to\infty$. This demonstrates that by assuming Φ and Ψ have the form (A6) the integral Eqs. (34) and (35) show that Φ and Ψ *do* have the form (A6) provided a dispersion relation is satisfied for ω as a function of m.

APPENDIX B

Consider

$$J_1(k,m) = 2\pi \int_{-1}^{+1} (1-\mu^2) \frac{E(|\mathbf{k}+\mathbf{m}|)}{|\mathbf{k}+\mathbf{m}|^2}\, d\mu, \tag{B1}$$

where $|\mathbf{k}+\mathbf{m}|^2 = k^2+m^2+2km\mu$.

Then

$$J_1(m,m) = \pi m^{-2} \int_{-1}^{+1} d\mu (1-\mu) E(m 2^{1/2}(1+\mu)^{1/2}) \quad \text{(B2)}$$

and

$$J_1'(m,m) = -\pi m^{-3} \int_{-1}^{+1} d\mu (1-3\mu) E(m 2^{1/2}(1+\mu)^{1/2}). \quad \text{(B3)}$$

Likewise,

$$J_2(m,m) = \pi \int_{-1}^{+1} d\mu (3+2\mu)(1-\mu) E(m 2^{1/2}(1+\mu)^{1/2}) \quad \text{(B4)}$$

and

$$J_2'(m,m) = -\pi m^{-1} \int_{-1}^{+1} d\mu (1-6\mu)(1+\mu) \times E(m 2^{1/2}(1+\mu)^{1/2}) \quad \text{(B5)}$$

Also,

$$J_3(m,m) = 2\pi \int_{-1}^{+1} d\mu (1-\mu^2) H(m 2^{1/2}(1+\mu)^{1/2}) \quad \text{(B6)}$$

and

$$J_3'(m,m) = -2\pi m^{-1} \int_{-1}^{+1} d\mu (1+\mu)(1-3\mu) \times H(m 2^{1/2}(1+\mu)^{1/2}). \quad \text{(B7)}$$

For small $k \to 0$ let $E(k) \to E_0 k^{2n}$. Then, by Cramér's theorem

$$H(k) \to \alpha E_0 k^{2n-1+r},$$

with $r \geqslant 0$; and, if $r = 0$, $|\alpha| \leqslant 1$, while if $r > 0$ then α is real but arbitrary as $k \to 0$.

Then as $m \to 0$ we have

$$J_1(m,m) = \pi E_0 2^n m^{2n-2} \int_{-1}^{+1} (1-\mu)(1+\mu)^n d\mu \equiv j_1 m^{2n-2}, \quad \text{(B8a)}$$

$$J_1'(m,m) = -\pi E_0 2^n m^{2n-3} \int_{-1}^{+1} (1-3\mu)(1+\mu)^n d\mu \equiv j_1' m^{2n-3}, \quad \text{(B8b)}$$

$$J_2(m,m) = \pi E_0 2^n m^{2n} \int_{-1}^{+1} (3+2\mu)(1-\mu)(1+\mu)^n d\mu \equiv j_2 m^{2n}, \quad \text{(B8c)}$$

$$J_2'(m,m) = -\pi E_0 2^n m^{2n-1} \int_{-1}^{+1} (1+6\mu)(1+\mu)^{n+1} d\mu \equiv j_2' m^{2n-1}, \quad \text{(B8d)}$$

$$J_3(m,m) = 2\pi\alpha E_0 2^{n-1/2+r/2} m^{2n-1+r} \int_{-1}^{+1} (1-\mu)(1+\mu)^{n+1/2+r/2} d\mu \equiv j_3 m^{2n-1+r}, \quad \text{(B8e)}$$

$$J_3'(m,m) = -2\pi\alpha E_0 2^{n-1/2+r/2} \times m^{2n-2+r} \int_{-1}^{+1} (1-3\mu)(1+\mu)^{n+1/2+r/2} d\mu \equiv j_3' m^{2n-2+r}. \quad \text{(B8f)}$$

At large $m (\gg 1)$ since $E(k)$ has a scale of unity we proceed as follows. Write $(1+\mu)2m^2 = x$ in Eqs. (B2)-(B7) so that, for example,

$$J_1(m,m) = \tfrac{1}{2}\pi m^{-4} \int_0^{4m^2} E(x^{1/2})(2 - x/2m^2) dx. \quad \text{(B9)}$$

Since E declines over a scale of unity then as $m \to \infty$, we have

$$J_1(m,m) \to \pi m^{-4} \int_0^\infty E(x^{1/2}) dx \equiv m^{-4}\mathcal{E}_0. \quad \text{(B10a)}$$

In like manner, we obtain as $m \to \infty$ that

$$J_1'(m,m) \to -(2/m) J_1(m,m), \quad \text{(B10b)}$$

$$J_2(m,m) \to J_1(m,m), \quad \text{(B10c)}$$

$$J_2'(m,m) \to -\frac{7\pi}{4} m^{-5} \int_0^\infty x E(x^{1/2}) dx \equiv -\mathcal{E}_1 m^{-5}, \quad \text{(B10d)}$$

$$J_3(m,m) \to \pi m^{-4} \int_0^\infty x H(x^{1/2}) dx \equiv H_1 m^{-4}, \quad \text{(B10e)}$$

$$J_3'(m,m) \to -2m^{-1} J_3(m,m). \quad \text{(B10f)}$$

[1] I. Lerche, J. Math. Phys. 14, 1381 (1973). Referred to in text as L1.
[2] R. H. Kraichnan, J. Math. Phys. 2, 124 (1961).
[3] U. Frisch, "Wave Propagation in Random Media" In *Probabilistic Methods in Applied Mathematics* edited by A. T. Bharucha-Reid (Academic, New York, 1968), Vol. I.
[4] F. Krause and P. H. Roberts, Astrophys. J. 181, 977 (1973).
[5] I. Lerche, Astrophys. J. 166, 627 (1971).
[6] F. Krause, Astron. Nachr. **294**, 83 (1972).
[7] I. Lerche and E. N. Parker, submitted to J. Math. Phys.
[8] I. Lerche, Astrophys. J. 181, 993 (1971).
[9] A. Hammerstein, Acta Math. **54**, 117 (1930).
[10] G. K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge, U. P., Cambridge, 1953).
[11] H. Cramer, Ann. Math. **41**, 215 (1940).
[12] See footnote 5 of Ref. 1 (LI).
[13] The point here is that if we investigate the Kraichnan equations with $H(k) \equiv 0$ (LI), then, at dispersion, G_{iu} is proportional to $k^2\delta_{iu} - k_i k_u$. There is no helical term in G_{iu} when $H \equiv 0$. For $H \neq 0$, G_{iu} is proportional to $k^2\delta_{iu} - k_i k_u \pm i\epsilon_{iu\lambda} k_\lambda k$ at dispersion.
[14] An integral equation is said to be in Hammerstein's normal form if it can be written $\Phi = a(k) + \int b(k,\kappa) f(\mathrm{K}, \Phi(\kappa)) d\mathrm{K}$.

Application of Kraichnan's direct interaction approximation to kinematic dynamo theory. III. Solution of the Kraichnan equations under Parker's "short-sudden" conditions

I. Lerche

Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637
(Received 4 May 1973)

Using Parker's "short-sudden" conditions, we solve the Kraichnan equations *exactly*. We find that the normal modes of the ensemble average magnetic field have very different properties in this case than when the velocity turbulence is static, which situation was investigated in earlier papers in this series. We have done this calculation for two reasons: first because exact solutions of the Kraichnan equations are few in number, and second because the nonlinearity and singularity of the Kraichnan equations is such as to emphasize the physical difference in the properties of the normal modes of the average magnetic field under a small change in the prescription of the turbulent velocity field.

I. INTRODUCTION

Since the invention of Kraichnan's[1] direct interaction approximation for investigating turbulence problems, considerable effort has gone into trying to solve the resulting nonlinear Kraichnan equations. This effort is warranted because the Kraichnan equations describe exactly an ensemble of physically possible dynamical systems. As such their exact solutions are of interest, for they can then be used as templates against which one can compare and contrast approximate treatments of the same problems. This in turn outlines the regime of applicability (if any) of the approximate methods. Further, Frisch[2] has noted that a model turbulence problem described by the Kraichnan equations is an accurate approximate description of the true turbulence problem for all values of the parameters involved. As such the model problem is both physically realizable and acceptable.

In the previous papers in this series [Lerche, Ref. 3 a, b—hereinafter referred to as L1 and L2] we have set up the Kraichnan equations describing kinematic dynamo activity in an infinite medium under a turbulent velocity field which is statistically homogeneous and stationary. We showed that when the velocity turbulence was *static* the Kraichnan equations were soluble and that (i) for incompressible isotropic, mirror symmetric velocity turbulence all the normal modes of the ensemble average magnetic field were degenerative i.e., decaying); (ii) including a helical component in the velocity turbulence, and maintaining incompressibility, gave rise to solely degenerative kinematic dynamo activity. In other words, the presence of helicity is not sufficient to guarantee regenerative dynamo action.

When the velocity turbulence is not static, the degree of nonlinearity, and the complexity, of the Kraichnan equations is such that, in general, we have not been able to obtain their general solution. We should also point out that the number of exact solutions of the Kraichnan equations is small.

There is, however, one time dependent form of velocity turbulence for which we have been able to construct exact solutions of the Kraichnan equations. We give here the results of those calculations for they give rise to a very different behavior than was obtained (L1, L2) under static velocity turbulence.

II. BASIC EQUATIONS

In the previous papers in this series (L1, L2) we obtained the nonlinear singular integral equations describing kinematic activity under incompressible, homogeneous and stationary velocity turbulence. We also remarked that we were unable to solve them in general, and, in fact, we could solve them then only when the velocity turbulence was static. Since that time we have found a method of solving the equations under Parker's[4] "short-sudden" conditions—which occurs when both the time-scale and spatial scale of the turbulent velocity field are infinitesimal compared to any other time and space scales under consideration.

In view of the interest in the problem of large scale magnetic field generation by turbulent velocity fluctuations, we give here the method of solution and the results; first, because they represent statistically exact solutions to the Kraichnan equations (and very few such solutions are known); secondly, because the method, or some variation of it, may be of more general use than for Parker's short-sudden limit—although we have not yet been able to generalize it; thirdly, because the results obtained from the Kraichnan equations describe exactly an ensemble of dynamically possible systems; and fourthly, because the results are *precisely* those obtained by Parker[4] using a very different approach.

The notation is the same as that in L1 and L2. Further, the two situations to be described are, *prime facie*, so different mathematically that we consider each separately, and compare and contrast the results later (Sec. III).

A. Incompressible, isotropic, mirror symmetric turbulence

From L1 Eq. (58) we have

$$\Phi(k,\omega) = 1 - i\omega k^{-2} + R^2 \int_0^\infty \kappa^2 J(k,\kappa,\omega-\omega')d\omega'/\Phi(\kappa,\omega'), \tag{1}$$

where the normal modes of the ensemble average magnetic field are given through

$$\Phi(k,\omega) = 0. \tag{2}$$

Further,

$$J(k,\kappa,\omega-\omega') = 2\pi \int_{-1}^{+1} d\mu E(|\boldsymbol{k}+\boldsymbol{\kappa}|, \times \omega-\omega')(1-\mu^2)|\boldsymbol{k}+\boldsymbol{\kappa}|^{-2},$$

with

$$|\boldsymbol{k}+\boldsymbol{\kappa}|^2 \equiv (k^2+\kappa^2+2k\kappa\mu).$$

Copyright © 1973 by the American Institute of Physics

Also

$$\langle v_i(\boldsymbol{x},t)v_j(\boldsymbol{x}',t')\rangle = R_{ij}(\boldsymbol{x}-\boldsymbol{x}', t-t'),$$

$$R_{ij}(\boldsymbol{k},\omega) \equiv \int d^3\boldsymbol{x}dtR_{ij}(\boldsymbol{x},t)\exp[i(\boldsymbol{kx}-\omega t)] = E(k,\omega)(\delta_{ij}-k_ik_jk^{-2}),$$

where $E(k,\omega) \geqslant 0$ by Cramér's[5] theorem.

Also R is the magnetic Reynolds' number.

When the velocity turbulence occurs over a time-scale which is infinitesimal compared to any other time-scale of interest we write $E(k,\omega)=E(k)$ to obtain from equation (1):

$$\Phi(k,\omega) = 1 - i\omega k^{-2} + R^2\int_0^\infty \kappa^2 J(k,\kappa)d\kappa \int_{-\infty}^{\infty} d\omega'/\Phi(\kappa,\omega'). \tag{3}$$

The integral over ω' in Eq. (3) cannot be completed by closure in an *arbitrary* domain of the complex ω' plane. This arises because for $t<0$, $G(t)=0$, where $G(t)$ is the Green's function [see L1, Eq. (3)]. The ω' integration path must pass above all poles and branch cuts (if any) of the integrand with closure in the lower half complex ω'-plane for $t>0$.

To effect a solution to Eq. (3) is now relatively simple. We start by assuming a form for $\Phi(k,\omega)$ and then prove that the assumed form does indeed satisfy Eq. (3). In other words we make *a priori* assumptions and justify them *a posteriori*. To this end assume *a priori* that $\Phi(k,\omega)$ has only a single, simple zero in the complex ω plane at $\omega=\Omega(k)$ and that $\Phi(k,\omega)$ can be written in the form

$$\Phi(k,\omega) = a(k)[\omega - \Omega(k)]. \tag{4}$$

Then

$$\int_{-\infty}^{\infty} d\omega'/\Phi(\kappa,\omega') = -i\pi/a(k). \tag{5}$$

When Eqs. (4) and (5) are used in Eq. (3) we obtain

$$a(k)[\omega-\Omega(k)] = 1 - i\omega k^{-2} - i\pi R^2\int_0^\infty \kappa^2 J(k,\kappa)d\kappa/a(\kappa). \tag{6}$$

But if the assumed form (4) is indeed a solution of Eq. (3), then Eq. (6) must be true for all values of ω. This demands that

$$a(k) = -i/k^2, \tag{7}$$

and that

$$i\Omega(k) = k^2\,(1 + \pi R^2\int_0^\infty \kappa^4 J(k,\kappa)d\kappa). \tag{8}$$

It is evident by definition that $J(k,\kappa) \geqslant 0$, for all real k and κ. Accordingly, $i\Omega(k)>0$. Now the normal modes of the large-scale magnetic field are given through $\Phi(k,\omega)=0$ [L1, Eq. (2)]. This yields the dispersion relation

$$\omega = \Omega(k). \tag{9}$$

The normal mode time dependence was chosen (L1) to be of the form $\exp(-i\omega t) \equiv \exp[-i\Omega(k)t]$. But $i\Omega(k)>0$. So all normal modes of the large scale magnetic field decay under incompressible, isotropic, mirror symmetric, velocity turbulence when Parker's "short-sudden" conditions are in force.

B. Incompressible, helical velocity turbulence

When the velocity turbulence contains a helical component as well as a mirror symmetric component, we have [L2, Eqs (34) and (35)].

$$1 - i\omega k^{-2} + R^2\int_0^\infty d\kappa\kappa^2 J_1(k,\kappa)\int_{-\infty}^{\infty} d\omega'/\Phi(\kappa,\omega')$$
$$= \Phi(k,\omega) - R^2\Phi(k,\omega)[F(k,\omega]^{-1}\int_0^\infty \kappa^2 d\kappa\int_{-\infty}^{\infty} d\omega'$$
$$\times [J_2(k,\kappa)/F(\kappa,\omega') + J_3(k,\kappa)/\Phi(\kappa,\omega')] \tag{10}$$

and

$$\Phi(k,\omega)(1 - i\omega k^{-2} + R^2\int_0^\infty \kappa^2 d\kappa J_1(k,\kappa)\int_{-\infty}^{\infty} d\omega'/\Phi(\kappa,\omega'))$$
$$= -R^2k^{-2}F(k,\omega)\int_0^\infty d\kappa\kappa^2\int_0^\infty d\omega'[J_2(k,\kappa)/F(\kappa,\omega')$$
$$+ J_3(k,\kappa)/\Phi(\kappa,\omega')]. \tag{11}$$

In this case

$$R_{ij}(k,\omega) = E(k,\omega)(\delta_{ij} - k_ik_jk^{-2}) + i\epsilon_{ij\lambda}k_\lambda H(k,\omega), \tag{12}$$

and, by Cramér's theorem,

$$E(k,\omega) \geqslant 0; \quad -E(k,\omega) \leqslant kH(k,\omega) \leqslant E(k,\omega) \tag{13}$$

Further,

$$J_1(k,\kappa,\omega) = 2\pi\int_{-1}^{+1} d\mu(1-\mu^2)|\boldsymbol{k}+\boldsymbol{\kappa}|^{-2}E(|\boldsymbol{k}+\boldsymbol{\kappa}|,\omega), \tag{14a}$$

$$J_2(k,\kappa,\omega) = 2\pi\int_{-1}^{+1} d\mu(1-\mu^2)|\boldsymbol{k}+\boldsymbol{\kappa}|^{-2}(k^2 + |\boldsymbol{k}+\boldsymbol{\kappa}|^2)E(|\boldsymbol{k}+\boldsymbol{\kappa}|,\omega), \tag{14b}$$

$$J_3(k,\kappa,\omega) = 2\pi\int_{-1}^{+1} d\mu(1-\mu^2)H(|\boldsymbol{k}+\boldsymbol{\kappa}|,\omega), \tag{14c}$$

with

$$|\boldsymbol{k}+\boldsymbol{\kappa}|^2 = (k^2+\kappa^2+2k\kappa\mu).$$

In this case the normal modes of the large-scale magnetic field are given through

$$\Phi(k,\omega) = 0 = F(k,\omega), \quad \Phi(k,\omega)k/F(k,\omega) = \pm 1. \tag{15}$$

We consider Eqs. (10) and (11) under Parker's "short-sudden" conditions, i.e., when the replacements $E(k,\omega) = E(k)$, $H(k,\omega) = H(k)$ are valid. Accordingly, we have written the $J_\alpha(k,\kappa,\omega)$ in Eqs. (10) and (11) ignoring the frequency dependence.

In order to solve Eqs. (10) and (11) under Parker's short-sudden conditions, we have to be more circumspect than in the case of incompressible, isotropic, mirror symmetric, velocity turbulence (case A). Note, once again, that the ω' integrals are to be completed by closure in the lower half complex ω' plane.

Assume *a priori* that

$$\int_{-\infty}^{\infty} d\omega'/F(\kappa,\omega') = 0. \tag{16}$$

Also we assume *a priori* that

$$\int_{-\infty}^{\infty} d\omega'/\Phi(\kappa,\omega') = -i\pi/a(\kappa), \tag{17}$$

where $a(\kappa)$ is, as yet, undetermined. We shall solve Eqs. (10) and (11) under *a priori* assumptions (16) and (17) and check to ensure that the solution does indeed satisfy the assumptions. Using Eqs (16) and (17), Eqs. (10) and (11) yield

$$1 - i\omega k^{-2} - i\pi R^2 \int_0^\infty \kappa^2 J_1(k,\kappa)d\kappa/a(\kappa)$$
$$= \Phi(k,\omega) + i\pi R^2 \Phi(k,\omega)[F(k,\omega)]^{-1} \times \int_{-\infty}^{\infty} \kappa^2 J_3(k,\kappa)d\kappa/a(\kappa) \tag{18}$$

and

$$\Phi(k,\omega)[1 - i\omega k^{-2} - i\pi R^2 \int_0^\infty \kappa^2 J_1(k,\kappa)d\kappa/a(\kappa)] = i\pi R^2 k^{-2} F(k,\omega) \int_0^\infty \kappa^2 J_3(k,\kappa)d\kappa/a(\kappa). \tag{19}$$

Use Eq. (19) to eliminate $F(k,\omega)$ from Eq. (18). Upon so doing we obtain

$$\Phi(k,\omega) = U^{-1}[U^2 - k^{-2}(i\pi R^2 \int_0^\infty \kappa^2 J_3(k,\kappa)d\kappa/a(\kappa))^2], \tag{20}$$

where

$$U = 1 - i\omega k^{-2} - i\pi R^2 \int_0^\infty \kappa^2 J_1(k,\kappa)d\kappa/a(\kappa). \tag{21}$$

Then it follows from Eq. (20) that

$$\int_0^\infty d\omega'/\Phi(\kappa,\omega') = \pi\kappa^{-2}, \tag{22}$$

so that

$$a(\kappa) = -i\kappa^{-2}. \tag{23}$$

Then

$$\Phi(k,\omega) = U^{-1}[U^2 - (\pi R^2 k^{-1} \int_0^\infty d\kappa\kappa^4 J_3(k,\kappa))^2], \tag{24}$$

where

$$U = 1 - i\omega k^{-2} + \pi R^2 \int_0^\infty \kappa^4 J_1(k,\kappa)d\kappa. \tag{25}$$

Equation (19) then gives

$$F(k,\omega) = -\Phi(k,\omega)Uk^2(\pi R^2 \int_0^\infty \kappa^4 J_3(k,\kappa)d\kappa)^{-1} \tag{26}$$

Now the dispersion relation for the normal modes of the large scale magnetic field is given through

$$\Phi = 0 = F, \quad \Phi k/F = \pm 1.$$

From Eqs. (24), (25), and (26) we see that this occurs when

$$U = \pm \pi k^{-1} R^2 \int_0^\infty \kappa^4 J_3(k,\kappa)d\kappa. \tag{27}$$

When (27) is satisfied, then $\Phi = 0 = F$ and $\Phi k/F = \pm 1$ occur simultaneously. The remaining question, then, is whether the *a priori* assumption (16) is satisfied. By using Eqs. (24) and (26) we see by inspection that it is indeed verified *a posteriori*.

Accordingly, Eqs. (24), (25), and (26) represent a solution to the Kraichnan equations under Parker's short-sudden conditions.

The dispersion relation (27) for the normal mode frequencies of the ensemble average magnetic field gives

$$i\omega = k^2(1 + \pi R^2 \int_0^\infty \kappa^4[J_1(k,\kappa) \pm k^{-1}J_3(k,\kappa)]d\kappa), \tag{28}$$

which is *precisely* Parker's result when the correlation length $L \to 0$. One of the modes given by Eq. (28) is unstable, i.e., regenerative dynamo action occurs at some wave number provided only that $\int_0^\infty \kappa^2 H(\kappa)d\kappa \neq 0$, as was first shown by Parker. He used a series expansion of the magnetic induction equation and then demonstrated convergence of the series in the infinitesimal scale length, infinitesimal time limit (for the turbulent velocity field). Hence the name "short-sudden" conditions.

III. COMPARISON OF THE RESULTS

There are two questions to be discussed here. First there is the question of the range of validity of the results (8) and (28), second there is the question of their physical interpretation.

A. The "short-sudden" limit

In the form used in this paper (viz. ignoring the frequency dependence of the turbulent velocity correlation tensor) it appears that the "short-sudden" conditions make use only of the "sudden" part, i.e., we have used an infinitesimal time-scale for the turbulent velocity fluctuations but we do not appear to have assumed an infinitesimal spatial scale for their occurrence.

This is, however, not completely correct. From Cramér's theorem we must have

$$R^2 \int |R_{ij}(k,\omega)| d^3k d\omega < \infty, \tag{29}$$

for physically acceptable turbulence spectra. Now when we choose an infinitesimal time-scale for the velocity turbulence we are taking $R_{ij}(k,\omega)$ to be independent of ω.

Then write $R_{ij}(k,\omega) = R_{ij}(k)$ to obtain $R^2 \int |R_{ij}(k,\omega)| d3kd\omega = R^2\beta\tau^{-1}$ with $\tau^{-1} = \int d\omega$, $\beta = \int |R_{ij}(k)| d3k < \infty$.

And β is finite, τ is infinitesimal.

The combination entering the dimensionless normal mode equations (8) and (28) (or for that matter entering the Kraichnan equations) is just R^2k^2. Then if we wish to have a finite correction to the free-decay modes in dimensional form, we require (write $k \to k/L$ where L is the correlation length of the velocity turbulence)

$$R^2/L^2 \equiv v^2/\eta^2 = \text{finite}.$$

But in order to satisfy Cramér's theorem we require

$$L^2 v^2/(\eta^2\tau) = \text{finite},$$

i.e., $L^2 \alpha \tau$.

But τ is infinitesimal. Accordingly so is L. And then we have to have infinitesimal turbulence occurring at an infinite rate ($L/\tau \propto \tau^{-1/2} \to \infty$) which is *precisely* Parker's short-sudden condition.

B. Interpretation of Equations (8) and (28)

It is correct to note that if the helicity is set to zero in Eq. (28), the helical modes dispersion relation reduces to Eq. (8)—the isotropic dispersion relation. It is *not* correct to argue that this limiting procedure is uniform. This arises because we have shown elsewhere (L2) that when any finite amount of helicity is present in the velocity turbulence, then the Green's magnetic stress tensor, which occurs in the Kraichnan equations, enjoys equipartition of its symmetric and antisymmetric parts in Fourier space at the normal mode values of the frequency. When the helicity is *zero* the antisymmetric part of the Green's tensor is zero under all conditions and so no equipartition statement is available. In other words, as far as the Green's magnetic stress tensor is concerned, passage to the limit of zero helicity is a singular limit. As far as the dispersion relations (8) and (28) are concerned passage to the limit of zero helicity is a well-behaved process. Accordingly, some care must be exercized when comparing and contrasting the implications and predictions of the normal mode dispersion relations with, and without, helicity in the turbulent velocity field.

We note here that in the absence of helicity there appears to be just one normal mode of the ensemble average magnetic field [Eq. (8)] which, under Parker's short-sudden conditions, is always decaying. When the helicity in the turbulent velocity field is nonzero the normal modes are two in number. The implication of this appears to be that the presence of helicity causes a bifurcation of a single mode (with no helicity) into two modes (an analogy which is suggestive is the lifting of degeneracy in atomic energy levels by L-S coupling or hyperfine splitting).

Further, note [Eq. (28)] that one of the bifurcated modes is unstable at some wave number provided only that $\int_0^\infty \kappa^2 H(\kappa) d\kappa \neq 0$. And this in turn is a somewhat singular behavior; for in the absence of helicity the mode is stable [Eq. (8)]; in the presence of helicity, *no matter how small*, the mode is unstable.

We have pointed out elsewhere (Lerche and Parker, Ref. 6) that the results obtained using first order smoothing theory are, in general, in error.

The work of Moffatt[7] is of interest in this regard. His early results[7a] use first order smoothing theory–but fortunately he eventually takes the "short-sudden" limit. Under that limit his results are identical to those obtained by Parker and agree with those obtained here. In his later work Moffatt[7b,c] showed that inertial waves in a rotating fluid possess helicity. Essentially his turbulent velocity field is a function of $x - x' - v_0(t - t')$ where v_0 is the inertial wave speed. Under the short-sudden limit he obtained kinematic dynamo action. This situation is not covered in the present paper, for it would entail writing $R_{ij}(\boldsymbol{k}, \omega) \propto \delta(\omega - \boldsymbol{k}.\boldsymbol{v}_0)$.

Needless to say we would be interested in seeing if Moffatt's (Ref. 7c) results follow from the Kraichnan equations under this specification for R_{ij}. So far we have been unable to solve the nonlinear Kraichnan equations under this form of turbulence.

IV. CONCLUSION

In this paper we have given *exact* solutions to the Kraichnan equations which describe kinematic dynamo activity in a physically realizable ensemble. The solutions, and their attendant normal modes, are valid under Parker's short-sudden conditions. And the normal modes in the presence of helical velocity turbulence are *precisely* these obtained by Parker who used an expansion procedure for the "true" turbulence problem. In other words, Kraichnan's model turbulence problem is indeed an accurate approximation to the true turbulence problem, by direct computation. And this bears out Frisch's remark (see the Introduction).

What we now require is a generalization of the method given here (or some other method) for solving the Kraichnan equations (1), (10), (11) exactly under time dependent velocity turbulence. Exact solutions are required for, as we have pointed out elsewhere (L1, L2), the Kraichnan equations do not satisfy Hammerstein's (1930) theorem; accordingly, they do not possess a uniformly convergent expansion in any variable.

Until such time as a general method of solution is available we point out that (i) the exact nature of the singular behavior as the helicity component of the turbulent velocity correlation tends to zero is not clearly delineated, in that we have only mapped out its behavior under the "short-sudden" conditions; (ii) the general class (both spatially and temporally) of incompressible, homogeneous, stationary turbulent velocity fields which will give rise to *regenerative* kinematic dynamo activity is not yet known [static turbulence (L1, L2) always give decay even when a helical contribution is included, short-sudden turbulence gives decay unless a helical contribution is included]. Somewhere between these extremes there must exist a general criterion which is both sufficient, and necessary, to guarantee either decay or growth of the normal modes of the ensemble average magnetic field for a given turbulent velocity field. And by this we mean that given $E(k, \omega)$ and $H(k, \omega)$ there must exist a relation which says that if certain integrals (as yet unknown) over k and ω are of a particular form, or exceed certain (unknown) values, then regenerative kinematic dynamo action is possible.

We would be interested in seeing any calculations relating to these points.

ACKNOWLEDGMENTS

The work reported here was supported by the United States Air Force, Air Force Systems Command under contract F-19628-72-C-0056.

[1] R. H. Kraichnan, J. Math. Phys. 2, 124 (1961).
[2] U. Frisch, "Wave Propagation in Random Media" in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968), Vol. I.
[3] I. Lerche, (a) J. Math. Phys. 14, 1381 (1973), hereafter called Paper L1; (b) J. Math. Phys. 14, 1579 (1973), hereafter called Paper L2.
[4] E. N. Parker, Astrophys. J. 122, 293 (1955).
[5] H. Cramer, Ann. Math. 41, 215 (1940).
[6] I. Lerche and E. N. Parker, J. Math. Phys. (to be published).
[7] H. K. Moffatt, (a) J. Fluid. Mech. 41, 435 (1970); (b) J. Fluid. Mech. 44, 705 (1970); (c) J. Fluid. Mech. 53, 385 (1972).

Uniformly valid asymptotic solution to a Volterra equation on an infinite interval

F. B. Hanson*

Department of Mathematics, University of Illinois at Chicago Circle, Chicago, Illinois

A. Klimas†

Aeronautical Research Associates of Princeton, Inc., 50 Washington Road, Princeton, New Jersey 08540

G. V. Ramanathan*

Department of Mathematics, University of Illinois at Chicago Circle, Chicago, Illinois

G. Sandri†

Aeronautical Research Associates of Princeton, Inc., 50 Washington Road, Princeton, New Jersey 08540
(Received 14 December 1972)

A Volterra equation on the infinite interval with a logarithmic kernel multiplied by a small parameter is analyzed. A Laplace transform solution is found; also, an approximate solution is given and proved to be uniformly asymptotic as the small parameter tends to zero for all time. The solution has an algebraic-logarithmic decay for large times. The Volterra equation is a model for the transport of charged particles in a random magnetic field.

1. INTRODUCTION

A model for cosmic ray transport has been constructed[1,2,3] which preserves the long-range character of the interaction given by the quasilinear approximation between charged particles and a random magnetic field. The kernel decays as the reciprocal of time t^{-1} and the standard adiabatic approximation is not valid.

The model governing the cosmic ray flux $F = F(t;\alpha)$ is the integrodifferential equation

$$\dot{F} = -\alpha K^*F - \sigma, \quad t > 0, \quad F(0;\alpha) = 1, \quad \alpha > 0, \quad (1.1)$$

where

$$\dot{F} = \frac{\partial F}{\partial t}(t;\alpha), \quad K^*F = \int_0^t d\tau K(t-\tau)F(\tau;\alpha)$$

and

$$K(t) = 1/(1+t).$$

The source $\sigma = \sigma(t)$ represents the density gradient and K in not integrable on $(0,\infty)$. The parameter α is small.

Our goal is to obtain an approximation to the solution of Eq. (1.1) that would be valid for all t, including when $t \to \infty$. One way to solve this problem is through Laplace transforms. The equation can be readily rewritten in Laplace transform variables and a solution can be obtained in terms of a contour integral. The explicit computation of this integral is a nontrivial task. In this paper, we calculate the long time behavior and prove that it is asymptotic as $\alpha \to 0$. We also give rigorous justification for the use of Laplace transforms.

Alternatively, we find an approximation for the kernel of the integral equation and the resultant equation is simply and exactly solvable. It turns out, perhaps fortuitously, that this solution is a uniformly valid asymptotic approximation to the exact solution; that is, the approximation differs from the exact solution by an amount which tends to zero with α, uniformly in the interval, $0 \le t < \infty$. Also, for long times, the relative error tends to zero with α. We prove this fact. We also verify that the long time behavior of this method coincides with the long time behavior of the Laplace transform solution.

2. HOMOGENEOUS EQUATION

Consider the homogeneous integrodifferential equation, with $\sigma = 0$,

$$\dot{f} = -\alpha K^*f, \quad t > 0, \quad \alpha > 0 \quad (2.1)$$

with initial condition

$$f(0;\alpha) = 1, \quad \alpha > 0.$$

If the solution $f = f(t;\alpha)$ is known, the solution of the inhomogeneous equation can be found by quadrature,[4]

$$F = f - \sigma^*f. \quad (2.2)$$

Hence, first we investigate the homogeneous integrodifferential equation. In particular, we are interested in the limit as $\alpha \to 0^+$. Upon integrating, Eq. (2.1) is written

$$f = 1 - \alpha\bar{K}^*f, \quad (2.3)$$

where $\bar{K}(t) = \int_0^t d\tau K(\tau)$. This is a Volterra equation of the second kind with kernel $-\alpha \ln(1+t)$.

3. LAPLACE TRANSFORM SOLUTION

We are interested in obtaining a uniformly valid asymptotic approximation to the solution of Eq. (2.1) that would accurately describe the decay of cosmic ray flux for all time. The usual Neumann series expansion in powers of α is a poor approximation. This is because for large t the terms of the series are majorized by a power of $\alpha t \ln t$ which becomes unbounded with t.

Since Eq. (2.1) can be written as a convolution-type, Volterra equation or renewal equation, as in Eq. (2.2), the Laplace transform is appropriate for finding a solution. This solution is found below as a contour integral and is put into a more usable form by means of an equivalent contour. While this Laplace transform solution is still rather complicated, it provides an exact solution which can be compared with the simpler uniformly valid approximation found in Sec. 4.

A. Justification for use of Laplace transforms

The following result justifies the use of the Laplace transform:

Lemma 3.1: Let $0 < \alpha < \alpha_0$. If $f(t;\alpha)$ is a solution of Eq. (2.3), then $f(t,\alpha)$ is of exponential order for $t \ge 0$ and is unique.

Copyright © 1973 by the American Institute of Physics

Proof: The first part is a direct consequence of Lemma 7.1 of Bellman and Cooke[3] and the uniqueness part relies on the same estimates as the first part. This is because for some $a > 0$,

$$\int_0^\infty dte^{-at}\,|-\alpha\bar{K}(t)| = \alpha a^{-1}e^{a}E_1(a),$$

where

$$E_1(a) = \int_a^\infty dte^{-t}t^{-1} \tag{3.1}$$

is the exponential integral and

$$\alpha a^{-1}e^{a}E_1(a) = \alpha a^{-2}(1 - e^{a}\int_a^\infty dte^{-t}t^{2}) \le \alpha a^{-2}$$

is less than unity when $a > \sqrt{\alpha_0}$. This establishes the exponential order of the solution. Uniqueness follows from the estimate

$$\sup_{0\le\tau\le t} |u(\tau)| \le \alpha a^{-1}e^{a}E_1(a) \sup_{0\le\tau\le t} |u(\tau)|,$$

so that $u(t) \equiv e^{-at}(f_1(t;\alpha) - f_2(t;\alpha)) = 0$, where f_1 and f_2 are two solutions of Eq. (2.3).

B. Formal Laplace transform solution

Transforming both sides of Eq. (2.3), one obtains,

$$\tilde{f}(s;\alpha) = s^{-1} - \alpha s^{-1}e^{s}E_1(s)\tilde{f}(s;\alpha),$$

where the Laplace transform is defined as

$$\tilde{f}(s;\alpha) = \int_0^\infty dte^{-st}f(t;\alpha).$$

It follows that

$$\tilde{f}(s;\alpha) = (s + \alpha e^{s}E_1(s))^{-1}, \tag{3.2}$$

provided

$$D(s;\alpha) \equiv (s + \alpha e^{s}E_1(s)) \ne 0. \tag{3.3}$$

The formal inverse is given by the contour integral

$$f_I(t;\alpha) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} dse^{ts}[D(s;\alpha)]^{-1}, \tag{3.4}$$

where $a > 0$ such that $[D(s;\alpha)]^{-1}$ has no singularities for $\mathrm{Re}[s] \ge a$. In Appendix A, we prove that $D(s;\alpha)$ has only two zeros and in Appendix B we prove that the inverse $f_I(t;\alpha)$ is the solution of the equation. In Appendix C, we find an equivalent contour where the solution is simpler but not simple to evaluate.

C. Boundedness of solution

The alternate form of the solution in terms of residues and a branch integral permits proof of the following assertion:

Theorem 3.1: f is bounded uniformly for $0 \le t < \infty$ and for sufficiently small positive α; in fact, $\dot{f}(t;\alpha)$ is absolutely integrable.

Proof: First, we show that $\dot{f}(t;\alpha)$ is absolutely integrable. Since the kernel, Eq. (C3) of Appendix C, of the branch cut integral behaves as

$$k_1(x;\alpha) \sim (\alpha\ \ln x)^{-2} \quad \text{as } x \to 0^+,$$

$$k_1(x;\alpha) \sim x^{-2}e^{-x} \quad \text{as } x \to +\infty,$$

and is otherwise continuous, the branch integral

$$I(t;\alpha) = -\alpha\int_0^\infty dxe^{-xt}k_1(x;\alpha)$$

converges uniformly and

$$\dot{f}(t;\alpha) = 2\ \mathrm{Re}[s_0\beta_0e^{s_0t}] + \alpha\int_0^\infty dxe^{-xt}xk_1(x;\alpha).$$

$\dot{f}(t;\alpha)$ is continuous on $[0, T]$ for any $T > 0$. Hence,

$$|\dot{f}(t;\alpha)| \le 2\ |s_0\beta_0|e^{\mathrm{Re}[s_0]t} + \alpha\int_0^\infty dxe^{-xt}xk_1(x;\alpha)$$

and

$$\int_0^t d\tau|\dot{f}(\tau;\alpha)| \le 2|s_0\beta_0|\,\mathrm{Re}\,[s_0]| + 2\ \mathrm{Re}[\beta_0] - 1,$$

since $k_1 > 0$, $\mathrm{Re}[s_0] < 0$ and $1 = f(0;\alpha) = 2\ \mathrm{Re}[\beta_0] - \alpha\int_0^\infty dxk_1(x;\alpha)$.

Using the estimates of s_0 given in Appendix A and the fact that $f(t;\alpha) = 1 + \int_0^\infty d\tau\dot{f}(\tau;\alpha)$, we find that

$$|f(t;\alpha)| \le 1 + \int_0^t d\tau|\dot{f}(\tau;\alpha)| < 6.5 \quad \text{for } \alpha < \alpha_1,$$

which states that $|f(t;\alpha)|$ is bounded uniformly.

The following fact is important enough to state by itself even though it was just used in proving the above.

Corollary: $\dot{f}(t;\alpha)$ is aboslutely integrable. An immediate consequence of the above absolute integrability of $\dot{f}$ is the boundedness of the solution of the inhomogeneous equation, Eq. (1.1). We may write[3]

$$F = 1 - \bar{\sigma} + (1 - \bar{\sigma})*\dot{f} \quad \text{where } \bar{\sigma} = \int_0^t d\tau\sigma(\tau). \tag{3.2'}$$

Hence we have the following:

Corollary: If the integral of the source term σ is bounded, then the solution F of Eq. (1.1) is bounded.

D. Decay of the solution as $t \to \infty$ for fixed α

Since $\mathrm{Re}[s_0] < 0$ for sufficiently small α, the residue contribution, $2\ \mathrm{Re}[\beta_0e^{s_0t}]$, is of exponentially small order as $t \to \infty$ and fixed α. Therefore, we need only investigate the decay of the branch integral, Eq. (C2) of Appendix C:

$$I(t;\alpha) = -\alpha\int_0^\alpha dxe^{-(t+1)x}/G_1(x;\alpha),$$

where we now write

$$G_1(x;\alpha) = g_1^2(x;\alpha) + \alpha^2\pi^2e^{-2x},$$

$$g_1(x;\alpha) = x + \alpha e^{-x}h_1(x),$$

$$h_1(x) = \ln x + \gamma + \int_0^x dy(e^{y} - 1)/y.$$

Because of the logarithmic singularity, Laplace's method is unsuitable for finding the asymptotic behavior of the integral, but the major contribution of the integrand still comes near the maximum $x_0 \simeq 2/(t\ \ln t)$, as it does in the application of Laplace's method.

Theorem 3.2: $I(t,\alpha) \sim -1/(\alpha t(\ln t)^2)$ as $t \to \infty$ for α fixed in $0 < \alpha_2 < \alpha < \alpha_0$ for some α_2 and α_0.

Proof: Let $z = (1 + t)$, $\mu_2 = \ln z/z$, $\mu_1 = 1/(z\ \ln z)$ and $\xi = zx$ so that as $t \to \infty$, $z \to \infty$, $z^{-1} \ll \mu_2 \ll 1$ and $0 < \mu_1 \ll z^{-1}$. The integral is decomposed into $I = I_1 + I_2 + I_3$, where

$$I_1 = -\alpha z^{-1}\int_{\mu_1 z}^{\mu_2 z} d\xi e^{-\xi}/G_1(\xi z^{-1};\alpha),$$

$$I_2 = -\alpha\int_0^{\mu_1} dxe^{-zx}/G_1(x;\alpha),$$

and

$$I_3 = -\alpha\int_{\mu_2}^\infty dxe^{-zx}/G_1(x;\alpha).$$

Now, $h_1'(x) = x^{-1}e^x, g_1'(x;\alpha) = 1 + \alpha x^{-1} - \alpha e^{-x}h_1$ and $G_1'(x;\alpha) = 2g_1(1 + \alpha/x + x - g_1) - 2\alpha^2\pi^2 e^{-2x}$. First, we show that the contribution from the tails of the integral is negligible. As $x \to 0^+, g_1 \sim \alpha \ln x < 0$ and $g_1' \sim \alpha/x > 0$ for $x < 1$ so that $G_1'(x;\alpha) < 0$ for sufficiently small x and $G_1(x;\alpha) \geq G_1(\mu_1;\alpha)$ when $0 \leq x \leq \mu_1$ for sufficiently large z. This leads to the estimate

$$|I_2| \leq (\alpha/G_1(\mu_1;\alpha))\int_0^{\mu_1} dx \sim \frac{\mu_1}{\alpha(\ln\mu_1)^2} \sim \frac{1}{\alpha z(\ln z)^3} \ll \frac{1}{\alpha z(\ln z)^2}$$

as $z \to \infty$ and α fixed. Since $G_1(x;\alpha) \geq \alpha^2\pi^2 e^{-2x}$,

$$|I_3| \leq \frac{1}{\alpha\pi^2}\int_{\mu_2}^{\infty} dx e^{-(z-2)x} \sim \frac{1}{\alpha\pi^2 z^2} \ll \frac{1}{\alpha z(\ln z)^2}$$

as $z \to \infty$ and α fixed.

Rewriting the main contribution,

$$I_1 = -\frac{1}{\alpha z(\ln z)^2} - \frac{1}{\alpha z(\ln z)^2} I_{12} + \frac{1}{\alpha z(\ln z)^2}(1 - e^{-\mu_1 z} + e^{-\mu_2 z}),$$

where $I_2 = \int_{\mu_1 z}^{\mu_2 z} d\xi e^{-\xi}[G_1(\xi/z;\alpha) - \alpha^2 \ln^2 z]/G_1(\xi/z;\alpha)$.

The error terms on the right are estimated as

$$(1 - e^{-\mu_1 z} + e^{-\mu_2 z}) \leq \frac{1}{\ln z} + \frac{1}{z} = 0(1) \quad \text{as } z \to \infty.$$

Since for $(\ln z)^{-1} = \mu_1 z \leq \xi \leq \mu_2 z = \ln z$, $|\ln\xi| \leq \ln\ln z$, it follows that $|G_1 - \alpha^2(\ln z)^2|/G_1 = 0(1)$ as $z \to \infty$ uniformly on $[\mu_1 z, \mu_2 z]$. Hence,

$$|I_{12}| = 0(1)\int_{\mu_1 z}^{\mu_2 z} d\xi e^{-\xi} = 0(1) \quad \text{as } z \to \infty.$$

Therefore,

$$I(t;\alpha) \sim -1/(\alpha z(\ln z)^2) \sim -1/(\alpha t(\ln t)^2) \quad \text{as } t \to \infty$$

and, hence,

$$f(t;\alpha) \sim -1/(\alpha t \ln^2 t) \quad \text{as } t \to \infty, \quad \text{for fixed } \alpha$$

in $0 < \alpha_2 < \alpha < \alpha_0$ for some α_2 and α_0. Thus, the long time behavior, which is important in the transport of cosmic rays, is *not* an exponential decay but a much slower, algebraic–logarithmic one. For short times, however, one could observe an exponential decay in this model. We show this in the following section.

E. Initial decay of the solution

We seek to show that the initial decay of $f(t;\alpha)$ is exponential. Although this is explicitly given in the residue terms, the contribution of the branch integral $I(t;\alpha)$ is not obvious. We show the following result:

Lemma 3.2: For some $T > 0$, $I(t;\alpha) \sim \exp(-\alpha t \ln 1/\alpha)$ as $\alpha \to 0^+$ when $t < T$. The same result is obtained by a formal "adiabatic" approximation (see, for instance, Ref. 4).

Proof: We make the following change of variables in the branch integral:

$$w = w(x) = \frac{1}{\alpha\pi}e^x g_1(x;\alpha) = \frac{1}{\pi}(\alpha^{-1}xe^x + \ln x + \gamma + e_1(x)),$$

where $e_1(x) = \int_0^x dy(e^y - 1)/y$. This tranformation has the properties that $w'(x) = (\alpha\pi)^{-1}e^x(2 + x + \alpha x^{-1} - \alpha x^{-2}) > 0$ for $x > 0$, $w(x) \to -\infty$ as $x \to 0^+$ and $w(x) \to +\infty$ as $x \to +\infty$. Hence, by the inverse function theorem, there exists a function $g_2(w)$ such that $x = g_2((\alpha\pi)^{-1}e^x g_1(x;\alpha))$ and $g_2'(w) = 1/w'(x)$.

The branch integral takes the form

$$I(t;\alpha) = -\pi^{-1}\int_{-\infty}^{\infty} dw g_3(w)/(1 + w^2),$$

where $g_3(w) = \exp(-tg_2(w))/(1 + g_2(w) + \alpha/g_2(w))$. For sufficiently small α and not too large t, the integral achieves its major contribution near $w = 0$. Let $x_0 = g_2(0)$. Since $w'(x) > 0$, $w(\alpha(L_1 - L_2)) \sim \gamma/\pi > 0$ as $\alpha \to 0^+$, $w(\alpha(L_1 - L_2 - \gamma)) \sim -L_2/(\pi L_1) < 0$ as $\alpha \to 0^+$, and $x_0 \sim \alpha(L_1 - L_2) \sim -\mathrm{Re}[s_0]$ as $\alpha \to 0^+$, where $L_1 = \ln 1/\alpha$ and $L_1 \equiv \ln\ln 1/\alpha$.

We choose $x_1 = \alpha L_1$ and $x_2 = \alpha(L_1 - 2L_2)$ so that the interval $[w_1, w_2]$, with $w_1 = w(x_1) \sim -L_2/\pi$ as $\alpha \to 0^+$ and $w_2 = w(x_2) \sim L_2/\pi$ as $\alpha \to 0^+$ includes almost all of the area of the branch integral. We write

$$I(t;\alpha) = -\frac{1}{\pi}\left(\int_{w_1}^{w_2} + \int_{w_2}^{\infty} + \int_{-\infty}^{w_1}\right) dw g_3(w)/(w^2 + 1) = I_1 + I_2 + I_3.$$

In the interval $-w_1(\alpha) \leq w \leq w_2(\alpha)$, we use the mean value theorem to obtain $g_2(w) = x_0 + g_2'(w_3)w$ where w_3 is between 0 and w and $g_3(w) = g_3(0) + g_3'(w_4)w$ (where w_4 is between 0 and w). We estimate the derivatives below, letting $x_3 = g_2(w_3)$ and $x_4 = g_2(w_4)$:

$$|g_2'(w_3)w| = |w|/w'(x_3) \leq \max(|w|)/w'(x_2) \sim \alpha L_2 \quad \text{as } \alpha \to 0^+$$

and

$$|g_3'(w_4)w| = (g_2'g_3|t + (1 - \alpha/g_2^2)/(1 + g_2 + \alpha/g_2)|)_{w=w_4}|w| \leq \exp(-x_1 t)(t + 1/(\alpha(L_1 - L_2)^2))\max(|w|)/w'(x_2) \sim (\alpha L_2 t + L_2/L_1^2) \times \exp(-\alpha L_1 t) \quad \text{as } \alpha \to 0^+.$$

We have used the facts that $w''(x) = (\alpha\pi)^{-1}e^x(2 + x + \alpha|x - \alpha|x^2) < 0$ on $[x_1, x_2]$ {i.e., $w'(x)$ is a decreasing function on $[x_1, x_2]$} and $(1 + g_2 + \alpha/g_2)^{-1} \leq 1$.

Now,

$$\int_{w_2}^{\infty} dw/(1 + w^2) \sim \int_{-\infty}^{w_1} dw/(1 + w^2) \sim 1/w_2 \sim \pi/L_2 = 0(1)$$

and

$$\pi^{-1}\int_{w_1}^{w_2} dw g_3'(w_4)w/(1 + w^2) \sim \exp(-\alpha L_1 t)(\alpha L_2 t + L_2/L_1^2) = \exp(-\alpha L_1 t)\cdot 0(1)$$

provided $t = 0(1/(\alpha L_2))$.

Hence, as $\alpha \to 0^+$,

$$I_1 = -\pi^{-1}g_3(0)\int_{-\infty}^{\infty} dw/(1 + w^2) + \pi^{-1}g_3(0)\left(\int_{-\infty}^{w_1} + \int_{w_2}^{\infty}\right) \times dw/(1 + w^2) - \frac{1}{\pi}\int_{w_1}^{w_2} dw g_3'(w_4)w/(1 + w^2) \sim -g_3(0) \sim -\exp(-\alpha(L_1 - L_2)t) \sim \exp(-\alpha L_1 t) \quad \text{if } t = 0\left(\frac{1}{\alpha L_2}\right),$$

$$|I_2| \leq \pi^{-1}e^{-x_2 t}\int_{w_2}^{\infty} dw/(1 + w^2) \leq (\pi w_2)^{-1}e^{-x_2 t} = 0(I_1),$$

$$|I_3| \leq \pi^{-1}\int_{-\infty}^{w_1} dw/(1 + w^2) \leq -(\pi w_1)^{-1} \sim 1/L_2 = 0(I_1) \quad \text{if } t = 0\left(\frac{1}{\alpha L_2}\right).$$

Finally, we have the following result giving the initial decay of $f(t;\alpha)$ since $2\,\mathrm{Re}[\beta_0 e^{s_0 t}] \sim 2\exp(-\alpha L_1 t)$:

Theorem 3.3: $f(t,\alpha) \sim \exp(-\alpha L_1 t)$ as $\alpha \to 0^+$ and $t = 0(1/\alpha L_2)$.

F. Solution of equation with constant source

When the source term in the inhomogeneous equation, Eq. (1.1) is constant, i.e., $\sigma(t) = \sigma_0$, then

$$\dot{F} = -\sigma_0 - \alpha K^* F, \qquad F(0,\alpha) = 1.$$

The solution of this equation can be written in terms of the solution $f(t;\alpha)$ of the homogeneous equation by Eq. (3.5) as

$$F = f - \sigma_0 \bar{f},$$

where $\bar{f}(t;\alpha) = \int_0^t d\tau f(\tau;\alpha)$. We seek the properties of $F(t;\alpha)$.

According to a Tauberian theorem,[5]

$$\bar{f}(+\infty;\alpha) = \tilde{f}(0^+;\alpha);$$

that is, the limit of the integral of f as $t \to \infty$ is equal to the limit of the Laplace transform of f as $s \to 0^+$. However, $\tilde{f}(s;\alpha) \to -(\alpha \ln s)^{-1} \to 0$ as $s \to 0^+$, so that $\bar{f}(+\infty;\alpha) = 0$ and we can write $\bar{f}(t;\alpha) = -\int_t^\infty d\tau f(\tau;\alpha)$ or

$$F(t;\alpha) = f(t;\alpha) + \sigma_0 \int_t^\infty d\tau f(\tau;\alpha).$$

Now, $f(t;\alpha)$ is integrable as $t \to \infty$ and, in fact,

$$f(t;\alpha) \sim -(\alpha t \ln^2 t)^{-1} \quad \text{as } t \to \infty, \quad \alpha_2 < \alpha < \alpha_1$$

since the branch integral has that behavior. Therefore, we can interchange asymptotic limit and integral,

$$\int_t^\infty d\tau f(\tau;\alpha) \sim -\frac{1}{\alpha}\int_t^\infty d\tau/(\tau (\ln\tau)^2) \sim -(\alpha \ln t)^{-1}$$

as $t \to \infty$ and we have the following result:

Theorem 3.4: $F(t;\alpha) \sim -\sigma_0/(\alpha \ln t)$ as $t \to \infty$, $\alpha_2 < \alpha < \alpha_1$ and fixed. Therefore, the decay of the solution of the inhomogeneous equation has a slower pure logarithmic decay, whereas the solution of the homogeneous equation has an algebraic–logarithmic decay as $t \to \infty$.

4. THE UNIFORMLY VALID APPROXIMATION

A. The approximation

Since $\bar{K}(t) = \ln(1+t)$, the homogeneous equation has the form

$$f(t;\alpha) = 1 - \alpha \int_0^t d\tau \ln(1+t-\tau) f(\tau;\alpha).$$

The kernel $\bar{K}$ is monotone-increasing and its derivative is monotone-decreasing. The major contribution of the kernel to the integral would come from large values of $t-\tau$. When τ is close to t, the kernel is small; and if we expect the solution to be bounded, it is not unreasonable to ignore the contribution from values of τ close to t and retain only the first term. Indeed, it turns out that this approximation of the kernel yields a uniform approximation to the solution. We prove this fact in the sequel. Letting $g(1+t;\alpha)$ by the simplifying approximation to $f(t;\alpha)$ and $z = 1+t$, $g(z;\alpha)$ satisfies the equation

$$g(z;\alpha) = 1 - \alpha \ln z \int_1^z d\zeta g(\zeta;\alpha).$$

This equation can be readily solved by viewing it as a linear first-order differential equation for $G(z;\alpha) = \int_1^z d\zeta g(\zeta;\alpha)$,

$$g(z;\alpha) = 1 - \alpha \ln z G(z;\alpha) \tag{4.1}$$

and

$$G(z;\alpha) = \exp(-\alpha z(\ln z - 1))\int_1^z d\zeta \exp(\alpha\zeta(\ln\zeta - 1)). \tag{4.2}$$

The error due to this approximation is $h(z;\alpha) = f(t;\alpha) - g(z;\alpha)$ and satisfies

$$h(z;\alpha) + \alpha \int_1^z d\zeta \ln(1+z-\zeta) h(\zeta;\alpha) = \phi(z;\alpha), \tag{4.3}$$

where

$$\phi(z;\alpha) = \alpha \ln z \int_1^z d\zeta g(\zeta;\alpha) - \alpha \int_1^z d\zeta \ln(1+z-\zeta) g(\zeta;\alpha). \tag{4.4}$$

B. Properties of $g(z;\alpha)$

We list and prove some useful properties of the function $g(z;\alpha)$.

(1) $G(z;\alpha) = \int_1^z d\zeta g(z;\alpha)$ is always nonnegative. This is transparent from Eq. (4.2).

(2) $g \to 0$ as $z \to \infty$ for fixed α. Proof follows from an application of L'Hopital's rule on $G(z;\alpha)$ which gives that $G(z;\alpha) \to 1/(\alpha \ln z)$ as $z \to \infty$.

(3) g has one and only one zero. Existence follows since $G(1;\alpha) = 0$ and $G(z;\alpha) \to 0$ as $z \to \infty$, there must be an intermediate point z_0 such that $1 < z_0 < \infty$ and $G'(z_0;\alpha) = 0$ where the prime denotes differentiation with respect to z. However, $G'(z;\alpha) = g(z;\alpha)$ so that $g(z_0;\alpha) = 0$. To establish uniqueness, note that $g' = -(\alpha \ln z)g - \alpha G/z$. Since $G(z;\alpha) > 0$ for $z > 1$, at any zero, say z_0, $g'(z_0;\alpha) = -\alpha G(z_0;\alpha)/z_0 < 0$. However, the sign of the derivative at adjacent zeros must be opposite or zero so that there can only be one zero of $g(z;\alpha)$ which we will call z_0.

Hence, g is unity at $z = 1$, decreases to zero at $z = z_0$, and remains negative as it asymptotically approaches zero as $z \to \infty$.

C. Estimate of the zero of g

Letting $g(z_0;\alpha) = 0$, we begin with a coarse estimate of z_0 and proceed to refine it.

(1) Coarse estimate: $z_0 \geq (\alpha \ln 1/\alpha)^{-1}$ for $\alpha < e^{-1}$. Let $U(z) = \exp(\alpha z(\ln z - 1))$ so that

$$g(z;\alpha) = 1 - \alpha \ln z (U(z))^{-1}\int_1^z d\zeta U(\zeta), \tag{4.5}$$

and when

$$z = z_0, 1 = \alpha \ln z_0 (U(z_0))^{-1}\int_1^{z_0} d\zeta U(\zeta). \tag{4.6}$$

For $1 \leq z \leq z_0$, $z(\ln z - 1) \leq z_0(\ln z_0 - 1)$; hence, $U(z) \leq U(z_0)$ and

$$1 \leq \alpha \ln z_0 \int_1^{z_0} d\zeta = \alpha \ln z_0 (z_0 - 1) \leq \alpha z_0 \ln z_0. \tag{4.7}$$

To show that $z_0 \geq 1/(\alpha \ln 1/\alpha)$ for sufficiently small α, suppose the contrary that $1 < z_0 < 1/\alpha \ln 1/\alpha)$. Let $L_1 = \ln 1/\alpha$ and $L_2 = \ln L_1 = \ln\ln 1/\alpha$, $0 < \alpha z_0 \ln z_0 < (1 - L_2/L_1) < 1$ for $0 < \alpha < e^{-1}$. This is a contradiction of Eq. (3.7) and hence $z_0 \geq 1/(\alpha \ln 1/\alpha)$.

(2) Second estimate: $z_0 > L_2/(\alpha L_1)$ for $\alpha < \alpha_1$. Since $\ln z \leq \ln z_0$ for $1 < z \leq z_0$, Eq. (4.6) becomes

$$1 \leq \frac{\alpha \ln z_0}{U(z_0)}\int_0^{z_0} dz \exp(\alpha z(\ln z_0 - 1)) \leq \frac{\ln z_0}{\ln z_0 - 1}\left(1 - \frac{1}{U(z_0)}\right),$$

and with further simplifications, $\ln z_0 \le U(z_0) \le \exp(\alpha z_0 \ln z_0)$. Upon taking logarithms,

$$\ln\ln z_0 \le \alpha z_0 \ln z_0, \tag{4.8}$$

where $\alpha < \alpha_1 \equiv 1/\exp(\exp(e)) < 3 \times 10^{-7}$ so that $e^e < L_1 < \alpha^{-1}$ and $e < L_2 < L_1$. If we suppose that $1/(\alpha L_1) \le z_0 \le L_2/(\alpha L_1)$, then $z_0 \ge 1/(\alpha_1 \ln(1/\alpha_1)) > e^e$. Since $z_0 \ln z_0/\ln\ln z_0$ is an increasing function of z_0 for $z_0 > e^e$,

$$\alpha z_0 \ln z_0/\ln\ln z_0 \le (1-\epsilon_1)/(1 + \ln(1-\epsilon_1)/L_2),$$

where $\epsilon_1 = (L_2 - L_3)/L_1$, $L_3 \equiv \ln L_2 = \ln\ln\ln 1/\alpha$ and $1 < L_3 < L_2$. But

$$(1 + \ln(1-\epsilon_1)/L_2)/(1-\epsilon_1) \ge 1 + \epsilon_1[1 - 1/(L_2(1-\epsilon_1))] \ge 1$$

since $\ln(1-\epsilon_1) \ge -\epsilon_1/(1-\epsilon_1)$ and $0 < \epsilon_1 < (e-1)/e^e < 0.15 < 1$. Hence, $\alpha z_0 \ln z_0/\ln\ln z_0 \ge 0$, which contradicts Eq. (4.8) so that $z_0 > L_2/(\alpha L_1)$.

(3) Third estimate: $z_0 < \alpha^{-1}$ for $\alpha < \alpha_1$. Integrating Eq. (4.5) by parts when $e < \zeta < z$, one obtains

$$g(z;\alpha) = 1 - \frac{\alpha \ln z}{U(z)}\left(\int_1^e d\zeta U(\zeta) + \frac{U(z)}{\alpha \ln z} + \frac{1}{\alpha} + \frac{1}{\alpha}\int_e^z \frac{d\zeta U(\zeta)}{\zeta \ln^2\zeta}\right) = \frac{\ln z}{U(z)}\left(1 - \int_e^z d\zeta \frac{U(\zeta)}{\zeta \ln^2\zeta} - \alpha\int_1^e d\zeta U(\zeta)\right).$$

Since $\alpha\zeta(\ln\zeta - 1) = \alpha\zeta(\ln z - 1) + \alpha(\ln\zeta - \ln z) \ge \alpha\zeta(\ln z - 1) - \alpha z/e$ for $e < \zeta < z$ and $(\zeta \ln^2\zeta)^{-1} \ge (z \ln^2 z)^{-1}$,

$$g(z) \le \frac{\ln z}{U(z)}\left(1 - \frac{\exp(-\alpha z/e)U(z)}{\alpha z \ln^3 z} \times [1 - \exp(-\alpha(z-e)(\ln z - 1))]\right) \tag{4.9}$$

and

$$g(1/\alpha) \le -\frac{1}{L_1^2 \exp(1/e)}[1 - \alpha e L_1^3 \exp((1/e) - \alpha e)] < 0$$

for $\alpha < \alpha_1$. Therefore, $z_0 < 1/\alpha$, since there can only be one change in sign of g with $g(1) = 1$.

(4) Fourth estimate: $z_0 < 4L_2/(\alpha L_1)$ for $\alpha < \alpha_1$. Using Eq. (4.9) with $z = z_0$, we have

$$1 \ge \frac{\exp(-\alpha z_0/e)}{\alpha(\ln z_0 - 1)z_0 \ln^2 z_0}[U(z_0) - e^{\alpha e(\ln z_0 - 1)}]$$

or

$$\ln^3 z_0 \ge (\ln z_0 - 1)\ln^2 z_0 \ge \frac{1}{\alpha z_0} e^{-\alpha z_0/e + \alpha z_0(\ln z_0 - 1)} \times (1 - e^{-\alpha(z_0 - e)(\ln z_0 - 1)})$$

or by taking logarithms

$$3\ln\ln z_0 \ge \alpha z_0 \ln z_0 - \alpha z_0(1 + 1/e) - \ln\alpha z_0 + \ln(1 - e^{-\alpha(z_0 - e)(\ln z_0 - 1)}). \tag{4.10}$$

Upon assuming the contrary hypothesis that $4L_2/(\alpha L_1) \le z_0 \le \alpha^{-1}$ for $\alpha < \alpha_1 = 1/\exp(\exp(e))$ and letting $\epsilon_2 = (L_2 - L_3 - \ln 4)/L_1$, $0 < \epsilon_2 < 0.025$, we have an estimate for the first term on the right of Eq. (4.10)

$$\alpha z_0 \ln z_0/(3\ln\ln z_0) \ge \frac{4}{3}(1-\epsilon_2)/\left(1 + \frac{1}{L_2}\ln(1-\epsilon_2)\right) \ge \frac{4}{3}(1-\epsilon_2) \ge 1.3,$$

for the second term

$$-\alpha z_0(1 + 1/e)/(3\ln\ln z_0) \ge -\frac{1}{3}(1 + 1/e)/\left[L_1\left(1 + \frac{1}{L_2}\ln(1-\epsilon_2)\right)\right] \ge 0.034$$

with

$$\epsilon_3 = \exp(-\alpha(z_0 - e)(\ln z_0 - 1)) < \exp(-4L_2(1-\epsilon_2)) + e\alpha L_1 < 4.6 \times 10^{-5},$$

and for the third term

$$\ln(1-\epsilon_3)/(3\ln\ln z_0) \ge -\epsilon_3/(3\ln\ln z_0(1-\epsilon_3)) \ge -6 \times 10^{-6}.$$

Adding the above inequalities, we have

$$\alpha z_0 \ln z_0/\ln\ln z_0 \ge 1.2 > 1,$$

which contradicts Eq. (4.10). Hence, $z_0 < 4L_2/L_1$; and summarizing the third and fourth estimates, we have

$$L_2/(\alpha L_1) < z_0 < 4L_2/(\alpha L_1) \quad \text{for } \alpha < \alpha_1,$$

or that the zero of the approximation $g(z;\alpha)$ has the order

$$z_0 = 0(L_2/(\alpha L_1)) \quad \text{as } \alpha \to 0^+.$$

D. Bound on $\phi(z;\alpha)$

In this subsection, we seek a bound of the function appearing in Eq. (4.4) which we write

$$\phi(z;\alpha) = -\alpha\int_1^z d\zeta g(\zeta)\ln(1 - (\zeta - 1)/z).$$

This is the source term in Eq. (4.3) for the error $h(z;\alpha)$ in the approximation $g(z;\alpha)$. We bound ϕ in order to bound h. We show

Lemma 4.1: $\phi(z;\alpha) = 0(1)$ as $\alpha \to 0^+$ uniformly in z. For $1 \le z \le z_0$, $g(z;\alpha) \ge 0$ and since $g(1) = 1$ and

$$g'(z;\alpha) = -\frac{\alpha}{zU(z)}\int_1^z d\zeta U(\zeta) - \alpha \ln z g(z;\alpha) \le 0,$$

We have that $g(z;\alpha) \le 1$ on $[1, z_0]$. Also, when $1 \le \zeta \le z$,

$$0 \le (\zeta - 1)/z < (z-1)/z < 1$$

and

$$|\phi(z;\alpha)| \le \alpha\int_1^z d\zeta |\ln(1 - (\zeta - 1)/z)| \le -\alpha z\int_{z^{-1}}^1 d\zeta \ln\zeta \le \alpha z \le \alpha z_0$$

for $z \le z_0$. However, $\alpha z_0 \le 4L_2/L_1 = 0(1)$ as $\alpha \to 0^+$, so that $\phi(z;\alpha) = 0(1)$ as $\alpha \to 0^+$ when $z \le z_0$.

When $z \ge z_0$, we write $\phi(z;\alpha) = \phi_1(z;\alpha) + \phi_2(z;\alpha)$ where

$$\phi_1(z;\alpha) = -\alpha\int_1^{z_0} d\zeta g(\zeta;\alpha)\ln(1 - (\zeta - 1)/z)$$

and

$$\phi_2(z;\alpha) = -\alpha\int_{z_0}^z d\zeta g(\zeta;\alpha)\ln(1 - (\zeta - 1)/z).$$

The first integral ϕ_1 has the properties

$$\phi_1'(z;\alpha) = -\alpha\int_1^{z_0} d\zeta g(\zeta;\alpha)(\zeta - 1)/(z(z + 1 - \zeta)) < 0$$

and $0 \le \phi_1(z;\alpha) \le \phi_1(z_0;\alpha) \le \alpha z_0 = 0(1)$ as $\alpha \to 0^+$, as before.

For the estimate of $\phi_2(z;\alpha)$, we need the decay of $g(z;\alpha)$ as $z \to \infty$. Still assuming that $z \ge z_0$ and integrating Eq. (4.5) by parts on $[z_0, z]$ only, we have

$$g(z;\alpha) = 1 - \frac{\alpha \ln z}{U(z)}\left(\int_1^{z_0} d\zeta U(\zeta) + \frac{U(z)}{\alpha \ln z} - \frac{U(z)}{\alpha \ln z_0} + \int_{z_0}^{z} \frac{d\zeta U(\zeta)}{\alpha\zeta \ln^2\zeta}\right)$$

The fact that z_0 is a zero leads to a cancellation,

$$g(z;\alpha) = -\frac{\alpha \ln z}{U(z)} \int_{z_0}^{z} \frac{d\zeta U(\zeta)}{\alpha\zeta \ln^2\zeta}.$$

A second integration by parts yields

$$g(z;\alpha) = \frac{-1}{\alpha z \ln^2 z} - \frac{\alpha \ln z}{U(z)} \int_{z_0}^{z} d\zeta \frac{(3+\ln\zeta)U(\zeta)}{\alpha^2\zeta^2 \ln^4\zeta} - \frac{\ln z U(z_0)}{\alpha z_0 \ln^3 z_0 U(z)}, \quad (4.11)$$

and the estimate

$$|g(z;\alpha)| \le \frac{1}{\alpha z \ln^2 z} + \frac{\alpha \ln z}{U(z)} \frac{(3+\ln z_0)}{\alpha z_0 \ln^2 z_0} \int_{z_0}^{z} d\zeta \frac{U(\zeta)}{\alpha\zeta \ln^2\zeta}$$

$$\le \frac{1}{\alpha z \ln^2 z} + \frac{(3+\ln z_0)}{\alpha z_0 \ln^2 z_0} |g(z;\alpha)|.$$

Now when $\alpha < \alpha_1$, $L_2/(\alpha L_1) < z_0 < 4L_2/(\alpha L_1)$, so that

$$(3+\ln z_0)/(\alpha z_0 \ln^2 z_0) < (1 + 3/(L_1(1-\epsilon_1))/(L_2(1-\epsilon_1)) < 0.58$$

and $|g(z;\alpha)| < 2.4/(\alpha z \ln^2 z)$. The estimate on the second integral is

$$|\phi_2'(z;\alpha)| \le 2.4 \int_{z_0}^{z} d\zeta |\ln(1-(\zeta-1)/z)| (\zeta \ln^2\zeta)$$

$$\le \frac{2.4}{\ln^2 z_0} \int_{z_0/z}^{1} d\xi |\ln(1-\xi+1/z)|/\xi$$

$$\le \frac{2.4}{\ln^2 z_0} \int_0^1 d\xi |\ln(1-\xi)|/\xi$$

or

$$|\phi_2(z;\alpha)| \le 2.4 \frac{\pi^2}{6}/\ln^2 z_0.$$

However, for $\alpha < \alpha_1$, $1/\ln^2 z_0 < 1/(L_1^2(1-\epsilon_1)^2) = 0(1)$ as $\alpha \to 0^+$. Therefore, $\phi(z;\alpha) = 0(1)$ as $\alpha \to 0^+$ uniformly in z.

E. Uniform validity of $g(z;\alpha)$

In order to show that $g(1+t;\alpha)$ is a uniformly valid approximation to $g(t;\alpha)$ as $\alpha \to 0^+$, we must show that the absolute error $h(z;\alpha)$ is uniformly small as $\alpha \to 0^+$.

Lemma 4.2: $\phi(z;\alpha) = 0(1)$ uniformly for $z \ge 1$ as $\alpha \to 0^+$ implies that $h(z;\alpha) = 0(1)$ uniformly for $z \ge 1$ as $\alpha \to 0^+$.

Since the kernel $\overline{K}$ in Eq. (4.3) for h is the same as Eq. (2.3) for f, the solution h may be written in terms of t and the source term ϕ so that[3]

$$h = \phi + \phi * \dot{f}. \quad (4.12)$$

From the result of Sec. 3C we have that $\dot{f}$ is continuous; and from the previous section we have that $\phi = 0(1)$ as $\alpha \to 0^+$ uniformly in z so that it is certainly bounded.

Therefore,

$$|h| \le |\phi| + |\phi| * |\dot{f}| \le \sup_z[|\phi|][1 + \int_0^\infty d\tau |\dot{f}(\tau;\alpha)|]$$

$$\le 12.5 \sup_z[|\phi(z;\alpha)|] = 0(1) \quad \text{as } \alpha \to 0^+.$$

Hence, $h = 0(1)$ uniformly in $z = 1 + t$ as $\alpha \to 0^+$. But $|h|$ is the absolute error between the solution $f(t;\alpha)$ and the approximation $g(1+t;\alpha)$ so that g is a uniformly valid approximation to $f(t;\alpha)$ for $t \ge 0$ as $\alpha \to 0^+$.

In addition, we can show that $h = f - g \to 0$ as $t \to \infty$ faster than either f or g for any fixed $\alpha > 0$. Returning to the twice-integrated-by-parts form for g, Eq. (4.11), we note that $U(z)$ is exponentially large compared to z and $\ln z$ and $z \to +\infty$. For this reason, the first term on the right of Eq. (4.11), $-1/(\alpha z \ln^2 z)$, is of larger order than the last since

$$\frac{\ln 2 U(z_0)}{\alpha z_0 \ln^3 z_0 U(z)} \Big/ \frac{1}{\alpha z \ln^2} \to 0 \quad \text{as } z \to +\infty.$$

The middle term on the right of Eq. (4.11) is also of smaller order than the first since by L'Hopital's rule

$$(\alpha z \ln^2 z)\frac{\alpha \ln z}{U(z)} \int_{z_0}^{z} d\zeta \frac{(3+\ln\zeta)U(\zeta)}{\alpha^2\zeta^2 \ln^4\zeta} \to \frac{3+\ln z}{(z\ln^2 z - \ln z - 3)} \to 0$$

as $z \to \infty$. Hence,

$$g(z;\alpha) = -\frac{1}{\alpha z \ln^2 z}(1 + 0(1)) \quad \text{as } z \to +\infty,$$

while $f(z;\alpha)$ has a similar behavior as $z \to +\infty$ so that

$$h = -\frac{0(1)}{\alpha z \ln^2 z} \quad \text{as } z \to +\infty,$$

i.e., decaying faster than either f or g as $z \to +\infty$. Thus, the relative error h/f tends to zero for every fixed $\alpha > 0$ and sufficiently small.

5. CONCLUSIONS

A uniformly valid solution has been found for a Volterra integral equation with a logarithmic kernel multiplied by a small parameter. We have proved that the absolute error tends to zero uniformly in the infinite interval as the small parameter tends to zero. Furthermore, for every fixed $\alpha > 0$ and sufficiently small, the relative error tends to zero for long times. The work was done in three steps. First, we found a suitable approximation to the kernel that would yield a uniformly valid solution. Second, we proved boundedness of the solution by using transform methods. Third, we gave a direct proof of the smallness of the error. We obtained a single expression valid not only for finite times but also exhibiting a slow algebraic-logarithmic decay for long times, which we proved to be the correct asymptotic approximation.

In this paper, we considered an equation that arises specifically in study of the diffusion of charged particles in a strong magnetic field.[1] It may not be amiss to remark that equations of the same type occur also in engineering, biology, and nuclear physics.

APPENDIX A. EXISTENCE OF ONLY TWO ROOTS OF $D(s;\alpha)$

The following result describes the singularities of $\bar{f}(s;\alpha)$ through the two zeros of $D(s;\alpha)$ in the left-hand plane and a branch point at $s = 0$.

Lemma A1: $D(s;\alpha) = s + \alpha e^s E_1(s)$ has only two zeros in $S = \{s = re^{i\theta}: 0 < r < \alpha, -\pi < \theta < \pi\}$ for α sufficiently small.

Proof: The exponential integral[6] can be written

$$E_1(s) = -\ln s - \gamma + e_1(s),$$

where γ is the Euler's constant and $e_1(s) = \int_0^s dy(1 - e^{-y})/y = -\sum_{n=0}^{\infty}(-s)^n/(n!n)$ is analytic. The only singularity of $D(s;\alpha)$ is the branch point of the logarithm at $s = 0$. D is analytic in s and we place the branch cut on the negative real axis.

The argument principle[7] is used to find the number of zeros of D. Consider the positive-oriented circular contour with a slit along the negative real axis $\Gamma = U_{i=1}^5 \Gamma_i$, where

$$\Gamma_1 = \{s = Re^{i\theta}: 0 \le \theta \le (\pi\text{-}\arctan(\epsilon/\sqrt{R^2 - \epsilon^2})\},$$

$$\Gamma_2 = \{s = x + i\epsilon: -\sqrt{R^2 - \epsilon^2} \le x \le 0\},$$

$$\Gamma_3 = \{s = \epsilon e^{i\theta}: -\pi/2 \le \theta \le +\pi/2\},$$

$$\Gamma_4 = \{s = x - i\epsilon: -\sqrt{R^2 - \epsilon^2} \le x \le 0\},$$

$$\Gamma_5 = \{s = Re^{i\theta}: (-\pi + \arctan(\epsilon/\sqrt{R^2 - \epsilon^2}) \le \theta \le 0\}.$$

(See Fig. 1.)

D is analytic in the interior of Γ so that the number of zeros of D is given by the change in the argument of D on Γ divided by 2π.

The zeros of D occur in conjugate pairs since $D^*(s;\alpha) = D(s^*;\alpha)$ and the change in argument will be the same on both portions of Γ in the upper and lower half planes, respectively.

Along Γ_1 with R sufficiently large, the asymptotic expansion of $E_1(s)$[5] yields

$$e^s E_1(s) \sim s^{-1} \sum_{n=0}^{\infty} n!(-s)^n \quad \text{as } |s| \to \infty| \arg(s)| < \frac{3\pi}{2},$$

so that for sufficiently large $|s|$ the change of the argument is

$$\Delta_1 \arg[D] = \Delta_1 \arg[s] + \Delta_1 \arg[1 + \alpha s^{-1} e^s E_1(s)] \sim \Delta_1 \arg[s] \cong \pi.$$

We need not be concerned with the error since the result will be an integral multiple of π.

On Γ_2 with $\epsilon \ll 1$ and $R \gg 1$, let $D = \rho e^{i\phi}$ and take $x \simeq re^{i\pi}$ ignoring ϵ so that $D \simeq U + iV$, where $U = -r + \alpha e^{-r}(-\ln r - \gamma - \int_0^r dy(e^y - 1)/y$ and $V = -\alpha\pi e^{-r}$. V is real and negative and U has only one zero, as can be seen from the asymptotic behavior

$$U \sim -r \quad \text{as } r \to \infty, \quad (\text{i.e., } \tan\phi \to 0^+),$$

$$U \sim \alpha \ln\frac{1}{r} \quad \text{as } r \to 0^+ \quad (\text{i.e., } \tan\phi \to 0^-)$$

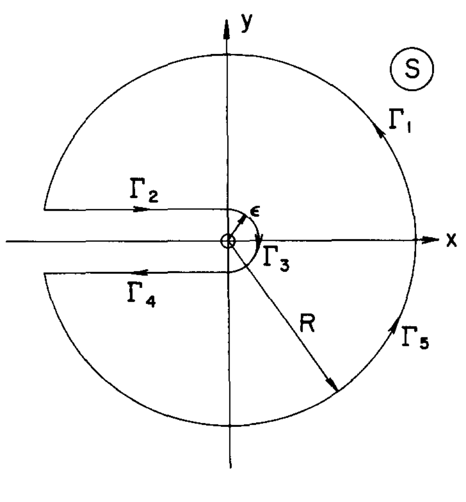

FIG. 1

and

$$\frac{\partial U}{\partial r} = -1 + \alpha u(r) - \alpha/r - \alpha e^{-r} E_1(r),$$

where $u(r) = e^{-r}\int_0^r dy(e^y - 1)/y$. $u(r)$ is positive and vanishes when $r \to 0^+$ or when $r \to \infty$ and otherwise can be no larger than 2. Hence, $\partial U/\partial r < 0$ for $\alpha < 1/2$ so that $U(r)$ is a decreasing function and can have only one zero. The change in argument on Γ_2 is

$$\Delta_2 \arg[D] \simeq \pi.$$

On Γ_3 with ϵ sufficiently small, $\mathrm{Re}[D] \sim \alpha \ln 1/\epsilon$ and $\mathrm{Im}[D] \sim \alpha\theta + \epsilon \sin\theta$ and hence $\tan(\arg[D] \sim (-\alpha\theta + \epsilon \sin\theta)/(\alpha \ln 1/\epsilon) \to 0$ as $\epsilon \to 0^+$. Thus on Γ_3, $\Delta_3 \arg[D]$ can be made arbitrarily small compared to π.

Since Γ, the sum of the partial contours, is a closed Jordan curve and $\Delta \arg[D] \simeq 2(\pi + \pi) = 4\pi$, $D(s;\alpha)$ has exactly two roots in S by the argument principle.

For later development, we will need the asymptotic behavior of the zeros, s_0 and s_0^* of $D(s;\alpha)$. Since we know these exist in conjugate pairs, we investigate only the upper half plane.

Let $s = re^{i\theta}$, $r > 0$ so that

$$E_1(re^{i\theta}) = \int_{r\exp(i\theta)}^{\infty} dy e^{-y}/y \tag{A1}$$

or

$$E_1(re^{i\theta}) = -\ln r - i\theta - \gamma + e_1(re^{i\theta}). \tag{A2}$$

Now, $D(re^{i0^+};\alpha) = r + \alpha e^r \int_0^r dy e^{-y}/y > 0$ and $\mathrm{Im}[D(re^{i\pi^-};\alpha)] = -\alpha\pi \exp(-r) < 0$ so that $0 < \theta < \pi$.

To find a lower bound on $|s_0|$ we observe that $D(re^{i\theta};\alpha) \to -\alpha \ln r$ as $r \to \theta$. By the triangular inequality,

$$|D(re^{i\theta};\alpha)| \ge ||\alpha \ln r| - |D(re^{i\theta};\alpha) + \alpha \ln r||. \tag{A3}$$

Since $|e^{r\exp(i\theta)}| \le e^r$, $|(e^{r\exp(i\theta)}e_1(r\exp(i\theta))| \le re^r$ and $|e^{r\exp(i\theta)} - 1| \le re^r$,

$$|D(re^{i\theta};\alpha) + \alpha \ln r| \le r + \alpha e^r(\pi + \gamma + r(1 + \ln r|)).$$

Now, suppose $0 < r \le 1/2\ \alpha \ln 1/\alpha$ and let $0 < \alpha < \alpha_1$, where $\alpha_1 = \exp(-\exp(e))$, then $|D(re^{i\theta};\alpha) + \alpha \ln r| < 0.55\alpha \ln 1/\alpha$ and $|\alpha \ln 1/r| > 0.85\alpha \ln 1/\alpha$ so that $|D(re^{i\theta};\alpha)| > 0$ by (A3). Therefore, $|s_0| = r_0$ cannot be contained in this interval and a lower bound must be $|s_0| > 1/2\alpha \ln 1/\alpha$.

We find an upper bound for $|s_0|$ by noting that $D(re^{i\theta};\alpha) \sim r$ as $r \to \infty$ and

$$|D(re^{i\theta};\alpha)| \ge |r - |D(re^{i\theta};\alpha) - r||. \tag{A4}$$

To estimate this, we relate $E_1(re^{i\theta})$ to $E_1(r)$ by

$$E_1(re^{i\theta}) = E_1(r) - i\int_0^\theta d\theta' \exp(-re^{i\theta'}),$$

by using an alternate contour. Hence, $\alpha| \exp(re^{i\theta}) E_1(re^{i\theta})| \le \alpha/r + \alpha\pi$. Suppose that $2\sqrt{\alpha} \le r < \infty$ when $0 < \alpha < \alpha_1$, then $(\alpha/r + \alpha\pi) < 0.6\sqrt{\alpha} < r$ and by Eq. (A4), $|D(re^{i\theta};\alpha)| > 0$, again, since on the contrary $D(s_0;\alpha) = 0$, $|s_0| < 2\sqrt{\alpha}$.

This upper bound can be refined by using the logarithmic form for the exponential integral, Eq. (A2), and we find that

$$0.5\alpha \ln 1/\alpha < |s_0| < 1.5\alpha \ln 1/\alpha.$$

With these bounds, we can use a Newton–Raphson-type method to iterate for approximations to s_0. Since D is continuous for $r > 0$ and $|\theta| < \pi$, the mean value theorem implies

$$D(s_0;\alpha) = D(s;\alpha) + D_s(s^+;\alpha)(s_0 - s), \tag{A5}$$

where $s^+ = s + \delta(s_0 - s)$ and $0 < \delta < 1$. The derivative

$$D_s(s;\alpha) = 1 - \alpha/s + \alpha e^{s}E_1(s)$$

is bounded uniformly away from zero by

$$|D_s(s;\alpha)| \geq \left|1 - \left|\frac{\alpha}{s} - \alpha e^{s}E_1(s)\right|\right|$$

with $|\alpha/s - \alpha e^{s}E_1(s)| \leq 0.14$ for $0.5\alpha \ln(1/\alpha) < r < 1.5\alpha \ln(1/\alpha)$ and $\alpha < \alpha_1$. Therefore, $D_s(s;\alpha) \geq 0.86$ and we write Eq. (A5) as

$$s_0 - s = -D(s;\alpha)/D_s(s^*;\alpha). \tag{A6}$$

As a one-term approximation, we try $s = e^{i\pi^-}\alpha \ln 1/\alpha$ in Eq. (A6) so that $|s_0 + \alpha L_1| < 1.7\,\alpha L_2$, where $L_1 = \ln 1/\alpha$, $L_2 = \ln\ln 1/\alpha$ and $s_0 \sim -\alpha L_1$ as $\alpha \to 0^+$. To obtain bounds on $\mathrm{Im}[s_0]$ we try a three-term approximation, $s = -\alpha L_1 + \alpha L_2 + \alpha(\gamma + i\pi)$ in Eq. (A6). We find $|s_0 - s| \leq 4.5\alpha L_2/L_1$ so that

$$s_0 \sim -\alpha L_1 + \alpha L_2 + \alpha(\gamma + i\pi) \quad \text{as } \alpha \to 0^+, \tag{A7}$$

and $\mathrm{Re}[s_0] \sim -\alpha L_1 + \alpha L_2 + \alpha\gamma$ and $\mathrm{Im}[s_0] \sim \alpha\pi$. The conjugate zero behaves as $s_0^* \sim -\alpha L_1 + \alpha L_2 + \alpha(\gamma - i\pi)$ as $\alpha \to 0^+$.

APPENDIX B. PROOF THAT THE INVERSE IS A SOLUTION

Even though we know the solution has a Laplace transform we must justify that the inverse is an exact solution by substitution.

Theorem B1: The inverse, Eq. (3.4), is a solution of Eq. (2.1) for sufficiently small α and $\mathrm{Re}[s_0] > 0$ and $t \leq T < \infty$.

Proof: Before substitution into Eq. (2.1), the integrand is manipulated into parts which have uniformly convergent integrals. Since $D(s;\alpha) \sim s^{-1}$ as $|s| \to \infty$ and is continuous otherwise if $\mathrm{Re}[s] > 0$, the integral converges conditionally, but not absolutely. We subtract the leading term for large $|s|$ and obtain

$$f_I(t;\alpha) = 1 - \frac{\alpha}{2\pi i}\int_{a-i\infty}^{a+i\infty} ds e^{ts} e^{s}E_1(s)/(sD(s;\alpha)), \tag{B1}$$

where we used the fact that $\int_{a-i\infty}^{a+i\infty} ds e^{ts}/s = 2\pi i$. For large $|s|$, $|e^{ts}e^{s}E_1(s)/(s(s + \alpha e^{s}E_1(s))| \sim e^{at}/|s|^3$ so that the integral in Eq. (B1) converges uniformly for $0 \leq t \leq T$ for some $T > 0$ and that $f_I(0;\alpha) = 1$ by Cauchy's theorem. Consequently, differentiation and integration may be performed under the integral sign

$$\dot{f}_I(t;\alpha) = \frac{\alpha}{2\pi i}\int_{a-i\infty}^{a+i\infty} ds e^{(t+1)s}E_1(s)/D(s;\alpha)$$

and

$$\alpha\int_0^t d\tau K(t-\tau)f_I(\tau;\alpha) = \alpha \ln(1+\tau)\frac{\alpha^2}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty} \times ds e^{s}E_1(s)e^{(t+1)s}\left(\frac{E_1(s) - E_1(s(1+t))}{sD(s;\alpha)}\right).$$

The contour integral on the right of the last equation can be simplified since

$$|e^{s}E_1(s)e^{s(t+1)}E_1(s(t+1))/(sD(s;\alpha))| \sim |1/((1+t)|s|^4)$$

as $|s| \to \infty$ so that its integral vanishes upon application of Cauchy's theorem. Also, $2\pi i \ln(1+t) = \int ds e^{(t+1)s}E_1(s)/s$ so that we have

$$\dot{f}_I = -\alpha K^* f_I. \qquad \text{QED}$$

APPENDIX C: EQUIVALENT CONTOUR FOR THE INVERSE

Although the inverse transform has been shown to be a solution in Appendix B, its form is not suitable for evaluation. A simpler, equivalent contour is sought which utilizes the poles and branch cut of $[D(s;\alpha)]^{-1}$.

We write the inverse Eq. (3.4) as a Cauchy principal value

$$f(t;\alpha) = \frac{1}{2\pi i}\lim_{R\to\infty}\int_{a-iR}^{a+iR} ds e^{st}/D(s;\alpha).$$

The integrand $e^{st}/D(s;\alpha)$ is continuous except for a logarithmic branch at $s = 0$ and two poles at the two zeros of $D(s;\alpha)$. The branch line is placed along the negative real axis. We define the following positively oriented arcs:

$$\Gamma_0 = \{s = a - iy: -R \leq y \leq R\},$$

$$\Gamma_1 = \{s = Re^{i\theta}: (\pi - \arctan(\delta/\sqrt{R^2-\delta^2})) \geq \theta \geq \arctan(\sqrt{R^2-a^2}/a)\},$$

$$\Gamma_2 = \{s = x + i\delta: -\sqrt{R^2-\delta^2} \leq x \leq \sqrt{\epsilon^2-\delta^2}\},$$

$$\Gamma_3 = \{s = \epsilon e^{i\theta}: (-\pi + \arctan(\delta/\sqrt{\epsilon^2-\delta^2})) \leq \theta \leq (\pi - \arctan(\delta/\sqrt{\epsilon^2-\delta^2}))\},$$

$$\Gamma_4 = \{s = x + i\delta: -\sqrt{R^2-\delta^2} \leq x \leq -\sqrt{\epsilon^2-\delta^2}\},$$

$$\Gamma_5 = \{s = Re^{i\theta}: (-\pi + \arctan(\delta/\sqrt{R^2-\delta^2})) \leq \theta \leq -\arctan(\sqrt{R^2-a^2}/a)\},$$

where $0 < \delta < \epsilon < (\alpha/e) < \mathrm{Im}[s_0]$, $\epsilon < a$ and $R > 1$. (See Fig. 2.)

According to Cauchy's residue theorem,

$$\frac{1}{2\pi i}\int_{\Gamma_0} ds e^{st}/D(s;\alpha) = \sum \mathrm{Res}[e^{st}/D] - \sum_{m=1}^{5}\frac{1}{2\pi i}\int_{\Gamma'_m} ds e^{st}/D(s;\alpha).$$

Using L'Hopital's rule, the residues at the two isolated poles, s_0 and s_0^*, are

$$\mathrm{Res}[e^{st}/D]|_{s=s_0} = \beta_0 e^{s_0 t}$$

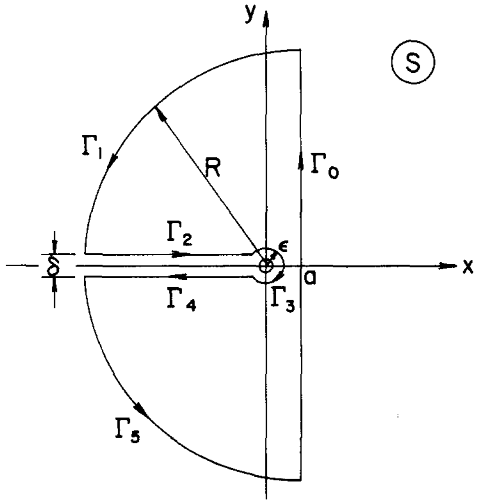

FIG. 2

and

$$\mathrm{Res}[e^{st}/D]|_{s=s_0^*} = \beta_0^* e^{s_0 t},$$

where $\beta_0 = 1/D_s(s_0; \alpha) = 1/(1 - \alpha/s_0 - s_0)$. The contribution of the contours Γ_1 and Γ_5 vanish as $R \to \alpha$ since $|D(s;\alpha)| \sim R$ as $R \to \alpha$. Also, the contribution of the contour Γ_3 is vanishing since $|D(s;\alpha)| \sim \alpha \ln 1/\epsilon$ as $\epsilon \to 0^+$.

On the contour adjacent to the branch cut, Γ_2 and Γ_3,

$$\mathrm{Im}[s - s_0] > 0.7\,\alpha\pi \quad \text{for } \alpha < \alpha_1,$$

where estimates of s_0 and the values of α_1 are given in Appendix A so that D is continuous and nonvanishing on Γ_2 and Γ_3. Hence the limit $\delta \to 0^+$ and the integral may be interchanged:

$$\lim_{\delta\to 0^+} \int_{\Gamma_2} ds e^{st}/D(\delta;\alpha) = -\int_\epsilon^R dx e^{-xt}/[x + \alpha e^{-x}(\ln x + i\pi + \gamma + e_2(x))]$$

where

$$e_2(x) \equiv \int_0^x dy (e^y - 1)/y$$

and the corresponding result for Γ_3 is minus the complex conjugate.

Finally, taking the limit as $R \to \infty$ and $\epsilon \to 0^+$, the solution takes the form

$$f(t;\alpha) = 2\,\mathrm{Re}[\beta_0 e^{s_0 t}] + I(t;\alpha), \tag{C1}$$

where

$$I(t;\alpha) = -\alpha \int_0^\infty dx e^{-xt} k_1(x;\alpha), \tag{C2}$$

$$K_1(x;\alpha) = e^{-x}/G_1(x;\alpha), \tag{C3}$$

and

$$G_1(x;\alpha) = (x + \alpha e^{-x}(\ln x + \gamma + e_2(x))^2 + \alpha^2\pi^2 e^{-2x}.$$

In Eq. (C1) the first term on the right is the contribution of the residues at s_0 and s_0^* and the second is the contribution of the branch cut.

*Work sponsored by the National Science Foundation, under Grant No. NSF GP 33368X.
†Work supported by the National Aeronautics and Space Administration under Contract No. NAS5-23020.

[1]A. Klimas and G. Sandri, Astrophys. J. **169**, 41 (1971).
[2]T. B. Kaiser, F. C. Jones, and T. J. Birmingham, NASA Preprint X-641-72-232 (1972). To appear in Astrophys. J.
[3]R. Bellman and K. L. Cooke, *Differential-Difference Equations* (Academic, New York, 1963).
[4]G. V. Ramanathan and G. Sandri, J. Math. Phys. **10**, 1763 (1969).
[5]D. V. Widder, *The Laplace Transform* (Princeton U.P., Princeton, N.J., 1946).
[6]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1964).
[7]R. V. Churchill, *Complex Variables and Applications* (McGraw-Hill, New York, 1960).

Deep inelastic scattering in a renormalized perturbative model

A. L. Mason

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England

(Received 16 November 1971)

We sum all the leading contributions to the form factor νW_2 in deep inelastic electron–proton scattering in a field theory with a neutral pseudoscalar gluon. This involves calculating ladder graphs with renormalization insertions. We find that we can express νW_2 in a simple way in terms of the asymptotic behavior of the renormalization parts. For ladders with only gluon rungs we obtain a result similar to Chang and Fishbane's for the Mellin transform of νW_2, namely $z(\ln\nu)(\exp\{(g^2/16\pi^2)[Y(\ln\nu)/(\lambda+1)(\lambda+2)]\}-1)$. Y and z are (calculable) functions depending on the renormalization parts. We also calculate the rather more complicated contribution from ladders with fermion rungs. Our results do not lend any support to the notion of anomalous dimensions.

1. INTRODUCTION

We investigate the deep inelastic electron–proton total cross section in a theory with a neutral pseudoscalar (or scalar) gluon. Chang and Fishbane[1] made the first extended investigation of this model, calculating the contribution of "rainbow" graphs by their infinite momentum technique. Gaisser and Polkinghorne[2] showed how to apply Feynman parametrization and Mellin transform techniques to calculate the asymptotic contribution of convergent graphs. Despite their conclusion that a cutoff model is a better description of the physical world —leading to Regge pole behavior and scale invariance in the Bjorken limit—it seems worthwhile to extend the investigation of the model without a cutoff to take account of divergent diagrams. We had some hope of discovering scaling with anomalous dimensions, as proposed by Wilson,[3] but this has not manifested itself.

We calculate the form factor νW_2 to leading order in $\ln\nu$ and the strong coupling constant g, calculating the contribution of every graph that is important in leading order. We find that such graphs are the dressed ladder graphs, that is, ladder graphs with renormalization parts inserted arbitrarily in them. We can sum these graphs in a fairly simple form in terms of the asymptotic behavior of renormalization parts. We find that the natural quantity to calculate is the Mellin transform of the form factor. For ladders with only gluon rungs we find

$$\mathfrak{M}_{1/\omega}(\nu W_2) \sim z\,(\ln\nu)\left[\exp\left(\frac{g^2}{16\pi^2}\int_0^{\ln\nu}\frac{y(x)\,dx}{(\lambda+1)(\lambda+2)}\right)-1\right] \tag{1.1}$$

The functions z and y depend upon the renormalization parts. If we set

$$y=1, \quad z(x)=\exp[(g^2/32\pi^2)x], \tag{1.2}$$

we reproduce the result Chang and Fishbane obtained by summing "rainbow graphs." Counting all leading renormalization parts, we arrive at

$$y(x)=[1-\tfrac{5}{16}(g^2/\pi^2)x]^{-1}, \quad z(x)=[1-\tfrac{5}{16}(g^2/\pi^2)x]^{1/10}, \tag{1.3}$$

reminiscent of results obtained by several authors[4-6] for the asymptotic behavior of renormalization parts.

These expressions for y and z are quite unsatisfactory, of course. Not only do they suffer from ghost cuts at large x, but also they do not give scaling. We can only hope that while our leading order calculation gives the wrong values for y and z yet the general form of (1.1) is correct. We discuss this further in Sec. 8.

In Sec. 2 we give a simple prescription for obtaining the contribution to νW_2 of dressed graphs with only gluon rungs. This relies upon replacing renormalization parts by their asymptotic values. In Sec. 3 we give a similar prescription for calculating the asymptotic behavior of the renormalization parts. This is similar in principle to the method used by Appelquist and Primack[6] for a charge symmetric model.

In Sec. 4 we calculate some simple examples by more rigorous methods and demonstrate that the results agree with those previously obtained. We also show that a diagram with fermion rungs can make an asymptotically important contribution. Sections 5 and 6 are devoted to justifying the prescription of Secs. 2 and 3, respectively.

In Sec. 7 we show how to calculate the leading contribution from the more general class of diagrams with fermion rungs. With renormalization parts as calculated in Sec. 3 the amplitude takes a fairly simple form.

In Sec. 8 we discuss briefly the validity of our results in other than leading order. We also show how formula (1.1) compares with experiment when we assume that the functions y and z behave in such a way as to allow scaling.

2. SUM OF DRESSED LADDER GRAPHS

In this section we present a prescription for calculating the leading behavior of the form factor νW_2. This prescription corresponds to the most naive way of doing the calculation, but turns out to give the same answer as the more rigorous method that will be described in Sec. 5.

We obtain the form factor from the imaginary part of the forward elastic $\gamma - p$ scattering amplitude. This is described in general by the diagrams of Fig. 1, but the diagrams we consider first are the dressed ladder diagrams of Fig. 2. We regard the blobs at the vertices of Fig. 2 as sums over all OPI (one particle irreducible) vertex parts, and the blobs on the propagators as sums over all (not necessarily OPI) self-energy diagrams. Then we need only sum over all possible number, n, of rungs, and we have counted all the diagrams of this class.

Suppose that the renormalized propagators have the following asymptotic forms (we neglect terms proportional to the rest masses):

Copyright © 1973 by the American Institute of Physics

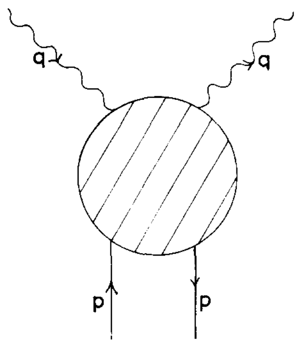

FIG. 1. Diagrams for forward elastic scattering.

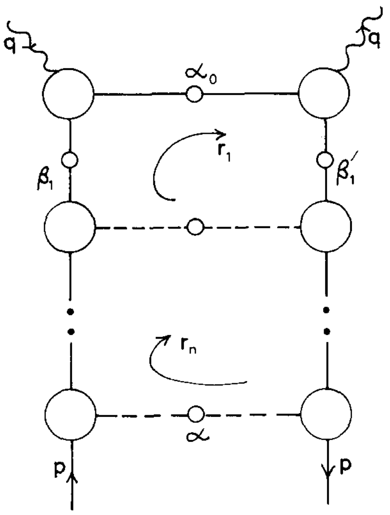

FIG. 2. Dressed ladder graph with gluon rungs only.

$$\tilde{S}_F'(Q) \sim \frac{i}{(2\pi)^4} \frac{\not{Q}}{Q^2 - m^2} \sum_{n=0}^{\infty} S_n(\ln(-Q^2))^n$$
$$= \frac{i}{(2\pi)^4} \frac{\not{Q}}{Q^2 - m^2} S(x) \quad (2.1)$$

$$\tilde{\Delta}_F'(Q) \sim \frac{i}{(2\pi)^4} \frac{1}{Q^2 - m^2} \sum_{n=0}^{\infty} T_n(\ln(-Q^2))^n$$
$$= \frac{i}{(2\pi)^4} \frac{1}{Q^2 - m^2} T(x), \quad (2.2)$$

where

$$x = \ln(-Q^2). \quad (2.3)$$

Further suppose that the photon vertex has the form

$$\Gamma_\mu \sim \frac{(2\pi)^4}{i} \gamma_\mu \sum_0^\infty u_n (\ln t)^n = \frac{(2\pi)^4}{i} \gamma_\mu U(y) \quad (2.4)$$

and the gluon vertex

$$\Gamma_5 \sim \frac{(2\pi)^4}{i} g\gamma_5 \sum_0^\infty v_n (\ln t)^n = \frac{(2\pi)^4}{i} g\gamma_5 V(y), \quad (2.5)$$

where

$$y = \ln t \quad (2.6)$$

and t is the invariant we chose to take asymptotic.

To calculate the behavior of a ladder graph, we replace each renormalization part by its asymptotic form. This is not a well-defined procedure for vertex parts since we may choose to take any of several variables as the asymptotic one. However, a prescription that will in Sec. 5 be shown to give the right answer is to take as asymptotic variable minus the squared momentum on the rung running into the vertex. This means that we can absorb the vertex renormalizing factors into the rung propagators, so that a dressed ladder graph looks just like a bare ladder graph with modified propagators. For the top (fermion rung) the effective propagator becomes

$$\frac{i}{(2\pi)^4} U^2(x) S(x) \frac{\not{Q}}{Q^2 - m^2} = \frac{i}{(2\pi)^4} \frac{\not{Q}}{Q^2 - m^2} z(x). \quad (2.7)$$

If we refer to the answer, (2.38), for the Mellin transform of νW_2, we see that we could equally well take $-q^2$ as the asymptotic variable for the electromagnetic vertices, since the function $z(\ln\nu)$ just factors out and $\ln\nu$ is asymptotically equal to $\ln(-q^2)$. As Appelquist and Primack[6] showed, the asymptotic form of the vertex is independent of which variable is taken asymptotic.

For the gluon rungs the effective propagator is

$$\frac{i}{(2\pi)^4} V^2(x) T(x) \frac{1}{Q^2 - m^2} = \frac{i}{(2\pi)^4} \frac{1}{Q^2 - m^2} t(x), \quad (2.8)$$

and for the (fermion) uprights it remains

$$\frac{i}{(2\pi)^4} \frac{\not{Q}}{Q^2 - m^2} S(x). \quad (2.9)$$

For a single term in any of these power series in $\ln(-Q^2)$ we can use the following parametrization

$$\frac{\ln^n(-Q^2)}{Q^2 - m^2} \sim (-1)^{n+1} \int d\rho \, \ln^n\rho \, \exp[\rho(Q^2 - m^2)]. \quad (2.10)$$

So for the complete propagator we have, say,

$$\frac{t(x)}{Q^2 - m^2} \sim \int d\rho \sum_0^\infty t_n (-1)^{n+1} \ln^n\rho \, \exp[\rho(Q^2 - m^2)], \quad (2.11)$$

and we may write this formally as

$$\frac{t(x)}{Q^2 - m^2} \sim - \int d\rho \, t(-\ln\rho) \exp[\rho(Q^2 - m^2)]. \quad (2.12)$$

With this parametrization we can write the dressed ladder contribution to γ-p scattering amplitude $T_{\mu\nu}$ in a form very closely resembling GP's expression for the contribution from the bare ladder. In fact, for the dressed n-rung ladder we write

$$T_{\mu\nu}^{(n)} = \prod_{i=1}^{n} \left(- \int_0^\infty t(-\ln\alpha_i) d\alpha_i \int_0^\infty S(-\ln\beta_i) d\beta_i \int_0^\infty \right.$$
$$\left. \times S(-\ln\beta_i') d\beta_i' ig^2 \int \frac{d^4 r_i}{(2\pi)^4} \right) \int_0^\infty z(-\ln\alpha_0) N_{\mu\nu}^{(n)}$$
$$\times \mathfrak{D}^{-1}(r_i, p, q) d\alpha_0 \quad (2.13)$$

while for the bare ladder we have GP's formula

$$T_{\mu\nu}^{(n)} = \prod_{i=1}^{n} \left(\mp \int_0^\infty d\alpha_i \int_0^\infty d\beta_i \int_0^\infty d\beta_i' \cdot ig^2 \int \frac{d^4 r_i}{(2\pi)^4} \right)$$
$$\times \int_0^\infty d\alpha_0 \, N_{\mu\nu}^{(n)} \mathfrak{D}^{-1}(r_i, p, q). \quad (2.14)$$

$N_{\mu\nu}^{(n)}$ the product of numerator factors, and $\mathfrak{D}^{-1}(r_i, p, q)$ the product of denominator factors from the propagators, are the same in both cases.

We now follow exactly the procedure of GP in choosing the dominant numerator factors proportional to q,

$$\Pi\,(-\tfrac{1}{2}r_i^2)(1/m)(p_\mu q_\nu + p_\nu q_\mu - g_{\mu\nu}\,p\cdot q), \tag{2.15}$$

and performing the symmetric integration. Defining the structure functions by

$$W_{\mu\nu} = \frac{Im}{\pi}\ T_{\mu\nu} = \frac{1}{m^2}\left(p_\mu - \frac{p\cdot q}{q^2}\,q_\mu\right)\left(p_\nu - \frac{p\cdot q}{q^2}\,q_\nu\right) \times W_2(q^2,\nu) - \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right) W_1(q^2,\nu), \tag{2.16}$$

where

$$\nu = (p\cdot q)/m,$$

we obtain

$$W_2^{(n)} = \frac{2m}{\omega}\left(\frac{g^2}{16\pi^2}\right)^n \frac{Im}{\pi} \prod_{i=1}^n \left(\int_0^\infty t(-\ln\alpha_i)d\alpha_i \int_0^\infty S(-\ln\beta_i)d\beta_i \int_0^\infty S(-\ln\beta_i')d\beta_i'\right) \int_0^\infty z\,[\ln(-\alpha_0)]\,d\alpha_0 \times \frac{1}{C^3}\exp\frac{D}{C} \tag{2.17}$$

for the leading contribution to W_2 from the n-rung dressed ladder. D and C are the standard parametric functions[7] for the corresponding bare ladder. The corresponding contribution from the bare ladder, calculated by GP, which we shall call B_n, is obtained by setting the functions t, S, and z equal to unity in (2.17).

We continue to use the methods of GP in investigating the asymptotic behavior of $W_2^{(n)}$. We have

$$D/C = 2m(g/C)\,\nu + d/C, \tag{2.18}$$

where

$$g = \alpha_0\left(\prod_{i=1}^n \alpha_i - \frac{1}{\omega}\,[C - \alpha_0 C']\right) \tag{2.19}$$

with

$$\omega = -2m\nu/q^2. \tag{2.20}$$

We scale according to

$$\delta_n \equiv \sigma_n \bar\delta_n,$$

$$\delta_{n-1} \equiv \sigma_n\sigma_{n-1}\bar\delta_{n-1},$$

$$\delta_1 \equiv \prod_{i=1}^n \sigma_i \bar\delta_1, \qquad \alpha_0 \equiv \prod_{i=0}^n \sigma_i\,, \tag{2.21}$$

where the set $\{\alpha_i, \beta_i, \beta_i'\}$ is represented by δ_i. Then we see that C scales as

$$C = \prod_{k=1}^n (\sigma_k)^k\,\bar C \tag{2.22}$$

so that g/C scales as

$$\frac{g}{C} = \prod_{k=1}^n \sigma_k\,\frac{\bar g}{\bar C} \tag{2.23}$$

The scaling changes terms like $\ln^p\alpha_j$ to $\ln^p(\bar\alpha_j \Pi_j^n \sigma_k)$. As we shall see, every power of a logarithm of a scaling factor leads to an extra logarithmic enhancement of the structure function. So the most important part of $\ln^p(\bar\alpha_j\,\Pi_j^n\,\sigma_k)$ is $\ln^p(\Pi_j^n\sigma_k)$. We obtain

$$W_2^{(n)} \sim \frac{2m}{\omega}\left(\frac{g^2}{16\pi^2}\right)^n \frac{Im}{\pi}\int \prod_{k=0}^n d\sigma_k \cdot z\left(-\ln\prod_{j=0}^n \sigma_j\right) \times \prod_{i=1}^n \left(y\Big(-\ln\prod_{j=i}^n \sigma_j\Big)\int d\bar\alpha_i d\bar\beta_i d\bar\beta_i'\,\delta(\bar\alpha_i + \bar\beta_i + \bar\beta_i' + \sigma_{i-1} - 1)\right)\frac{1}{\bar C^3}\exp\left(\frac{\bar g}{\bar C}\,\nu\prod_0^n \sigma_i\right)\exp\frac{\bar d}{\bar C}\,\sigma_n\,, \tag{2.24}$$

where

$$y = S^2 t. \tag{2.25}$$

Since y and z represent power series, Eq. (2.24) has expressed $W_2^{(n)}$ in a form to which we can apply the work of Appendix A. We see at once that we can write

$$W_2^{(n)} \sim \frac{1}{\omega}\left(\frac{g^2}{16\pi^2}\right)^n \frac{Im}{\pi}\cdot\frac{1}{\nu}\ln^n\nu \sum_{n_0,n_1\cdots n_n} \times\left\{\frac{(z_{n_0}\ln^{n_0}\nu)(y_{n_1}\ln^{n_1}\nu)\cdots(y_{n_n}\ln^{n_n}\nu)}{(n_n+1)(n_n+n_{n-1}+2)\cdots(n_n+\cdots+n_1+n)}\right\} \times f_n(\sigma,\beta)\big|_{\sigma=0,\beta=1}\,, \tag{2.26}$$

where, in fact,

$$\frac{Im}{\pi}\,f_n(0,-1) = \prod_{j=1}^n \left[\int d\bar\alpha_j d\bar\beta_j d\bar\beta_j'\,\delta(\bar\alpha_j + \bar\beta_j + \bar\beta_j' - 1)\right] \times\delta\left(\frac{\bar g}{\bar C}\right). \tag{2.27}$$

Formally we can rewrite (2.26) as

$$\nu W_2^{(n)} \sim \frac{z(\ln\nu)}{\omega}\prod_{i=1}^n\left(\int_0^{x_{i+1}} dx_i\,y(x_i)\cdot\frac{g^2}{16\pi^2} \times\int d\bar\alpha_i d\bar\beta_i d\bar\beta_i'\,\delta(\bar\alpha_i + \bar\beta_i + \bar\beta_i' - 1)\right)\frac{Im}{\pi}\left(\Pi\bar\alpha_i - \frac{1}{\omega}\right)^{-1}, \tag{2.28}$$

and the contribution from the bare ladder

$$\nu B_n \sim \frac{1}{\omega}\prod_{i=1}^n\left(\int_0^{x_{i+1}} dx_i\,\frac{g^2}{16\pi^2} \times\int d\bar\alpha_i d\bar\beta_i d\bar\beta_i'\,\delta(\bar\alpha_i + \bar\beta_i + \bar\beta_i' - 1)\right)\frac{Im}{\pi}\left(\Pi\bar\alpha_i - \frac{1}{\omega}\right)^{-1}, \tag{2.29}$$

where we interpret x_{n+1} as $\ln\nu$.

Define the Mellin transform of $g(x)$ by

$$\mathfrak{M}_x(g) = \int_0^1 x^{\lambda-1} g(x)\,dx \tag{2.30}$$

following Refs. 1 and 2. Transforming (2.28) with respect to $1/\omega$, the parametric integrals factorize and we can write

$$\mathfrak{M}_{1/\omega}(\omega\nu W_2^{(n)}) = z\cdot\int_0^{\ln\nu} dx_n\,\kappa y(x_n)\cdots\int_0^{x_2}\kappa y(x_1)\,dx, \tag{2.31}$$

where

$$\kappa = \frac{g^2}{16\pi^2}\ \frac{1}{\lambda(\lambda+1)}. \tag{2.32}$$

Finally we sum over n. Writing

$$\mathfrak{M}_{1/\omega}(\omega\nu W_2^{(n)}) = X_n \tag{2.33}$$

and X for the sum, and regarding them as functions of $\ln\nu$, λ being held fixed, we have

$$\frac{X}{z} = \int_0^x \frac{dx'}{z}\,\kappa y(X+1) \tag{2.34}$$

since

$$\int_0^x dx_{n+1}\,\kappa y(x_{n+1})\,\frac{X_n(x_{n+1})}{z}=\frac{X_{n+1}}{z}. \tag{2.35}$$

Equation (2.34) can be solved immediately to yield

$$X = z\left[\exp\left(\int_0^x dx'\,\kappa y(x')\right)-1\right] \tag{2.36}$$

So, writing

$$Y=\int_0^x y(x')\,dx', \tag{2.37}$$

we obtain

$$\mathfrak{M}_{1/\omega}(\omega\nu W_2)\sim z(\ln\nu)\left[\exp\left(\frac{g^2}{16\pi^2}\,\frac{Y(\ln\nu)}{\lambda(\lambda+1)}\right)-1\right] \tag{2.38}$$

for the contribution to νW_2 of the ladder graphs with gluon rungs only. All the difference between the bare ladders and the dressed ladders is summed up in these two functions z and Y.

3. ASYMPTOTIC BEHAVIOR OF RENORMALIZATION PARTS

We use methods similar to those of the previous section to calculate the asymptotic behavior of renormalization parts. The renormalization parts that are important in our dressed ladder graphs are those that can be drawn in the forms of Fig. 3. Other irreducible vertex parts, and box diagrams, lead to lower than leading behavior and need not be calculated. This is because, as we shall see, every divergent loop integral leads to a logarithmic enhancement in the asymptotic behavior, and only the diagrams of Fig. 3 have the maximum number of such loops.

The principle of the calculation is to replace every subdiagram by its asymptotic form and to sum over all possible subdiagrams. Graphs with overlapping divergences will be counted more than once, according to the number of ways they can be partitioned into skeletons. For example, the graph of Fig. 4 will be counted twice, according to the partitions shown in Fig. 5.

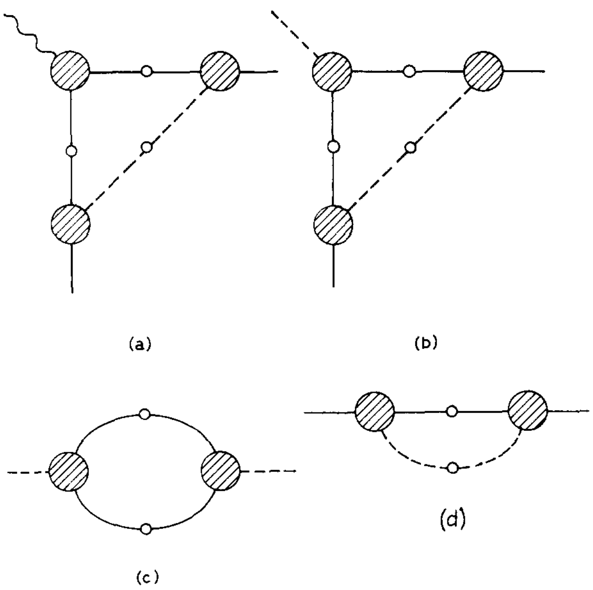

FIG. 3. Important renormalization parts.

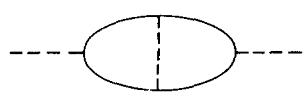

FIG. 4. Diagram with overlapping divergences.

A natural interpretation of this prescription would be that for such graphs contributions to the leading behavior come from several distinct regions of integration, and a particular partition focuses attention upon a particular region.

The above procedure defines inductively the asymptotic behavior of every graph. Working in terms of the sums to all orders, we obtain simple integral equations for the leading contributions of all the graphs of Fig. 3.

We use again expressions (2.1), (2.2) and (2.4), (2.5) for the renormalized propagators and vertices. We also require expressions for the sums of all OPI self-energy graphs. For the fermion we write

$$\Sigma(Q)\sim\frac{(2\pi)^4}{i}\,\not{Q}\sum_1^\infty f_n\,\ln^n(-Q^2) =\frac{(2\pi)^4}{i}\,\not{Q}\,F(x) \tag{3.1}$$

and for the gluon

$$K(Q)\sim\frac{(2\pi)^4}{i}\,Q^2\sum_1^\infty g_n\,\ln^n(-Q^2) =\frac{(2\pi)^4}{i}\,Q^2 G(x). \tag{3.2}$$

We obtain the propagators by summing over all possible OPI insertions, and so we have relations between F and S, G and T:

$$S=[1-F]^{-1}, \tag{3.3}$$

$$T=[1-G]^{-1}. \tag{3.4}$$

Let us now calculate the asymptotic behavior of the photon vertex of Fig. 6. Once again the contribution from the vertices is not well defined. We must take as asymptotic variable minus the square momentum of one of the lines in the loop. It will be shown to make no difference which one. Naively the Feynman integral we obtain is

$$\begin{aligned}\int d^4r\,&\frac{g^2\gamma_5(\not{p}+\not{r})\gamma_\mu(\not{p}+\not{q}+\not{r})\gamma_5}{[(p+r)^2-m^2][(p+q+r)^2-m^2][r^2-\lambda^2]}\\ &\times V(\ln[-(p+r)^2])S(\ln[-(p+r)^2])U(\ln[-(p+r)^2])\\ &\times S(\ln[-(p+q+r)^2])V(\ln(-r^2))\,T(\ln(-r^2)),\end{aligned} \tag{3.5}$$

and this represents the sum of all important vertex parts except for the bare vertex. So expression (3.5) is asymptotically equal to

$$[(2\pi)^4/i]\gamma_\mu[U(\ln t)-1]. \tag{3.6}$$

The term $\not{r}\gamma_\mu\not{r}$ in the numerator leads to a logarithmic divergence. This is the term that gives the leading asymptotic behavior. Since we are using pseudoscalar gluons, we get no enhancement from the infrared region, but we would not be interested in that region anyway when we insert the vertex in a ladder. Only the ultraviolet region leads to enhancements in the ladder diagrams. We pick out the $\not{r}\gamma_\mu\not{r}$ term, parametrize as in Sec. 2, and naively perform the symmetric integration. We obtain

$$\frac{g^2}{16\pi^2}\gamma_\mu \int_0^\infty \frac{dxdydz}{C^3}\,(VSU)(-\ln x)S(-\ln y)(VT)(-\ln z)e^{D/C}, \tag{3.7}$$

where D and C are the parametric functions for the simple bare vertex loop. We may consider the integral to be regularised according to Hepp's[8] prescription by cutting off the parametric integration at some small value. We renormalize by subtracting in the integrand the constant term in the Taylor expansion in powers of the momenta. The point we take the expansion about is unimportant. It corresponds to a finite renormalization which will introduce only an additive constant negligible compared with the leading logarithmic behavior. If we also scale over x, y, and z, we obtain

$$\begin{aligned}[U(\ln t)-1] \sim \frac{g^2}{16\pi^2}\int \frac{d\rho}{\rho}\,(V^2S^2TU)(-\ln\rho) \\ \times \int d\bar{x}d\bar{y}d\bar{z}\,\delta(\bar{x}+\bar{y}+\bar{z}-1)\,(\exp(\rho\bar{D}/\bar{C}) \\ -\exp\{-\rho[(\bar{x}+\bar{y})m^2+\bar{z}\lambda^2]\})\end{aligned} \tag{3.8}$$

We have once again discarded the terms in $\ln\bar{x}$ from $(\ln\rho\bar{x})^P$ etc. as we did before Eq. (2.24). Our integral is now convergent and we may integrate by parts with respect to ρ. $(UV^2S^2T)(-\ln\rho)$ represents a power series in $\ln\rho$, and so we may consider one term at a time. It is easy to see that

$$\int_0^\infty d\rho\,\frac{\ln^n\rho}{\rho}f(\rho)=\frac{1}{n+1}\int_0^\infty d\rho\,\ln^{n+1}\rho\left(-\frac{\partial}{\partial\rho}\right)f(\rho) \tag{3.9}$$

provided the surface term

$$[\ln^{n+1}\rho f(\rho)]_0^\infty \tag{3.10}$$

vanishes. Remarking that the transformation

$$\ln^n\rho \to \ln^{n+1}\rho/(n+1) \tag{3.11}$$

looks like an integration, we can now rewrite Eq. (3.8) as

$$\begin{aligned}U(\ln t)-1 \sim (g^2/16\pi^2)\int d\rho\,R_U(-\ln\rho)\,\frac{\partial}{\partial\rho} \\ \times \int d\bar{x}d\bar{y}d\bar{z}\,\delta(\bar{x}+\bar{y}+\bar{z}-1)\exp\rho\,\bar{D}/\bar{C},\end{aligned} \tag{3.12}$$

where

$$R_U(x)=\int_0^x (UV^2S^2T)(x')dx' \tag{3.13}$$

and we have dropped what remains of the counterterm since it gives only a term independent of the momenta. We may write

$$\bar{D}/\bar{C}=(\bar{g}/\bar{C})t+\bar{d}/\bar{C}. \tag{3.14}$$

t is the variable we wish to take asymptotic. Performing the differentiation in (3.12) and picking out the term t, we see that

$$\begin{aligned}U(\ln t)-1 \sim (g^2/16\pi^2)t\int d\rho\,R_U(-\ln\rho) \\ \times \int d\bar{x}d\bar{y}d\bar{z}\,\delta(\bar{x}+\bar{y}+\bar{z}-1)\bar{g}/\bar{C}\exp[\rho(\bar{g}/\bar{C})t] \\ \times \exp(\rho\,\bar{d}/\bar{C}).\end{aligned} \tag{3.15}$$

And from Appendix A this gives

$$U(\ln t)-1 \sim \frac{-g^2}{16\pi^2}R_U(\ln t)\int d\bar{x}d\bar{y}d\bar{z}\,\delta(\bar{x}+\bar{y}+\bar{z}-1). \tag{3.16}$$

The integral in (3.16) equals $\frac{1}{2}$, and so we have derived the following equation connecting the renormalization parts:

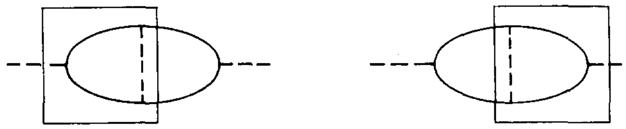

FIG. 5. Two distinct partitions of graph of Fig. 4.

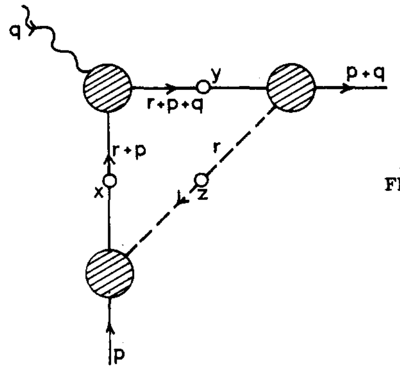

FIG. 6. General photon vertex.

$$U(x)-1=\frac{-g^2}{32\pi^2}\int_0^x U(\xi)V^2(\xi)S^2(\xi)T(\xi)d\xi. \tag{3.17}$$

The only way in which the gluon vertex of Fig. 3(b) differs from the photon vertex is that the important numerator factor is $\not{r}\gamma_5\not{r}$ instead of $\not{r}\gamma_\mu\not{r}$. This gives an extra factor of -2, so that

$$V(x)-1=\frac{g^2}{16\pi^2}\int_0^x V^3(\xi)S^2(\xi)T(\xi)d\xi. \tag{3.18}$$

The gluon self-energy part (Fig. 7) leads to the unrenormalized Feynman integral

$$\begin{aligned}-g^2\int d^4r\,\mathrm{Tr}[\gamma_5(\not{Q}+\not{r})\gamma_5\not{r}]\int dxdy(VS)(-\ln x) \\ \times (VS)(-\ln y)\exp[x(Q+r)^2+yr^2-(x+y)m^2].\end{aligned} \tag{3.19}$$

Both the quadratic and the logarithmic divergences give important contributions here.

Taking account of displacement terms, we have that the important part of the trace is

$$4\{r'^2-[xy/(x+y)^2]Q^2\}, \tag{3.20}$$

where r' is the symmetrized loop momentum. Performing the symmetric integration and renormalizing gives

$$\begin{aligned}[(2\pi)^4/i]Q^2G(\ln(-Q^2)) \\ \sim -4g^2 i\pi^2\int(d\rho/\rho)(V^2S^2)(-\ln\rho) \\ \times\int d\bar{x}d\bar{y}\,\delta(\bar{x}+\bar{y}-1)[-(2/\rho)(\exp\rho\bar{x}\bar{y}Q^2-1-\rho\bar{x}\bar{y}Q^2) \\ -\bar{x}\bar{y}Q^2(\exp\rho\bar{x}\bar{y}Q^2-1)].\end{aligned} \tag{3.21}$$

Integrating by parts reduces this to

$$\begin{aligned}[(2\pi)^4/i]G(\ln(-Q^2)) \sim 12g^2 i\pi^2\int d\rho\,R_G(-\ln\rho)\,\partial/\partial\rho \\ \times\int d\bar{x}d\bar{y}\bar{x}\bar{y}\,\delta(\bar{x}+\bar{y}-1)\exp(\rho\bar{x}\bar{y}Q^2-\rho m^2),\end{aligned} \tag{3.22}$$

where

$$R_G(x)=\int_0^x (V^2S^2)(x')dx', \tag{3.23}$$

whence

$$\begin{aligned}G(\ln(-Q^2)) \sim +12(g^2/16\pi^2)R_G(\ln(-Q^2)) \\ \int d\bar{x}d\bar{y}\bar{x}\bar{y}\,\delta(\bar{x}+\bar{y}-1).\end{aligned} \tag{3.24}$$

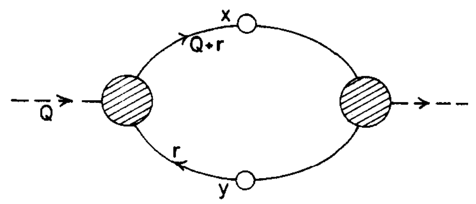

FIG. 7. Gluon self-energy part.

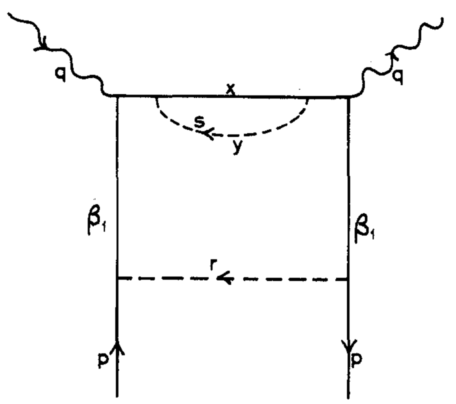

FIG. 8. Simple "rainbow" graph.

We can now use the relation (3.4) to rewrite this

$$1-\frac{1}{T(x)}=\frac{g^2}{8\pi^2}\int_0^x V^2(\xi)S^2(\xi)d\xi. \tag{3.25}$$

A similar procedure with the fermion self-energy part gives us

$$F(x)=1-\frac{1}{S(x)}=\frac{g}{32\pi^2}\int_0^x V^2(\xi)S(\xi)T(\xi)d\xi. \tag{3.26}$$

Equations (3.17), (3.18), (3.25), and (3.26) form a solvable set of integral equations for U, V, S, and T. By writing

$$y=V^2S^2T, \quad z=U^2S \tag{3.27}$$

as in Sec. 2, the equations give, upon differentiation,

$$\begin{aligned}U'/U&=-(g^2/32\pi^2)y, \quad V'/V=(g^2/16\pi^2)y,\\ S'/S&=(g^2/32\pi^2)y, \quad T'/T=(g^2/8\pi^2)y,\end{aligned} \tag{3.28}$$

and

$$y'/y=(5g^2/16\pi^2)y. \tag{3.29}$$

These are easily solved to give

$$\begin{aligned}U(x)&=[1-(5g^2/16\pi^2)x]^{1/10},\\ V(x)&=[1-(5g^2/16\pi^2)x]^{-1/5},\\ S(x)&=[1-(5g^2/16\pi^2)x]^{-1/10},\\ T(x)&=[1-(5g^2/16\pi^2)x]^{-4/10},\end{aligned} \tag{3.30}$$

and

$$z(x)=[1-(5g^2/16\pi^2)x]^{1/10} \tag{3.31}$$

$$Y(x)=\int_0^x y(x')dx'=-(16\pi^2/5g^2)\ln[1-5g^2/16\pi^2)x]. \tag{3.32}$$

Results resembling (3.30) have been obtained by a number of authors. Most recently Appelquist and Primack[6] obtained expressions for the form factors in a charge symmetric model by calculations similar in principle to ours. (They also quote results for the theory with only a neutral gluon. These are identical with equation (3.30)). Their results agreed with earlier work by Landau, Abrikosov, and Khalatnikov[4], who solved integral equations for the Green's function in a cutoff model. J. C. Taylor[5] also obtained the same results by renormalization group arguments. Solving functional equations connecting cutoff and renormalized form factors leads to expressions of the form

$$[1-ax]^{n_i} \tag{3.33}$$

for the functions V, S, and T. The quantities a and n_i are calculated from second-order perturbation expansions. We remark that our technique also reduces just to calculating second-order graphs.

4. SIMPLE EXAMPLES

It is not obvious at first sight that the important contributions to the scattering amplitude should come from the asymptotic regions for the renormalization part. In this section we demonstrate that it is true for some simple examples. We also demonstrate [example (D)] that diagrams containing fermion boxes are also important.

(A) We consider first the diagram of Fig. 8. Ignoring the term proportional to the fermion mass, the renormalized amplitude for the self-energy part can be written

$$\begin{aligned}&i\pi^2g^2\not{Q}\int_0^\infty d\rho\,\ln\rho\left(-\frac{\partial}{\partial\rho}\right)\int d\bar{x}d\bar{y}\,\delta(\bar{x}+\bar{y}-1)\\ &\quad\times[\exp\rho\bar{x}\bar{y}\,Q^2-1]\exp[-(\rho\bar{x}m^2+\rho\bar{y}\lambda^2)],\end{aligned} \tag{4.1}$$

where

$$Q=p+q+r, \tag{4.2}$$

following the procedure of the previous section. The second term in the square bracket leads only to a constant when integrated. We shall see that it can again be neglected. This throws the self-energy contribution into what we shall find is a characteristic form, namely,

$$i\pi^2g^2\not{Q}\int d\rho\,\ln\rho\left(-\frac{\partial}{\partial\rho}\right)\int d\xi f(\xi)\exp\rho\,\frac{\bar{D}}{\bar{C}}, \tag{4.3}$$

where ξ represents the internal variables of the renormalization part.

To evaluate the complete amplitude, we note that, neglecting the terms proportional to m, we can cancel (Q^2-m^2) from the denominator of one of the propagators in the top line, against Q^2, the product of numerator factors. The contribution to $T_{\mu\nu}$ becomes

$$\begin{aligned}&\left(-\frac{ig^2}{(2\pi)^4}\right)\int\frac{d^4r\bar{u}(p)\gamma_5(\not{r}+\not{p})\gamma_\mu(\not{r}+\not{p}+\not{q})\gamma_\nu(\not{r}+\not{p})\gamma_5u(p)}{[r^2-\lambda^2][(r+p)^2-m^2][(r+p+q)^2-m^2]}\\ &\quad\times\frac{g^2}{16\pi^2}\int d\rho(-\ln\rho)\left(-\frac{\partial}{\partial\rho}\right)\int d\bar{x}d\bar{y}\,\delta(\bar{x}+\bar{y}-1)\exp\frac{D}{C}.\end{aligned} \tag{4.4}$$

The most important contribution arises from contracting the $\not{p}$ from the wavefunction with the $\not{q}$ from the top line and taking the loop momenta from the other terms. That is, the important numerator term containing a factor of q is

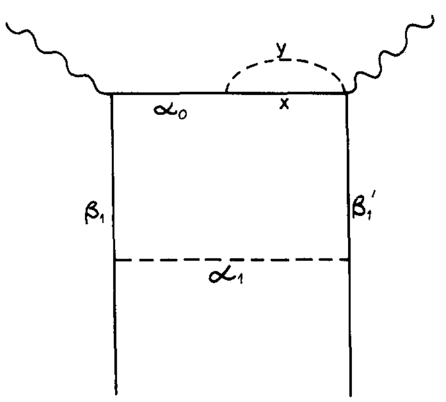

FIG. 9. Cancelled graph corresponding to Fig. 8.

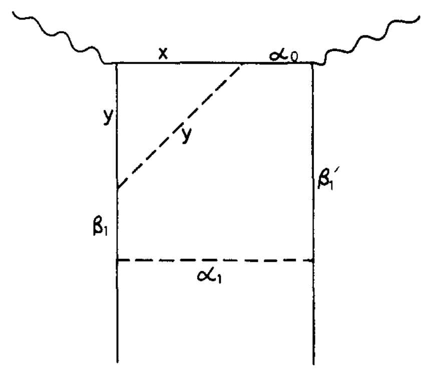

FIG. 10. Ladder graph with dressed vertex.

$$(1/4m)\ \mathrm{Tr}[\not p \not r \gamma_\mu \not q \gamma_\nu \not r] \sim (-\tfrac{1}{2}r^2)(1/m)(p_\mu q_\nu + p_\nu q_\mu - g_{\mu\nu}\, p\cdot q), \qquad (4.5)$$

neglecting displacement terms. Performing the loop integration, we find the contribution C to W_2

$$C \sim \frac{2m}{\omega}\left(\frac{g^2}{16\pi^2}\right)^2 \frac{Im}{\pi} \int d\alpha_0 d\alpha_1 d\beta_1 d\beta_1' \int d\rho(-\ln\rho)\left(-\frac{\partial}{\partial\rho}\right) \times \int d\bar{x} d\bar{y}\, \bar{y}\, \delta(\bar{x}+\bar{y}-1)\frac{\exp(F/E)}{(\alpha_0+\alpha_1+\beta_1+\beta_1'+\rho\bar{x}\bar{y})^3}. \qquad (4.6)$$

We write E and F for the parametric functions C and D for the "cancelled" graph (Fig. 9) corresponding to our original graph (Fig. 8). We have

$$F/E = 2m\nu\, g/E + d/E \qquad (4.7)$$

with

$$\frac{g}{E} = (\alpha_0 + \rho\bar{x}\bar{y})\frac{[\alpha_1 - (1/\omega)(\alpha_1+\beta_1+\beta_1')]}{(\alpha_0+\alpha_1+\beta_1+\beta_1'+\rho\bar{x}\bar{y})}. \qquad (4.8)$$

If we perform the scaling of (2.21) with the difference that

$$\alpha_0 = \sigma_1\sigma_0\bar{\alpha}_0, \qquad \rho = \sigma_1\sigma_0\bar{\rho}, \qquad (4.9)$$

we see that g/E scales as

$$g/E = \sigma_1\sigma_0\, \bar{g}/\bar{E} \qquad (4.10)$$

Keeping, as usual, only $\ln\sigma_1\sigma_0$ from the term $\ln\sigma_1\sigma_0\bar{\rho}$, we obtain

$$C \sim \frac{2m}{\omega}\left(\frac{g^2}{16\pi^2}\right)^2 \frac{Im}{\pi}\int d\sigma_1 d\sigma_0(-\ln\sigma_1\sigma_0)\, d\bar{\alpha}_0 d\bar{\rho}\,\delta(\bar{\alpha}_0+\bar{\rho}-1) \times \int d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}_1'\delta(\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}_1'+\sigma_0-1)\left(-\frac{\partial}{\partial\bar{\rho}}\right) \times \int d\bar{x}d\bar{y}\bar{y}\delta(\bar{x}+\bar{y}-1)\frac{\exp(2m\nu\sigma_0\sigma_1\bar{g}/E)\exp(\sigma_1\bar{d}/\bar{E})}{[\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}_1+\sigma_0(\bar{\alpha}_0+\bar{\rho}\bar{x}\bar{y})]^3} \qquad (4.11)$$

and the usual invocation of Appendix A gives us the asymptotic behavior

$$C \sim \frac{1}{\omega\nu}\left(\frac{g^2}{16\pi^2}\right)^2 \frac{Im}{\pi}\ln^2\nu \int d\bar{\alpha}_0 d\bar{\rho}\,\delta(\bar{\alpha}_0+\bar{\rho}-1) \times \int d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}_1'\delta(\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}_1'-1)\left(-\frac{\partial}{\partial\bar{\rho}}\right) \times \int d\bar{x}d\bar{y}\bar{y}\delta(\bar{x}+\bar{y}-1)\left(\frac{\bar{g}}{\bar{E}}\right)^{-1} \qquad (4.12)$$

At this stage we can see that no other terms in the amplitude could have given as big a contribution as the term we have chosen. Writing the imaginary part of $(\bar{g}/\bar{E})^{-1}$ as a delta function, we have

$$C \sim \frac{\ln^2\nu}{\omega\nu}\left(\frac{g^2}{16\pi^2}\right)^2 \int d\bar{x}d\bar{y}\bar{y}\delta(\bar{x}+\bar{y}-1)\int d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}_1'\delta(\bar{\alpha}_1 + \bar{\beta}_1+\bar{\beta}_1'-1)\left\{\int d\bar{\alpha}_0 d\bar{\rho}\,\delta(\bar{\alpha}_0+\bar{\rho}-1)\left(-\frac{\partial}{\partial\bar{\rho}}\right) \times \frac{1}{(\bar{\alpha}_0+\bar{\rho}\bar{x}\bar{y})}\right\}\delta\left(\bar{\alpha}_1-\frac{1}{\omega}\right), \qquad (4.13)$$

and by Appendix B we can remove all the ρ's from the last integral, whereupon it reduces to 1. So we have factorized the amplitude

$$\nu C \sim (g^2/16\pi^2)\int d\bar{x}d\bar{y}\bar{y}\delta(\bar{x}+\bar{y}-1)\cdot\ln\nu \times (1/\omega)(g^2/16\pi^2)\ln\nu\int d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}_1'\delta(\bar{\alpha}_1+\bar{\beta}_1 + \bar{\beta}_1'-1)\delta(\bar{\alpha}_1-1/\omega) \qquad (4.14)$$

into a product of the bare ladder contribution and the asymptotic behavior of the propagator, agreeing with the prescription of Sec. 2.

(B) A similar analysis may be made of the diagram of Fig. 10. The ultraviolet term in the vertex part, when renormalized can be written

$$-i\pi^2 g^2\gamma_\mu\int d\rho\,\ln\rho\left(-\frac{\partial}{\partial\rho}\right)\int d\bar{x}d\bar{y}d\bar{z}\delta \times(\bar{x}+\bar{y}+\bar{z}-1)\exp(\rho\bar{D}/\bar{C}). \qquad (4.15)$$

Plugging this into the rest of the graph and following the same procedure as before, we find that g/E scales in the same way, and we have

$$[\bar{g}/\bar{E}]^{-1}_{\sigma=0} = \{1/[\bar{\alpha}_0+\bar{\rho}(\bar{x}\bar{y}+\bar{x}\bar{z})]\}(\bar{\alpha}_1-1/\omega)^{-1}. \qquad (4.16)$$

So we can again write the contribution to νW_2 in the factorized form

$$\nu C \sim (g^2/16\pi^2)\int d\bar{x}d\bar{y}d\bar{z}\delta(\bar{x}+\bar{y}+\bar{z}-1)\ln\nu \times(1/\omega)(g^2/16\pi^2)\ln\nu\int d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}_1'\delta(\bar{\alpha}_1 + \bar{\beta}_1+\bar{\beta}_1'-1)\delta(\bar{\alpha}_1-1/\omega) \qquad (4.17)$$

(C) We now consider a diagram where the renormalization part is not on the top rung—Fig. 11. The gluon self-energy part gives the renormalized contribution

$$- 12r^2 i\pi^2 g^2 \int d\rho \ln\rho \left(-\frac{\partial}{\partial\rho}\right) \int d\bar{x}d\bar{y}\bar{x}\bar{y}\delta \times (\bar{x}+\bar{y}-1)\exp\rho\frac{\bar{D}}{\bar{C}}, \quad (4.18)$$

and we use the r^2 to cancel the denominator of one of the neighboring gluon propagators.

Then the corresponding contribution to W_2 is

$$C \sim \frac{2m}{\omega}\left(\frac{g^2}{16\pi^2}\right)\frac{Im}{\pi}\int d\alpha_0 d\alpha_1 d\beta_1 d\beta'_1 \times (-12)\left(\frac{g^2}{16\pi^2}\right)\int d\rho\,(-\ln\rho)\left(-\frac{\partial}{\partial\rho}\right) \times \int d\bar{x}d\bar{y}\bar{x}\bar{y}\delta(\bar{x}+\bar{y}-1)\frac{\exp(F/E)}{(\alpha_0+\alpha_1+\beta_1+\beta'_1+\rho\bar{x}\bar{y})^3} \quad (4.19)$$

where E and F are the parametric functions of the cancelled graph, Fig. 12. The coefficient of $2m\nu$ in F/E is given by

$$\frac{g}{E} = \frac{\alpha_0[\alpha_1+\rho\bar{x}\bar{y}-(1/\omega)(\alpha_1+\beta_1+\beta'_1+\rho\bar{x}\bar{y})]}{\alpha_0+\alpha_1+\beta_1+\beta'_1+\rho\bar{x}\bar{y}} \quad (4.20)$$

So if we make the scaling

$$\alpha_0 = \sigma_1\sigma_0, \qquad \delta = \sigma,\bar{\delta}, \quad (4.21)$$

where δ represents $\{\alpha_1, \beta_1, \beta'_1, \rho\}$,

g/E scales in the usual way, and we get

$$C \sim \frac{2m}{\omega}\left(\frac{g^2}{16\pi^2}\right)^2\frac{Im}{\pi}\int d\sigma_1 d\sigma_0(-\ln\sigma_1) \times \int d\bar{\rho}d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}'_1\delta(\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}'_1+\bar{\rho}+\sigma_0-1)\left(-\frac{\partial}{\partial\bar{\rho}}\right) \times \left(-12\int d\bar{x}d\bar{y}\bar{x}\bar{y}\delta(\bar{x}+\bar{y}-1) \times \frac{\exp(2m\nu\sigma_1\sigma_0\bar{g}/\bar{E}+\sigma_1\bar{d}/\bar{E})}{(\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}'_1+\bar{\rho}\bar{x}\bar{y}+\sigma_0)^3}\right) \quad (4.22)$$

From this we get, applying Appendix A and taking the imaginary part,

$$\nu C \sim \frac{1}{\omega}\left(\frac{g^2}{16\pi^2}\right)^2\frac{\ln^2\nu}{2}\int d\bar{\rho}d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}'_1\delta(\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}'_1+\bar{\rho}-1)\left(-\frac{\partial}{\partial\bar{\rho}}\right)(-12)\int d\bar{x}d\bar{y}\bar{x}\bar{y} \times \delta(\bar{x}+\bar{y}-1)\,\delta\left(\frac{\bar{\alpha}_1+\bar{\rho}\bar{x}\bar{y}}{\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}_1+\bar{\rho}\bar{x}\bar{y}}-\frac{1}{\omega}\right). \quad (4.23)$$

The argument of the δ function is homogeneous of degree zero so that we can again apply Appendix B to remove $\bar{\rho}$'s. The result can be written

$$\nu C \sim \frac{1}{\omega}\int d(\ln\nu)\left((-12)\frac{g^2}{16\pi^2}\int d\bar{x}d\bar{y}\bar{x}\bar{y}\delta(\bar{x}+\bar{y}-1)\right) \times \ln\nu\left(\frac{g^2}{16\pi^2}\right)\int d\bar{\alpha}_1 d\bar{\beta}_1 d\bar{\beta}'_1\delta(\bar{\alpha}_1+\bar{\beta}_1+\bar{\beta}'_1-1) \times \delta\left(\bar{\alpha}_1-\frac{1}{\omega}\right) \quad (4.24)$$

in agreement with the result of Sec. 2.

(D) Considering the Bjorken–Johnson–Low limit, Polkinghorne[9] found that diagrams like Fig. 13 gave

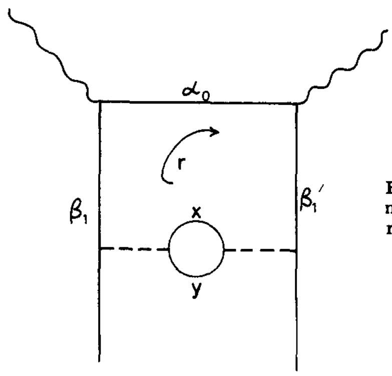

FIG. 11. Diagram with the renormalization part off the top rung.

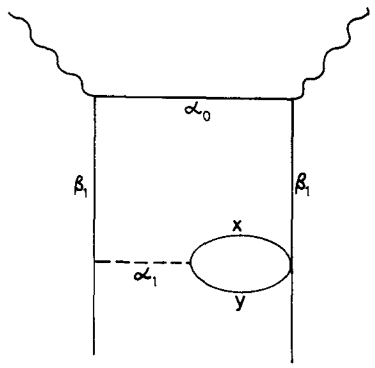

FIG. 12. Cancelled graph corresponding to Fig. 11.

contributions greater by a power of q_0 than ladder diagrams. In the deep inelastic limit we find that they give contributions of exactly the same order as ladder graphs. For we can write the integral in the form (2.14), with

$$N_{\mu\nu} \sim (-1)\,\mathrm{Tr}[(\not{r}_1-\not{r}_2)\not{r}_1\gamma_\mu(\not{r}_1+\not{q})\gamma_\nu\not{r}_1]\frac{p\cdot(p-r_2)}{m}. \quad (4.25)$$

We have for the displaced loop momenta r'_1 and r'_2

$$r_1 = r'_1 + \frac{\alpha_1}{\alpha_1+\beta_1+\beta'_1}\,r_2 - \frac{\alpha_0}{\alpha_1+\beta_1+\beta'_1}\,q + O(\sigma),$$

$$r_2 = r'_2 + \frac{\alpha_2}{\alpha_2+\beta_2+\beta'_2} \times p - \frac{\alpha_1\alpha_0}{(\alpha_1+\beta_1+\beta'_1)(\alpha_2+\beta_2+\beta'_2)}\,q + O(\sigma). \quad (4.26)$$

The term corresponding to the divergence of the box graph is proportional to $g_{\mu\nu}$ so does not contribute to νW_2 (and its imaginary part is zero anyway as we shall see in Sec. 7). Important terms in $N_{\mu\nu}$ quadratic in r'_1 and r'_2 are

$$\{r'^2_1[1+\alpha_1/(\alpha_1+\beta_1+\beta'_1)]4(r'_{2\mu}q_\nu+r'_{2\nu}q_\mu-g_{\mu\nu}q\cdot r'_2) - 2r'_1\cdot r'_2[1-\alpha_1/(\alpha_1+\beta_1+\beta'_1)] \times 4(r'_{1\mu}q_\nu+r'_{1\nu}q_\mu-g_{\mu\nu}q\cdot r'_1)\}\,r'_2\cdot p/m. \quad (4.27)$$

Symmetrizing, this becomes

$$(-\tfrac{1}{2}r'^2_1)(-\tfrac{1}{2}r'^2_2)[2+6\alpha_1/(\alpha_1+\beta_1+\beta'_1)](p_\mu q_\nu + p_\nu q_\mu - g_{\mu\nu}p\cdot q), \quad (4.28)$$

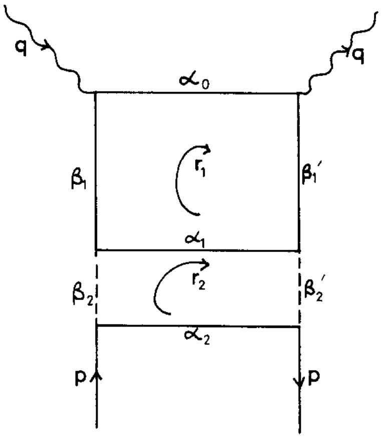

FIG. 13. Diagram containing fermion box.

and we see that this differs from the numerator for the ladder graph only by the factor $[2 + 6\alpha_1/(\alpha_1 + \beta_1 + \beta_1')]$, which does not affect the scaling. So there is at least one term in the graph of Fig. 13 which is as important as the corresponding ladder graph. In fact we shall see in Sec. 7 that we can find several other important terms as well.

5. JUSTIFICATION OF SEC. 2

As we shall prove in the next section, the leading contribution from a renormalization part takes the form

$$C_r R_r = C_r \frac{(2\pi)^4}{i} \int (-\ln\rho_r)^{n_r} d\rho_r \left(\frac{\partial}{\partial\rho_r}\right) \times \int d\xi\, f_r(\xi) \exp \rho_r \bar{F}_r/\bar{E}_r, \quad (5.1)$$

where r labels the particular renormalization part, F_r and E_r are the parametric functions for the corresponding cancelled graph, and C_r is $Q^2, \not{Q}, g\gamma_5$, or γ_μ depending on what sort of renormalization part it is.

Taking this form for the insertions, we can write down a convergent Feynman integral for any dressed ladder graph. If we cancel the C_r terms of self-energy parts against the denominators of neighboring propagators as in the examples of Sec. 4, we obtain the amplitude contributing to $T_{\mu\nu}$

$$C_{\mu\nu} \propto \prod_{i=1}^{n} \left(\int_0^\infty d\alpha_i d\beta_i d\beta_i' \int d^4 r_i\right) \int_0^\infty d\alpha_0 N_{\mu\nu}^{(n)} \times \mathfrak{D}^{-1}(r,p,q) \cdot \prod_r R_r. \quad (5.2)$$

As we see, this resembles (2.13) and (2.14) quite closely. We can perform the symmetric integrations under the integrals shown in (5.1). We can write

$$\prod_{i=1}^{n} \left(\int d^4 r_i\right) N_{\mu\nu}^{(n)} \mathfrak{D}^{-1}(r,p,q) \prod_r e^{F_r/E_r} = f_{\mu\nu}(p,q,\alpha) e^{F/E}, \quad (5.3)$$

where F and E are the parametric functions for the complete cancelled graph. We show in Appendix C that g/E, the coefficient of $2m\nu$ in F/E, scales in the same way as g/C in the bare ladder graph, and so the same terms in $N_{\mu\nu}$ will be important, namely the terms

$$\propto \prod_{i=1}^{n} r_i^2 (p_\mu q_\nu + p_\nu q_\mu - g_{\mu\nu}\, p \cdot q). \quad (5.4)$$

Let us define the scaling more explicitly. We write [cf. (2.21)]

$$\begin{aligned} \delta_n &= \sigma_n \bar{\delta}_n, \\ \delta_{n-1} &= \sigma_n \sigma_{n-1} \bar{\delta}_{n-1}, \\ &\vdots \\ \delta_0 &= \prod_{i=0}^{n} \sigma_i \bar{\delta}_0, \end{aligned} \quad (5.5)$$

where δ_k represents the set $\{\alpha_k, \beta_k, \beta_k', \rho_{rk}\}$ and rk indexes those renormalization parts lying on the lines with index k.

Then integrate successively with respect to $r_1, r_2 \cdots r_n$. Completing the square and shifting the origin, we find the coefficient of $r_1'^2$ has the form

$$A_1 = \alpha_1 + \beta_1 + \beta_1' + \sum a_{r1}\rho_{r1} + O(\sigma_0) \quad (5.6)$$

and, continuing,

$$A_m = \alpha_m + \beta_m + \beta_m' + \sum a_{rm}\rho_{rm} + O(\sigma_{m-1}). \quad (5.7)$$

So we find, for the contribution to W_2,

$$C \propto \frac{1}{\omega} \frac{Im}{\pi} \int d\alpha_0 \prod_{i=1}^{n} \left(\int d\alpha_i d\beta_i d\beta_i'\right) \times \prod_r \int d\rho_r (-\ln\rho_r)^{n_r} \left(-\frac{\partial}{\partial\rho_r}\right) \int d\xi f_r(\xi) \frac{e^{F/E}}{(\Pi A_i)^3} \quad (5.8)$$

Now scaling as in (5.5) throws the integral into the standard form of Appendix A (since from Appendix C

$$\frac{g}{E} = \prod_{i=0}^{n} \sigma_i \frac{\bar{g}}{\bar{E}}\bigg) \quad , \quad (5.9)$$

leading to the asymptotic behavior

$$C \propto \frac{1}{\omega} \frac{Im}{\pi} \frac{1}{2m\nu} \times \frac{(\ln\nu)^{\sum_r n_r + n}}{(N_n + 1)(N_n + N_{n-1} + 2) \cdots (N_1 + \cdots + N_n + n)} \times \int d\bar{\alpha}_0 \prod_{i=0}^{n} \left(\int d\bar{\alpha}_i d\bar{\beta}_i d\bar{\beta}_i'\right) \times \prod_r \left\{\int d\rho_r \left(-\frac{\partial}{\partial\bar{\rho}_r}\right) \int d\xi f_r(\xi)\right\} \times \prod_{i=0}^{n} \delta(\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}_i' + \sum \bar{\rho}_{ri} - 1) \cdot \frac{(\bar{g}/\bar{E})^{-1}}{(\Pi \bar{A}_i)^3}\bigg|_{\sigma=0}. \quad (5.10)$$

From Appendix C we have

$$\frac{(\bar{g}/\bar{E})^{-1}}{(\Pi \bar{A}_i)^3}\bigg|_{\sigma=0} = \frac{1}{(\bar{\alpha}_0 + \sum a_{r0}\bar{\rho}_{r0})} \times \prod_{i=1}^{n} \frac{1}{(\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}_i' + \sum_r a_{ri}\bar{\rho}_{ri})^3} H_0(\bar{\alpha}, \bar{\beta}, \bar{\rho}), \quad (5.11)$$

where H_0 is homogeneous of degree zero with respect to each group of variables (indexed by i) separately. We have

$$H(\bar{\alpha}, \bar{\beta}, 0) = \left(\prod_{i=1}^{n} \frac{\bar{\alpha}_i}{\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}_i'} - \frac{1}{\omega}\right)^{-1}. \quad (5.12)$$

We can therefore use the result of Appendix B to simplify the integral of (5.10). Applying it to each group

$$\int d\bar{\alpha}_i d\bar{\beta}_i d\bar{\beta}'_i \prod_r d\bar{\rho}_{ri}\,\delta(\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}'_i + \sum \bar{\rho}_{ri} - 1) \times H_0(\bar{\alpha},\bar{\beta},\bar{\rho}) \cdot (\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}'_i + \sum a_{ri}\rho_{ri})^{-3} \quad (5.13)$$

successively, we can remove every ρ factor (and α_0) and reduce the amplitude to

$$C \propto \frac{1}{\omega}\frac{1}{\nu}\frac{(\ln\nu)^{\sum n_r + n}}{(N_n + 1)\cdots(N_n + \cdots + N_1 + n)} \times \prod_{i=1}^{n} \int d\bar{\alpha}_i d\bar{\beta}_i d\bar{\beta}'_i\,\delta(\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}'_i - 1) \times \delta\left(\prod \bar{\alpha}_i - \frac{1}{\omega}\right) \prod_r \int d\xi f_r(\xi) \quad (5.14)$$

Comparing equation (5.14) with (2.26) and (2.27), we see that if we can identify the factors

$$\int d\xi f_r(\xi) \quad (5.15)$$

with the coefficients of the logarithms in the asymptotic expansions of the renormalization parts, we have indeed justified the prescription of Sec. 2.

6. CONTRIBUTION OF RENORMALIZATION PARTS

As we have seen in Sec. 5, the terms from renormalization parts which are most important in the deep inelastic limit are those with the highest powers of the logarithm of the over-all scaling parameter. These are also the terms controlling the ultraviolet asymptotic behavior of the renormalization parts.[6] We shall calculate the important terms inductively, assuming for lower-order graphs the canonical form of (5.1).

We consider first the vertex parts. These are simplest because they have superficially no overlapping divergences. The important vertex graphs can be written uniquely in the skeleton form of Figs. 3(a) and (b). We calculate the renormalized vertex amplitude by plugging in the renormalized values of the subgraphs and making a single subtraction from the resulting Feynman amplitude.[10] If we cancel as many denominators as we can against the numerator factors, we obtain for the naive Feynman amplitude (neglecting m terms) for the vector vertex of Fig. 6

$$g^2 \int d^4r \frac{\gamma_5(\not{p} + \not{r})\gamma_\mu(\not{p} + \not{q} + \not{r})\gamma_5}{[(p + r)^2 - m^2][(p + q + r)^2 - m^2][r^2 - \lambda^2]} \times \prod_r \left[\int d\rho_r (-\ln\rho_r)^{n_r}\left(-\frac{\partial}{\partial\rho_r}\right)\int d\xi f_r(\xi)\, e^{F_r/E_r}\right]. \quad (6.1)$$

The same manipulations as in Sec. 3, with the scaling

$$\delta = \rho\bar{\delta}, \quad \delta \in \{x, y, z, \rho_r\}, \quad (6.2)$$

give for the important part of the renormalized amplitude

$$\Lambda_\mu \sim \gamma_\mu\, i\pi^2 g^2 \int d\rho \frac{(-\ln\rho)^{\sum n_r + 1}}{\sum n_r + 1}\left(-\frac{\partial}{\partial\rho}\right)\int d\bar{x}d\bar{y}d\bar{z} \times \prod_r \left(\int d\bar{\rho}_r\right)\delta(\bar{x} + \bar{y} + \bar{z} + \sum\bar{\rho}_r - 1)\prod_r\left[\left(-\frac{\partial}{\partial\bar{\rho}_r}\right) \times \int d\xi f_r(\xi)\right]\frac{\exp(\rho\bar{F}/\bar{E})}{\bar{B}^3}, \quad (6.3)$$

where

$$B = x + y + z + a_r\rho_r \quad (6.4)$$

and

$$\frac{F}{E} = \frac{xzp^2 + xyq^2 + yz(p + q)^2 + O(\rho_r) + O(m^2)}{x + y + z + O(\rho r)} \quad (6.5)$$

We see that the amplitude is now in the canonical form

$$\frac{(2\pi)^4}{i}\gamma_\mu \int d\rho(-\ln\rho)^{N+1}\left(-\frac{\partial}{\partial\rho}\right)\int d\xi f_U(\xi)\exp(\rho\bar{F}_U/\bar{E}_U). \quad (6.6)$$

And as in Sec. 3, we can write down the asymptotic behavior as any invariant is taken asymptotic

$$\Lambda_\mu \sim \frac{(2\pi)^4}{i}\gamma_\mu \int d\xi f_U(\xi)(\ln t)^{N+1}. \quad (6.7)$$

So we can indeed identify $\int d\xi f_U(\xi)$ as a contribution to the coefficient of $(\ln t)^{N+1}$ in the asymptotic expansion of Eq. (2.4). Now, of course, we use Appendix B to remove the $\bar{\rho}_r$. We can write

$$\int d\bar{x}d\bar{y}d\bar{z}\prod_r (d\bar{\rho}_r)\delta(\bar{x} + \bar{y} + \bar{z} + \sum\bar{\rho}_r - 1) \times \prod_r\left[\left(-\frac{\partial}{\partial\bar{\rho}_r}\right)\int d\xi f_r(\xi)\right](\bar{x} + \bar{y} + \bar{z} + \sum a_r\bar{\rho}_r)^{-3} = \left[\int d\bar{x}d\bar{y}d\bar{z}\,\delta(\bar{x} + \bar{y} + \bar{z} - 1)\right]\prod_r\left[\int d\xi f_r(\xi)\right], \quad (6.8)$$

whence

$$\int f_u(\xi)d\xi\,(\ln t)^{N+1} = \frac{-g^2}{32\pi^2}\frac{\prod_r\left[\int d\xi f_r(\xi)\right]}{N + 1}(\ln t)^{N+1}. \quad (6.9)$$

And we recognize this as a single term from the expansion of Eqs. (3.17). So, for the vector vertex we have justified the prescription of Sec. 3. The gluon vertex can be treated in just the same way.

The self-energy parts present more difficulty because they generally do have superficial overlapping divergences. We follow Appelquist and Primack in using Ward identities to treat them. For the fermion self-energy parts we have the identity

$$\partial^\mu \sum(p) = -\Lambda^\mu(p,p). \quad (6.10)$$

It is convenient to use this in the integrated form[10]

$$\sum(p) - \sum(p') = -\int_0^1 d\lambda (p - p')_\mu \Lambda^\mu(p^\lambda, p^\lambda), \quad (6.11)$$

where

$$p^\lambda = \lambda p + (1 - \lambda)p'. \quad (6.12)$$

Neglecting terms with a factor of m we have

$$\not{p}F(p^2) \sim \not{p}'F(p'^2) - \int_0^1 d\lambda(p - p')_\mu \Lambda^\mu(p^\lambda, p^\lambda) \quad (6.13)$$

and since this holds for all p and p' the terms proportional to p' on the right-hand side must cancel, giving

$$\not{p}F(p^2) \sim -\int_0^1 d\lambda p_\mu \Lambda^\mu(p^\lambda, p^\lambda). \quad (6.14)$$

We have seen that Λ^μ can be written in the canonical form of (6.6). Now we must show that it can be rewritten in a form corresponding to a self-energy diagram.

With each self-energy diagram the Ward identity associates all those vertex parts obtained by inserting a photon vertex in a fermion line. The important vertex graphs, of the form of Fig. 6, we obtain by inserting the photon vertex in the line x of Fig. 14. The photon vertex may be inserted either between two self-energy subdiagrams or directly into one, so that we may represent the graphs arising from Fig. 14 by Fig. 15.

We use the Ward identity (6.10) to write the subvertex in Fig. 15(b) as $-\partial^\mu \sum(p^\lambda + r)$ and looking at our canonical form for $\sum(p)$, Eq. (5.1), we see that we get the main contribution by differentiating the factor $(\not p^\lambda + \not r)$ outside the integral. If we now write out the amplitudes of Fig. 15 in the form (6.1), we see that graphs (a) and (b) give exactly equal and opposite contributions. And there is one more of type (a) than of type (b). So, corresponding to each skeleton of Fig. 14, we have the unrenormalized amplitude

$$g^2 \int d^4 r \frac{\gamma_5(\not p^\lambda + \not r)\gamma_\mu(\not p^\lambda + \not r)\gamma_5}{[(p^\lambda + r)^2 - m^2]^2[r^2 - \lambda^2]} \times \prod_r \left[\int d\rho_r (-\ln\rho_r)^{n_r}\left(-\frac{\partial}{\partial\rho_r}\right)\int d\xi f_r(\xi)\, e^{F_r/E_r}\right], \tag{6.15}$$

where the product runs over the subgraphs of Fig. 14. We choose to parametrize by

$$[(p^\lambda + r)^2 - m^2]^{-2} = \int_0^\infty dx\, x \exp\{x[(p^\lambda + r)^2 - m^2]\}. \tag{6.16}$$

Then symmetric integration gives us the exponential factor e^{F_S/E_S}, the factor corresponding to the self-energy graph of Fig. 14. Renormalizing and scaling, we find the contribution to $\Lambda^\mu(p^\lambda, p^\lambda)$ corresponding to a single self-energy diagram is

$$\sum_\delta i\pi^2 g^2 \gamma^\mu \int d\rho \frac{(-\ln\rho)^{\sum n_r + 1}}{\sum n_r + 1}\left(-\frac{\partial}{\partial\rho}\right)\int d\bar x d\bar y\, \bar x \prod_r d\bar\rho_r \times \delta(\bar x + \bar y + \sum\bar\rho_r - 1)\prod_r\left\{\left(-\frac{\partial}{\partial\bar\rho_r}\right)\int d\xi f_r(\xi)\right. \times \frac{\exp[\rho \bar F_S(p^\lambda)/\bar E_S]}{\bar B'^3}. \tag{6.17}$$

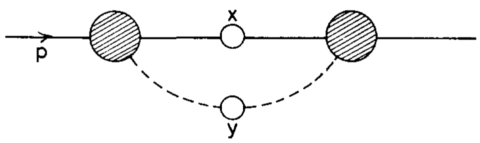

FIG. 14. Diagram giving important vertex part.

The sum runs over all possible skeletons,

$$B' = x + y + \sum a'_r \rho_r \tag{6.18}$$

and

$$\frac{F_S(p^\lambda)}{E_S} = \frac{xy(p^\lambda)^2 + O(\bar\rho_r) + O(m^2)}{x + y + \sum a'_r\rho_r}. \tag{6.19}$$

Now since (6.17) holds for all p' we can replace $(p^\lambda)^2$ by $[\lambda^2 p^2 + (1-\lambda)^2 O(m^2)]$, and we might as well drop the second term since mass terms in the exponential do not affect any of our results. Substituting (6.17) in (6.14), we find we can write $\sum(p)$ in the form

$$\sum(p) \sim \frac{(2\pi)^4}{i} \not p \int_0^1 d\lambda \int d\rho (-\ln\rho)^{N+1}\left(-\frac{\partial}{\partial\rho}\right) \times \int d\xi f_F(\xi)\exp\left(\frac{F_S(\lambda p)}{E_S}\right), \tag{6.20}$$

differing from the canonical form by the λ factors. However, no singularity arises from the λ integration so that we can perform all the usual manipulations under the integral. We can see that

$$\sum(p) \sim \frac{(2\pi)^4}{i} \not p \int d\xi f_F(\xi)\left(\int_0^1 d\lambda\right)\ln^{N+1}(-p^2) \tag{6.21}$$

as $p^2 \to -\infty$ so that we can identify $\int d\xi f_F(\xi)$ as a contribution to the coefficient f_{N+1} in Eq. (3.1).

The λ factors do not affect the arguments of Sec. 5. We can simply wait until the $\bar\rho$ terms have been removed and then trivially integrate. It is also easy to see that they do not affect the argument when the expression of (6.20) is inserted in another renormalization part. So we may replace (6.20) by the canonical form.

As before we use Appendix B to remove the $\bar\rho_r$ from $\int d\xi f_F(\xi)$. We find

$$\int f_F(\xi)d\xi \cdot \ln^{N+1}(-p^2) = \frac{g^2}{32\pi^2}\frac{\prod_r[\int d\xi f_r(\xi)]}{N+1}\ln^{N+1}(-p^2), \tag{6.22}$$

which has the form of a term from the expansion of (3.26).

Gluon self-energy parts are more complicated still. We first show that the important graphs have the incoming and outgoing lines attached to the same fermion loop. For if not, they can be divided into subgraphs joined by two or more gluon lines, and except for box graphs such

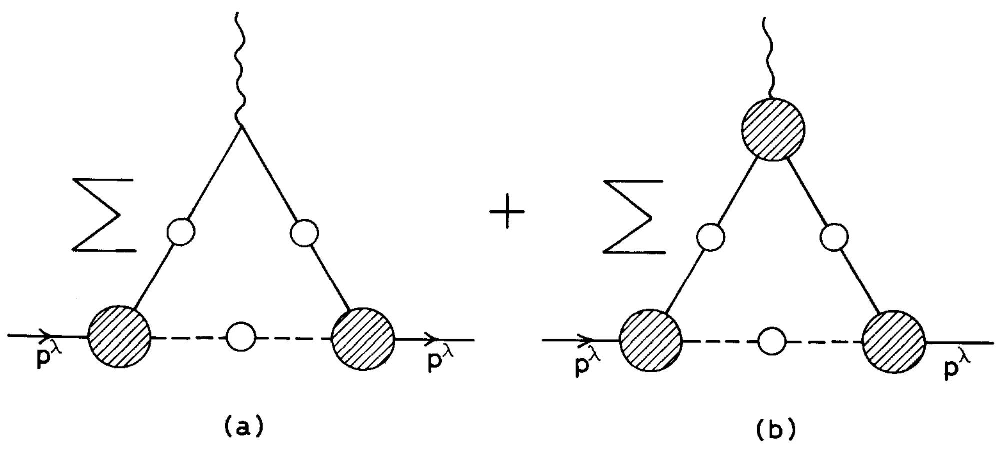

FIG. 15. Vertex diagrams arising from Fig. 14.

graphs are superficially convergent and therefore do not lead to a logarithmic enhancement of the highest order for the whole diagram. At first sight, box graphs, e.g., Fig. 16, look important. It looks as if we must make renormalization subtractions at four levels, and in our previous experience each subtraction adds a logarithm to the asymptotic behavior. However, as Salam pointed out,[11] counterterms of the form of Fig. 17 are independent of the momentum p and may therefore be absorbed in the over-all subtraction. This reduces the number of subtractions to three with no overlapping divergences so we see that the diagram only gives a $[\ln(-p^2)]^3$, nonleading, enhancement. For the same reason we need not worry about box diagrams occurring in the form of Fig. 18. The counterterm corresponding to the box is of the form of Fig. 17 and may therefore be neglected. We might have deduced that box graphs are unimportant from the fact that Appelquist and Primack obtained agreement with renormalization group calculations without considering them.

We can now write a Ward identity connecting the remaining self-energy diagrams and photon-2 gluon vertices. We have

$$-\partial^\mu K(p) = V^\mu(p,p). \tag{6.23}$$

$V^\mu(p,p)$ represents the sum of all vertex graphs obtainable by inserting a photon vertex in, say, the top fermion line of a self-energy graph. Naturally, if we made all possible insertions, the amplitude would be zero by Furry's theorem. (6.23) closely resembles an identity found by Salam[12] for a charge symmetric theory. He, however, had to insert vector vertices in meson lines also.

Just as before, we rewrite (6.23) as

$$K(p) \sim -p_\mu \int_0^1 d\lambda\, V^\mu(p^\lambda,p^\lambda). \tag{6.24}$$

The important graphs contributing to V^μ are shown in Fig. 19(a) and are obtained by inserting the photon vertex in the line x of Fig. 19(b). By the same arguments as above we write $V^\mu(p^\lambda,p^\lambda)$ in terms of the parametric functions F_T/E_T of the self-energy graph. We can write the contribution to $V^\mu(p^\lambda,p^\lambda)$ from a single self-energy diagram:

$$\begin{aligned}\lambda \cdot \sum_\delta -i\pi^2 g^2 p^\mu \int d\rho\, &\frac{(-\ln\rho)^{\sum n_r+1}}{\sum n_r+1}\left(-\frac{\partial}{\partial\rho}\right)\\ \times \int d\bar{x}d\bar{y}\, \Pi\, d\bar{\rho}_r\, &\frac{\bar{x}(-8)[2\bar{y}+\tfrac{1}{2}\bar{x}+O(\bar{\rho}_r)]}{\bar{B}''}\\ \times\, \delta(\bar{x}+\bar{y}+\textstyle\sum\bar{\rho}_r-1)\, &\prod_r\left[\left(-\frac{\partial}{\partial\bar{\rho}_r}\right)\right.\\ \times \left.\int d\xi f_r(\xi)\right] &\frac{\exp(\rho\bar{F}_T/\bar{E}_T)}{\bar{B}''^3}\end{aligned} \tag{6.25}$$

almost identical to (6.17). This leads to

$$\begin{aligned}K(p) \sim \frac{(2\pi)^4}{i} p^2 \int 2\lambda d\lambda \int d\rho(-\ln\rho)^{N+1}\left(-\frac{\partial}{\partial\rho}\right)\\ \times \int d\xi f_G(\xi) \exp[F_T(\lambda p)/E_T]\end{aligned} \tag{6.26}$$

with

$$\int d\xi f_G(\xi) = \frac{g^2}{8\pi^2}\prod_r\left[\int d\xi f_r(\xi)\right] \tag{6.27}$$

agreeing with (3.25). Once again the λ factors may be removed without affecting anything, to give us the canonical form for the gluon self-energy part. This completes the induction.

7. LADDER GRAPHS WITH EXCHANGE OF FERMIONS

As we saw from the example D of Sec. 4 there is another important class of diagrams beside those we have already considered. This is the class of ladder graphs with exchange of fermions. We use Fig. 20 to define our notation for a graph with $N+1$ rungs.

Let us first consider the Feynman integral corresponding to a bare graph with no numerator factors. Scaling as in (2.21), we find the amplitude takes the form

$$\begin{aligned}A \propto \int_0 \prod^m (d\sigma_i\sigma_i{}^i) \prod_{i=1}^m [\int d\bar{\alpha}_i d\bar{\beta}_i d\bar{\beta}'_i\, \delta(\bar{\alpha}_i+\bar{\beta}_i+\bar{\beta}'_i\\ +\sigma_{i-1}-1)] \exp\left[2m\nu\left(\prod_0^m \sigma_i\right)\frac{\bar{g}}{\bar{C}}+\sigma_m\frac{\bar{d}}{\bar{C}}\right]\bar{C}^2.\end{aligned} \tag{7.1}$$

In this case the leading behavior comes completely from the σ_0 scaling.

Now we consider the effect of possible numerator factors. We write r'_j for the jth loop momentum with origin displaced. Then a factor $r_j'^2$ leads, upon symmetric integration to an extra factor

$$[\alpha_j+\beta_j+\beta'_j+O(\sigma)]^{-1}, \tag{7.2}$$

which contains scaling factors $(\Pi_j^m \sigma_i)^{-1}$. So if we have a factor $r_j'^2$ for each segment of the ladder, we exactly cancel the factor $\Pi_0^m \sigma_i{}^i$ in (7.1). This is the case in the simple ladder graph.

Another important factor is $p\cdot q\, \Pi_0^{j-1}\sigma_i$, which we associate with the jth segment. Assuming a factor $r_k'^2$ for every other segment, this multiplies the integral of (7.1) by $p\cdot q$ and changes the product of scale factors to $\Pi_0^m \sigma_i$. Looking at Appendix A, we see that the factor $\Pi_0^m \sigma_i$ moves the pole in the Mellin transform one unit to the left, and the factor $p\cdot q$ moves the pole

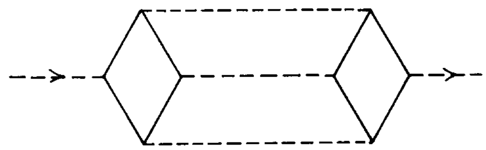

FIG. 16. Important-looking diagram.

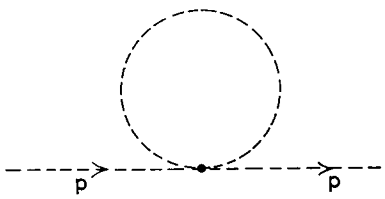

FIG. 17. Diagram arising during renormalization.

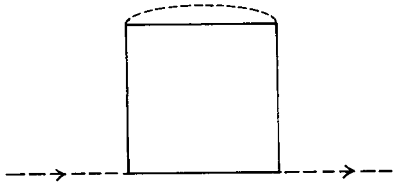

FIG. 18. Graph containing single box diagram.

one unit to the right. So asymptotically this amplitude is as important as the previous one. We can construct factor $p\cdot q\,\Pi_0^{j-1}\,\sigma_i$ from displacement terms in the graph of Fig. 20. In simple ladders they cancel themselves out as we see in Appendix D.

We must also consider the class of numerator factors giving rise to divergent box graphs. We may represent an amplitude corresponding to such a numerator by Fig. 21, where the blob represents the renormalized box graph. We can write the leading contribution from the box graph in the usual form,

$$\int d\rho\,\ln^n\rho\left(-\frac{\partial}{\partial\rho}\right)\int d\xi f(\xi)\,\exp(\rho^{\bar{F}/\bar{g}}). \tag{7.3}$$

It is intuitively fairly clear that the amplitude defined by Fig. 21 cannot have an imaginary part, since when we cut it across the middle in optical theorem fashion we do not obtain any recognizable $\gamma-p$ scattering reaction. This intuition can readily be verified by a calculation similar to that of Sec. 5.

Now that we have disposed of the divergences, we can use the prescription of Sec. 2 to calculate the contribution of the remaining terms. The chief difficulty lies in enumerating the important numerator factors.

The numerator $N_{\mu\nu}$ is the product of $n+1$ traces

$$N_{\mu\nu}=(-2)^m\prod_{i=0}^{m}T_i. \tag{7.4}$$

The factor 2^m arises because we must count boxes with the fermion lines directed in both senses.[13] We use GP's Fierz transformation technique to rewrite the traces in a tractable form. As before only the vector couplings give important contributions, and so we can write

$$T_i\sim D_i^{\alpha_0}\prod_{j=1}^{n_i}(2r_{ij\,\alpha_{j-1}}r_{ij}^{\alpha_j}-r_{ij}^2\,g_{j-1}^{\alpha_j})\,C_{i\,\alpha_{n_i}}, \tag{7.5}$$

where

$$D_i^{\alpha}=[g^{\alpha}{}_{\mu}(q+r_{01})_{\nu}+g^{\alpha}{}_{\nu}(q+r_{01})_{\mu}-g_{\mu\nu}(q+r_{01})^{\alpha}]\quad\text{if } i=0 \tag{7.6}$$

$$=-(r_{i0}-r_{i1})^{\alpha}\quad\text{otherwise}$$

and

$$C_{i\alpha}=p_{\alpha}/m\quad\text{if } i=m$$
$$=4[2r_{if\,\alpha}\,r_{if}\cdot(r_{if}-r_{i+10})-(r_{if}-r_{i+10})_{\alpha}\,r_{if}^2] \tag{7.7}$$

otherwise.

Now the only important part of the factors $(2r_{\alpha_{j-1}}r^{\alpha_j}-r^2g_{\alpha_{j-1}}{}^{\alpha_j})$ comes from taking the displaced loop momentum r' from r as we show in Appendix D. Symmetrizing then gives us the factor $(-\frac{1}{2}r'^2g_{\alpha_{j-1}}{}^{\alpha_j})$, just as for the simple ladder. For the other factors, however, we must consider also the displacement terms. We have

$$r_j=r'_j+\frac{\alpha_j}{\sum_j}r'_{j+1}+\frac{\alpha_j\alpha_{j+1}}{\sum_j\sum_{j+1}}r'_{j+2}+\cdots+\prod_j^N\frac{\alpha_k}{\sum_k}p-\left(\prod_0^{j-1}\alpha_k\Big/\prod_1^j\sum_k\right)q+O(\sigma). \tag{7.8}$$

As we see in Appendix D, important contributions can arise from loop momenta linking factors $D_{i+1}^{\beta}C_{i\alpha}$ and $D_{j+1}^{\gamma}C_{j\delta}$. We shall first of all, however, consider the contributions which do not involve such cross linkings. We find that such contributions to $(-1)D_{i+1}^{\beta}C_{i\alpha}$ can be written symmetrized as

$$\begin{aligned}g_{\alpha}^{\beta}\Bigg[&(-\tfrac{1}{2}r_{if}'^2)(-\tfrac{1}{2}r_{i+10}'^2)\left(6\frac{\alpha_{if}}{\sum_{if}}+2\right)\\&+(-\tfrac{1}{2}r_{if}'^2)\,2p\cdot q\left(\frac{\alpha_{00}}{\sum_{i+10}}\prod_1^N\frac{\alpha_k}{\sum_k}\right)\left(6\frac{\alpha_{if}}{\sum_{if}}+2\right)\\&+2p\cdot q\left(\frac{\alpha_{00}}{\sum_{if}}\prod_1^N\frac{\alpha_k}{\sum_k}\right)(-\tfrac{1}{2}r_{i+10}'^2)\,4\frac{\alpha_{if}}{\sum_{if}}\\&+2p\cdot q\left(\frac{\alpha_{00}}{\sum_{if}}\prod_1^N\frac{\alpha_k}{\sum_k}\right)2p\cdot q\left(\frac{\alpha_{00}}{\sum_{i+10}}\prod_1^N\frac{\alpha_k}{\sum_k}\right)2\frac{\alpha_{if}}{\sum_{if}}\Bigg]\end{aligned} \tag{7.9}$$

So, effectively, D_0 and C_m couple together to give a factor

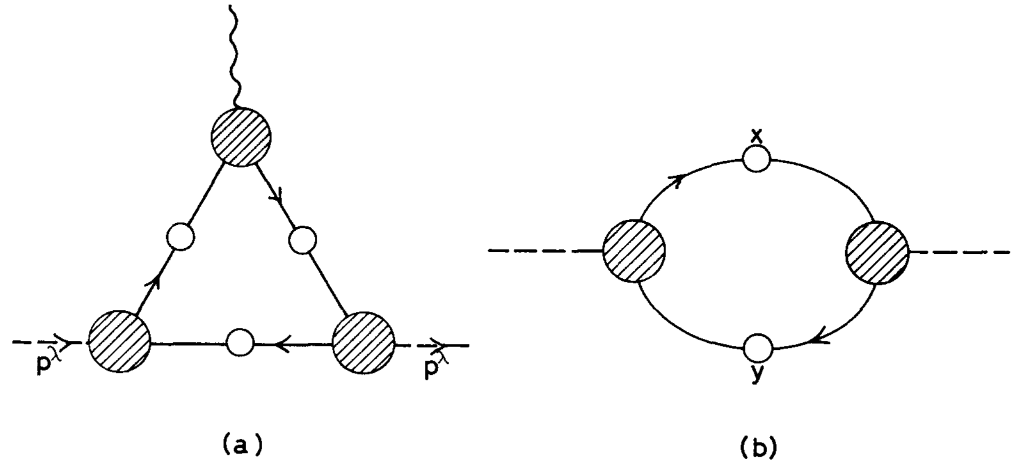

FIG. 19. Important vertex graphs and the self-energy graphs from which they are obtained.

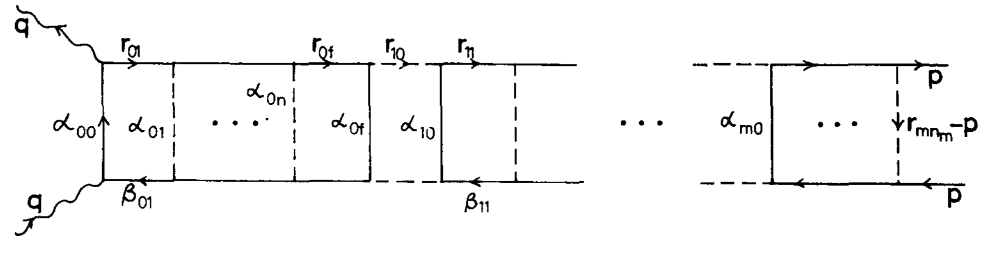

FIG. 20. Ladder graph with fermion exchange.

$$(1/m)\,(p_\mu q_\nu + p_\nu q_\mu - g_{\mu\nu} p\cdot q), \tag{7.10}$$

where we have considered only those terms of the tensor decomposition which contain a factor q.

We must introduce two more functions besides y to describe the effects of the renormalization parts in the segments labelled f and 0. We write

$$u = S^3 V^2, \qquad v = SV^2T^2. \tag{7.11}$$

Then we can write the contribution to W_2 in exactly the form of (2.24) except that corresponding to the segments (if) and $(i+1\ 0)$ we have the operation

$$u\left(-\ln \prod_{(if)}^{N} \sigma_k\right) \int d\bar{\alpha}_f d\bar{\beta}_f d\bar{\beta}'_f\, \delta(\bar{\alpha}_f + \bar{\beta}_f + \bar{\beta}'_f + \sigma_{in_i} - 1)$$
$$\times\, v\left(-\ln \prod_{(i+1\ 0)}^{N} \sigma_k\right) \int d\bar{\alpha}_0 d\bar{\beta}_0 d\bar{\beta}'_0 \delta(\bar{\alpha}_0 + \bar{\beta}_0 + \bar{\beta}'_0 + \sigma_{if} - 1)$$
$$\times\left[\left(2p\cdot q \prod_0^N \sigma_k \prod_1^N \frac{\bar{\alpha}_0}{\overline{\sum}_k}\right)^2 2\,\frac{\bar{\alpha}_{if}}{\overline{\sum}_{if}} + \left(2p\cdot q \prod_0^N \sigma_k \prod_1^N \frac{\bar{\alpha}_n}{\overline{\sum}_k}\right)\left(10\,\frac{\bar{\alpha}_{if}}{\overline{\sum}_{if}} + 2\right) + 6\,\frac{\bar{\alpha}_{if}}{\overline{\sum}_{if}} + 2\right)\right]. \tag{7.12}$$

From Appendix A we see that the effect of a factor $(2p\cdot q\ \Pi_0^N\, \sigma_k)$ is to make the change

$$(\bar{g}/\bar{C})^{-1} \to (\bar{g}/\bar{C})^{-1-p}\, \Gamma(1+p) \tag{7.13}$$

in the coefficient of the leading term. And we have the relation[14]

$$\frac{Im}{\pi}\,(x - i0)^{-p-1} = \frac{(-1)^p}{\Gamma(p+1)}\,\delta^{(p)}(x). \tag{7.14}$$

So, writing for the factor in square brackets in (7.12) the expression

$$\left[\left(\prod_1^N \bar{\alpha}_k\right)^2 \left(-\frac{\partial}{\partial\tau}\right)^2 2\bar{\alpha}_{if} + \left(\prod_1^N \bar{\alpha}_k\right)\left(-\frac{\partial}{\partial\tau}\right)(10\bar{\alpha}_{if} + 2) + (6\bar{\alpha}_{if} + 2)\right], \tag{7.15}$$

we can express the contribution to νW_2 in the form of (2.28). It is straightforward to verify that the arguments of Sec. 5 still apply, so that the above procedure is valid. In particular, we can still use the reduction procedure of Appendix B.

We can reduce the amplitude to a factored form now by Mellin transforming with respect to $\tau = 1/\omega$. Integrating by parts, we can exchange the operators $(-\partial/\partial\tau)$ acting upon $\delta(\tau - \Pi\,\bar{\alpha}_k)$ for operators $(\partial/\partial\tau)$ acting upon τ^λ. We then find that we can perform the parametric integrations explicitly, and we can sum up all the terms we have just considered in the form

$$E(s) = z(\ln\nu)\left[\sum_{i=0}^{\infty}\left\{\sum_{n_i=0}^{\infty} (\textstyle\int \kappa y)^{n_i} L_i(s) \int \kappa u \int \kappa v\right\}^i \times \sum_{n=0}^{\infty} (\textstyle\int \kappa y)^n - 1\right], \tag{7.16}$$

where

$$\kappa = (g^2/16\pi^2)[(\lambda+1)(\lambda+2)]^{-1} \tag{7.17}$$

and for the terms with no cross linking of loop momenta

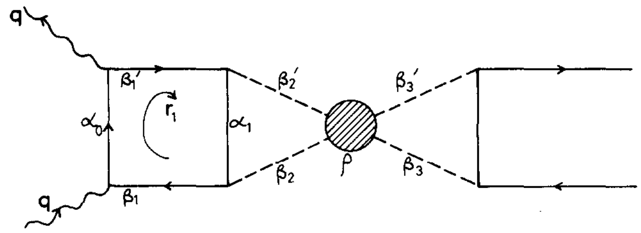

FIG. 21. Graph containing divergent box.

$$L_i(s) = [2/(\lambda+3)][(2\lambda+2)s^2 + (12\lambda+6)s + 8\lambda + 12]. \tag{7.18}$$

The integrals must be regarded as operating on everything to their right and the final argument is $\ln\nu$. We obtain the contribution to $\mathfrak{M}_{1/\omega}(\nu W_2)$ from $E(s)$ by the rule—expand $E(s)$ in powers of s and replace s^n by $\Gamma(\lambda+1)/\Gamma(\lambda-n+1)$.

We see in Appendix D that the terms involving the cross linking of loop momenta give rise to factors similar in form to (7.9). For every different linking, then, we obtain an expression of the form of (7.16), but with different coefficients of s^n in (7.18). To find the complete contribution to $\mathfrak{M}(\nu W_2)$, we must then sum $E(s)$ over all possible cross linkings.

We can simplify our expressions somewhat as follows. First it is convenient to make the substitution

$$x \to Q = \int_0^x \kappa y(x')\,dx'. \tag{7.19}$$

Then, since

$$\sum_{n=0}^{\infty} (\textstyle\int dQ)^n X = e^Q \int_0^Q e^{-Q'} dQ' X(Q') - X, \tag{7.20}$$

we have

$$E(s) = z\left\{\sum_{i=0}^{\infty} [L_i(e^Q \textstyle\int e^{-Q} + 1) \int w^{-1} \int w]^i \times (e^Q \int e^{-Q} + 1) - 1\right\}. \tag{7.21}$$

An attempt to evaluate

$$T = \sum_{i=0}^{\infty} \{(e^Q \textstyle\int e^{-Q} + 1)\Lambda \int w^{-1} \int w\}^i X \tag{7.22}$$

leads to the differential equation

$$\frac{d^2Z}{dQ^2} - \left(1 - \frac{d}{dQ}(\ln w)\right)\frac{dZ}{dQ} - \left(\Lambda + \frac{d}{dQ}(\ln w)\right) Z = X \tag{7.23}$$

for $Z = T + X$. We cannot solve this for arbitrary w, but if w takes the form e^{kQ} (as in the case when we use the amplitudes calculated in Sec. 3), it reduces to an equation with constant coefficients and we can solve it. Then we can write

$$E(\Lambda) = z\cdot\exp[\tfrac{1}{2}(1-k)Q]\left(\frac{\tfrac{1}{2}(1+k)}{[\tfrac{1}{4}(1+k)^2 + \Lambda]^{1/2}} \times \sinh Q[\tfrac{1}{4}(1+k)^2 + \Lambda]^{1/2} + \cosh Q[\tfrac{1}{4}(1+k)^2 + \Lambda]^{1/2}\right) - z, \tag{7.24}$$

where $E(\Lambda)$ is defined by replacing L_i by Λ in (7.16) or (7.21). We can expand this in powers of Λ as

$$\sum_{r=0}^{\infty} \frac{\Lambda^r}{r!}\, z\cdot\exp[\tfrac{1}{2}(1-k)Q]\left[\left(\frac{1}{2}\frac{d}{d\mu}\frac{1}{\mu}\right)^r \sinh\mu Q + \left(\frac{1}{2}\frac{1}{\mu}\frac{d}{d\mu}\right)^r \times \cosh\mu Q\right]_{\mu=(1+k)/2} - z. \tag{7.25}$$

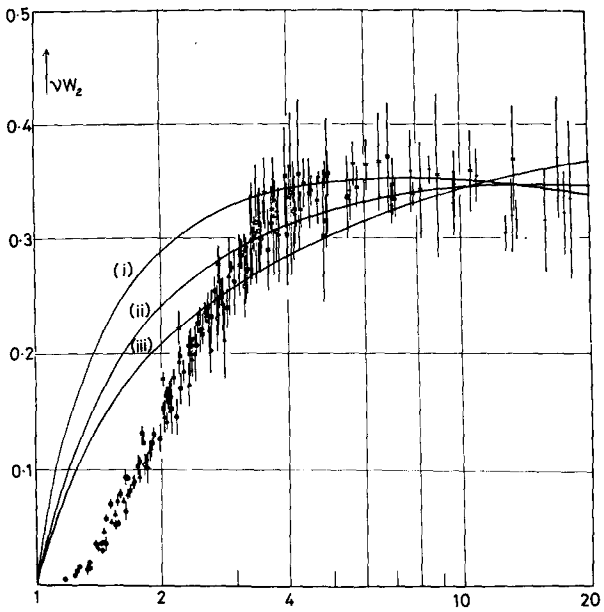

FIG. 22. Comparison of the experimental curve for νW_2 with curves described by the model with $(g^2/16\pi^2)\, Y =$ (i) 5 (ii) 6 (iii) 7.

Then, replacing Λ^r by

$$\sum\left(\prod_{i=1}^{r} L_i(s)\right), \tag{7.26}$$

where the sum runs over all possible cross linkings, we have a prescription for calculating $\mathfrak{M}(\nu W_2)$ as a series parametrized by the number of fermion rungs in the corresponding ladder graphs.

It is possible, with a great deal of algebra, to evaluate expression (7.26) in a fairly compact form. When the substitution

$$S^n \to \lambda(\lambda-1)\cdots(\lambda-n+1) \tag{7.27}$$

is made, the almost miraculous result

$$\sum\left(\prod^{r} L_i(s)\right) \to [4(\lambda+1)(\lambda+2]^r \tag{7.28}$$

is obtained. This means that, with $\Lambda = 4(\lambda+1)(\lambda+2)$, Eqs. (7.24) and (7.25) give $\mathfrak{M}(\nu W_2)$ directly.

8. DISCUSSION

In this section we discuss what we can say about the amplitude νW_2 on the basis of the foregoing results.

As we remarked in the Introduction the expressions for y and z calculated in Sec. 3 are unsatisfactory. Formally we can easily take account of nonleading contributions from renormalization parts. We may choose to consider the term involving $\ln^{n-r}\sigma \ln^r \bar{\rho}$ from the amplitude

$$\int d\bar{\rho}\, \ln^n \sigma\bar{\rho}\left(-\frac{\partial}{\partial\bar{\rho}}\right)\int d\xi f(\xi)\, e^{F/E}. \tag{8.1}$$

Then we can carry the argument of Sec. 5 as far as Eq. (5.10), but we can no longer apply Appendix B to remove the $\bar{\rho}$'s. So the contributions from different renormalization parts in the same segment can no longer be factorized and separated from the basic ladder parameters. We can still sum up all the terms in one segment in a single function, however, to write

$$\nu C \sim \frac{1}{\omega}\int d\chi\zeta(\chi)\cdot\prod_{i=1}^{n}\left\{\int_0^{x_{i+1}} x_i^{n_i}\, dx_i \int \eta_i(\chi)\,d\chi\right\}\delta\left(\frac{g}{E}\right). \tag{8.2}$$

A Mellin transform factorizes the δ function just as before, and we obtain

$$\mathfrak{M}(\omega\, \nu W_2) = \zeta(\ln\nu, \lambda)\,[\exp \int_0^{\ln\nu} \eta(x,\lambda)\,dx - 1] \tag{8.3}$$

from the graphs with gluon rungs only. We can no longer identify the functions η and ζ with products of asymptotic form factors. We shall take this result as an excuse to use formula (2.38) making whatever assumptions we like about the form of y and z.

It is convenient to regard the Mellin transform with respect to $1/\omega$ as a Laplace transform with respect to $t = \ln\omega$. Then we can write the contribution from the gluon ladders in the form

$$\nu W_2 \sim z\,\mathcal{L}^{-1}\left[\sum_1^{\infty}\left(\frac{a}{(\lambda+1)(\lambda+2)}\right)^n \Big/ n!\right], \tag{8.4}$$

which can be rewritten as a sum of convolutions of $a(e^{-t} - e^{-2t})$. In fact, instead of using the Mellin transform to factorize the expression (2.27), we might have written it as a convolution using the property

$$\delta(\alpha_1 - e^{-t}) * \delta(\alpha_2 - e^{-t}) = \delta(\alpha_1\alpha_2 - e^{-t}). \tag{8.5}$$

By working in terms of repeated convolutions, it is easy to see that

$$\int_1^{\infty} \frac{\nu W_2(\omega, \ln\nu)}{\omega}\, d\omega = z\left(\exp\frac{g^2}{32\pi^2}\cdot Y - 1\right) \tag{8.6}$$

and that the contribution to this integral from the n-rung ladder has Poisson distribution with respect to n. This is because, writing

$$\| f \| = \int_0^{\infty} f(t)\,dt, \tag{8.7}$$

we have

$$\| f * g \| = \| f \| \cdot \| g \| . \tag{8.8}$$

If we want to preserve scaling in this model, we must assume that the functions Y and z both tend to constants limits as $\ln\nu$ goes to infinity. Making this assumption, we find that we can obtain the right general shape for νW_2 (see Fig. 22), but with unsatisfactory behavior near $\omega = 1$, where our amplitude behaves like

$$(g^2/16\pi^2)\cdot Y\cdot z\cdot(1/\omega - 1/\omega^2). \tag{8.9}$$

If we are prepared to make assumptions about the form of u and v, we can also calculate the contribution from graphs with fermion rungs. Using the formula (7.25) and taking k to be small, we find that the contribution from ladders with one fermion pair is of the same order as the contribution from the graphs with gluon rungs only. This would predict that the present scaling law should break down when antibaryon production becomes important, as was recently suggested by Wilson.[15]

ACKNOWLEDGMENTS

I would like to thank Professor J. C. Polkinghorne for suggesting this problem and reading the manuscript. I also thank the Science Research Council for a maintenance grant.

APPENDIX A: ASYMPTOTIC VALUE OF A CLASS OF INTEGRALS

We find that the amplitudes whose asymptotic behavior we wish to calculate can be reduced to an integral of the following form:

$$A \sim \int \prod_{i=0}^{n}\left[d\sigma_i \sigma_i^{p} \ln^{n_i}\left(\prod_{j=i}^{n} \sigma_j\right)\right] \mathfrak{F}\left(\exp\left(-\prod_{j=0}^{n} \sigma_j \cdot \bar{g}\nu\right)\right), \tag{A1}$$

where we have taken ν to be the asymptotic variable. $\mathfrak{F}$ may be almost any linear operation—multiplication or integration or differentiation with respect to internal variables of $\bar{g}$.

Performing a Mellin transform

$$\mathfrak{M}(A) = \int_0^\infty \nu^{-\beta-1} A d\nu \tag{A2}$$

(we now use a different definition from that of Sec. 2 to conform with Ref. 7) gives

$$\mathfrak{M}(A) \sim \Gamma(-\beta) \int \prod_0^n \left[d\sigma_i \sigma_i^{p} \ln^{n_i}\left(\prod_i^n \sigma_j\right)\right] \mathfrak{F}\left(\bar{g}\left(\prod_0^n \sigma_j\right)^{\beta}\right)$$

$$= \Gamma(-\beta) \int \prod_0^n \left[\sigma_i^{\beta} \ln^{n_i}\left(\prod_i^n \sigma_j\right) \cdot d\sigma_i\right] f(\sigma,\beta). \tag{A3}$$

Then we assert that the behavior of this integral near its leading singularity in β (at $\beta = -p-1$) is

$$\mathfrak{M}(A) = \frac{\Gamma(p+1)(-1)^{\Sigma_0^n n_i}}{(\beta+p+1)^{n+1+\Sigma_0^n n_i}} \times \frac{(n_0+\cdots+n_n+n)!\, f(\sigma,\beta)|_{\sigma=0,\,\beta=-1-p}}{(n_n+1)\cdots(n_n+\cdots+n_1+n)} \tag{A4}$$

Clearly this requires some restriction on $\mathfrak{F}$ to ensure that $f(\sigma,\beta)|_{\sigma=0,\,\beta=-1-p}$ is not zero or infinity. So the asymptotic behavior of the amplitude itself is given by

$$A \sim \frac{\Gamma(1+p)}{\nu^{1+p}} \frac{(\ln\nu)^{\Sigma_0^n n_i+n}(-1)^{\Sigma_0^n n_i} f(0,-1-p)}{(n_n+1)\cdots(n_n+\cdots+n_1+n)}. \tag{A5}$$

We prove Eq. (A4) by induction. First we remark that we can expand the factors $\ln^{n_i}(\Pi_i^n \sigma_j)$ in multinomial series, so that we can rewrite (A3) as

$$\mathfrak{M}(A) = \sum C(\mathbf{m},\mathbf{n})\, \Gamma(-\beta) \int \prod_{i=0}^{n} (\sigma_i^{p+\beta} \ln^{m_i}\sigma_i\, d\sigma_i) f(\sigma,\beta) \tag{A6}$$

with

$$\sum_{i=0}^{n} m_i = \sum_{i=0}^{n} n_i. \tag{A7}$$

Now, integrating by parts repeatedly with respect to each variable in turn and keeping only the most singular terms, we arrive at

$$\mathfrak{M}(A) \sim \sum_{\mathbf{m}} C(\mathbf{m},\mathbf{n})\, \Gamma(-\beta) \frac{(-1)^{\Sigma_0^n m_i}}{(\beta+p+1)^{\Sigma_0^n m_i+n+1}} \times \left(\prod_0^n m_i!\right) \int \prod_0^n \left[\sigma_i^{\beta+1+p}\, d\sigma_i \left(-\frac{\partial}{\partial\sigma_i}\right)\right] f(\sigma,\beta). \tag{A8}$$

The integral is now convergent when we set $\beta = -1-p$ and so we obtain

$$\mathfrak{M}(A) = \Gamma(p+1) \sum_{\mathbf{m}} C(\mathbf{m},\mathbf{n}) \left(\prod_0^n m_i!\right) \times \frac{(-1)^{\Sigma n_i}}{(\beta+p+1)^{\Sigma n_i+n+1}} f(0,-1-p). \tag{A9}$$

Here we have established the general form of Eq. (A4). We need only evaluate the sum over the multinomial coefficients.

Clearly (A4) holds for $n=0$. Suppose it is true up to $n = m-1$. Then for $n=m$ we can write

$$\mathfrak{M}(A) \sim \sum_{r=0}^{n_0} \binom{n_0}{r} \int (d\sigma_0 \sigma_0^{p+\beta} \ln^{n_0-r}\sigma_0) \times \Gamma(-\beta) \int \prod_{i=1}^{m} \left[d\sigma_i \sigma_i^{p+\beta} \ln^{n_i'}\left(\prod_i^n \sigma_j\right)\right] f(\sigma,\beta) \tag{A10}$$

where

$$\begin{aligned} n_i' &= n_i, & i &\neq 1, \\ &= n_i + r, & i &= 1. \end{aligned} \tag{A11}$$

Then the second factor in (A10) has the form of (A3) for $n = m-1$ if we just rename the index, i. So we can write

$$\mathfrak{M}(A) \sim \sum_r^{n_0} \binom{n_0}{r} (n_0-r)!\, \frac{(-1)^{n_0-r}}{(\beta+p+1)^{n_0+1-r}} \times \frac{\Gamma(-\beta)(-1)^{\Sigma_1^m n_i}}{(\beta+p+1)^{\Sigma_1^m n_i'+m}} \times \frac{(n_1+r+n_2+\cdots+n_m+m-1)!\, f(0,-1-p)}{(n_m+1)\cdots(n_m+\cdots+n_2+m-1)}, \tag{A12}$$

$$\sum_{r=0}^{n_0} \binom{n_0}{r} (n_0-r)!\,(n_1+\cdots+n_m+r+m-1)! = n_0!(n_1+\cdots+n_m+m-1)! \sum_{r=0}^{n_0} \times \binom{n_1+\cdots+n_m+r+m-1}{n_1+\cdots+n_m+m-1}. \tag{A13}$$

The sum on the right-hand side is equal to the coefficient of $x^{n_1+\cdots+n_m+m-1}$ in

$$(1+x)^{n_1+\cdots+n_m+m-1} \sum_{r=0}^{n_0} (1+x)^r \tag{A14}$$

and that is just

$$\binom{n_0+\cdots+n_m+m}{n_1+\cdots+n_m+m}. \tag{A15}$$

If we substitute this back in (A13), we get

$$\sum_{r=0}^{n_0} \binom{n_0}{r} (n_0-r)!\,(n_1+\cdots+n_m+r+m-1)! = \frac{(n_0+\cdots+n_m+m)!}{(n_1+\cdots+n_m+m)} \tag{A16}$$

and substituting this into (A12) gives us (A4), completing the induction.

APPENDIX B

In this appendix we show how to simplify a class of

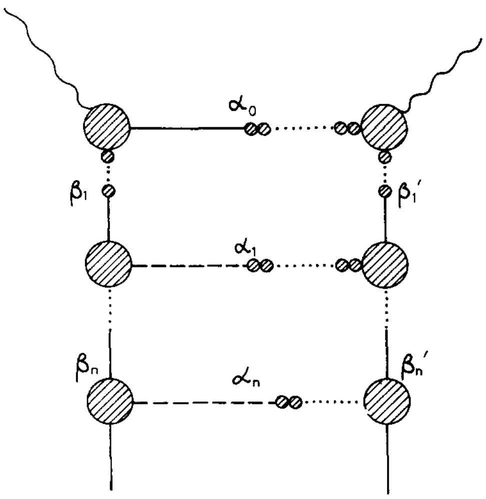

FIG. 23. Cancelled, dressed ladder.

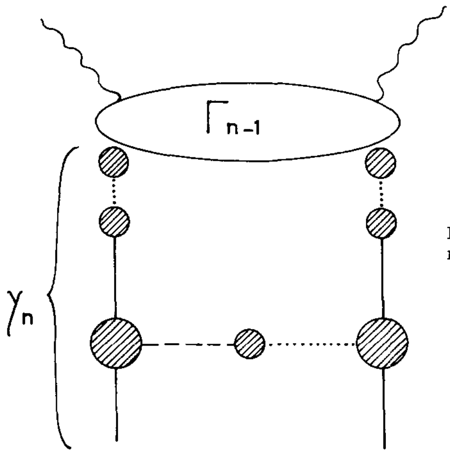

FIG. 24. Inductive representation of Γ_n.

integrals which appears very frequently in the above work. We have

$$I(m,n) = \int_0^1 \prod_{i=1}^{n} dx_i \prod_{j=1}^{m} d\rho_j \; \delta\left(\sum x_i + \sum \rho_j - 1\right) \times \prod_{j=1}^{m}\left(-\frac{\partial}{\partial \rho_j}\right)\left(\frac{H_0(\rho,x)}{(\sum x_i + \sum a_j\rho_j)^n}\right), \tag{B1}$$

where H_0 is homogeneous of degree zero.

We shall show that we can just remove all the integrations and differentiations with respect to the ρ's and set all remaining ρ's to zero.

Consider the expression

$$\frac{1}{\Gamma(n)}\int_0^\infty \prod^{n} d\tilde{x}_i \prod^{m}\left[d\tilde{\rho}_j\left(-\frac{\partial}{\partial\tilde{\rho}_j}\right)\right] \{H_0(\tilde{\rho},\tilde{x}) \times \exp[-(\sum \tilde{x}_i + \sum a_j\tilde{\rho}_j)]\} \tag{B2}$$

If we scale by

$$\tilde{x}_i = ux_i, \quad \tilde{\rho}_j = u\rho_j, \tag{B3}$$

(B2) reduces to

$$\frac{1}{\Gamma(n)}\int \prod^{n} dx_i \prod^{m} d\rho_j \; \delta(\sum x_i + \sum \rho_j - 1) \prod^{m}\left(-\frac{\partial}{\partial\rho_j}\right) \times \left(\frac{H_0(\rho,x)}{(\sum x_i + \sum a_j\rho_j)^n}\int_0^\infty du\, u^{n-1} \times (\sum x_i + \sum a_j\rho_j)^n \exp[-u(\sum x_i + \sum a_j\rho_j)]\right) = I(m,n) \tag{B4}$$

as we see upon performing the n integration. But in expression (B2) we can immediately perform the $\tilde{\rho}$ integrations. This reduces it to the form we would obtain by transforming $I(O,n)$. So we have proved

$$I(m,n) = I(O,n). \tag{B5}$$

That is,

$$I(m,n) = \frac{1}{\Gamma(n)}\int_0^\infty \prod^{n} d\tilde{x}_i\, H_0(O,\tilde{x}) \exp(-\sum \tilde{x}_i) \tag{B6}$$

$$= \int_0^1 \prod dx_i\, \delta(\sum x_i - 1)\, \frac{(H_0(O,x)}{(\sum x_i)^n}. \tag{B7}$$

APPENDIX C: PARAMETRIC FUNCTIONS FOR CANCELLED, DRESSED LADDERS

We want to investigate the scaling properties of g/E, the coefficient of $2m\nu$ in F/E, with respect to the scaling parameters σ_i, defined in (5.5). The terms in E and g that determine the scaling behavior are those that scale least. These are the ones we need to calculate. We shall use the rules given in Ref. 7 to calculate E and g. E and F are the C and D functions for the cancelled diagrams of Fig. 23, where the blobs represent cancelled OPI renormalization parts. So to find E we must consider sets of cuts that will just leave Γ_n, the diagram of Fig. 23, connected. We can represent Γ_n as in Fig. 24 so that it is built up from γ_n and Γ_{n-1}. We see that we may either leave Γ_{n-1} connected and divide γ_n in two or vice versa to obtain C_{Γ_n}. But to divide Γ_{n-1} requires one more cut than leaving it just connected, which implies one more parameter scaling as σ_{n-1}. So such terms may be neglected, and we may write

$$C_{\Gamma_n} = C_{\Gamma_{n-1}} \prod_{rn} E_{rn}\left(\alpha_n + \beta_n + \beta_n' + \sum_{rn}\frac{d_{rn}}{E_{rn}}\right) + O(\sigma_{n-1}), \tag{C1}$$

where d_{rn} represents some terms of F_{rn}. Repeating this procedure n times, we obtain

$$E = C_{\Gamma_n} = \prod_{r} E_r \prod_{i=1}^{n}\left(\alpha_i + \beta_i + \beta_i' + \sum_{ri}\frac{d_{ri}}{E_{ri}}\right) \tag{C2}$$

$$+ \text{ higher orders in } \sigma.$$

Consider now the coefficient of $2p\cdot q$ in D_{Γ_n}, λ_n, say. This is obtained by partitioning Γ_n so that the incoming momentum in one part is $p + q$ (see Fig. 25). So, in the same way as before, we obtain

$$\lambda_n = \lambda_{n-1}\prod_{rn} E_{rn}\left(\alpha_n + \sum_{rn}\frac{d'_{rn}}{E_{rn}}\right), \tag{C3}$$

whence

$$\lambda_n = \prod_{r} E_r\left(\alpha_0 + \sum_{r0}\frac{d'_{r0}}{E_{r0}}\right) \times \prod_{i=1}\left(\alpha_i + \sum_{ri}\frac{d'_{ri}}{E_{ri}}\right) + \text{ higher orders in } \sigma. \tag{C4}$$

The coefficient of q^2, κ_n arises from three types of partition (Fig. 26). We obtain

$$\kappa_n = \kappa_{n-1} \prod_{rn} E_{rn}\left(\alpha_n + \beta_n + \beta'_n + \sum_{rn} \frac{d_{rn}}{E_{rn}}\right), \tag{C5}$$

whence

$$\kappa_n = \prod_r E_r \left(\alpha_0 + \sum_{r0} \frac{d_{r0}}{E_{r0}}\right) \prod_{i=1}^{n}\left(\alpha_i + \beta_i + \beta'_i + \sum_{ri} \frac{d_{ri}}{E_{ri}}\right)$$

$$+ \text{ highest orders in } \sigma. \tag{C6}$$

Putting the results (C2), (C4), and (C6) together, we can write

$$\frac{g}{E} = \left(\alpha_0 + \sum_{r0} \frac{d'_{r0}}{E_{r0}}\right)\prod_{i=1}^{n} \frac{(\alpha_i + \sum_{ri} d'_{ri}/E_{ri})}{(\alpha_i + \beta_i + \beta'_i + \sum_{ri} d_{ri}/E_{ri})} - \frac{1}{\omega}\left(\alpha_0 + \sum_{r0} \frac{d_{r0}}{E_{r0}}\right)$$

$$+ \text{ higher orders} \tag{C7}$$

or

$$\bar{g}/\bar{E} = (\bar{\alpha}_0 + \sum_{r0} a'_{r0}\bar{\rho}_{r0}) \prod_{i=1}^{n} \frac{\bar{\alpha}_i + \sum_{ri} a'_{ri}\bar{\rho}_{ri}}{\alpha_i + \beta_i + \beta'_i + \sum_{ri} a_{ri}\bar{\rho}_{ri}} + \frac{1}{\omega}(\bar{\alpha}_0 + \sum_{r0} a_{r0}\bar{\rho}_{r0}), \tag{C8}$$

where the a's are functions of the internal variables of the renormalization parts.

We remark that setting all the $\bar{\rho}$'s to zero gives the result for a bare ladder

$$\bar{\alpha}_0 \prod_{i=1}^{n}\left(\frac{\bar{\alpha}_i}{\bar{\alpha}_i + \bar{\beta}_i + \bar{\beta}'_i} - \frac{1}{\omega}\right). \tag{C9}$$

APPENDIX D

We can use equations (7.5)–(7.8) to write the numerator (7.4) as a sum of inner products of the momenta p, q and the displaced loop momenta r'_j. Consider one term in such an expansion. The r'_j's must occur in pairs—we shall see that terms containing four or six or more are negligible. Generally the r'_j's will not occur in the form r'^2_j but each will be contracted with another vector. The symmetrization then will contract these vectors together. So we obtain chains of vectors held together by symmetrization and inner products.

We find that the dominant terms are characterized by the disposition of their chains on the ladder. We call a segment those lines of a ladder with the same index and we say that one segment is higher than another if it has a lower index. Now, if in a particular chain we can replace a pair of loop momenta r'_j by another pair with a lower index then the first chain is negligible. This means, from (7.8), that the index of such a pair must be the same as the index of the lower of the segments from which the two momenta are taken. Because of the vector nature of the coupling this means that terms quartic in r'_j are negligible, since if one vector in a segment couples to a higher vector the other must couple to a lower. We can show that dominant chains are of the form either r'^2_j with both terms from the same segment or are monotonic, with a q at the top and a p at the bottom. By monotonic we mean that every succeeding vector in the chain comes from a segment not higher than its predecessor. We can see this by breaking up any nonmonotonic chain into monotonic sub-chains and substituting q's and p's at their ends. This procedure gives a permissible term that dominates the original one.

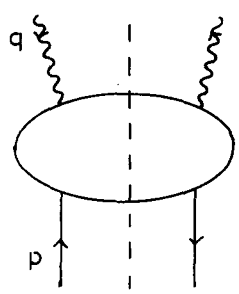

FIG. 25. Partition of Γ_n to give λ_n.

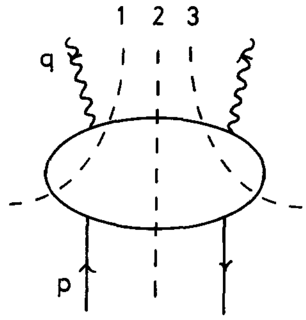

FIG. 26. Partitions for κ_n.

Taking account of the above requirements the important factors from a segment with a gluon rung are

$$2\left(\prod_j^N \frac{\alpha_k}{\sum_k} p + r'_j\right)_{\alpha_{j-1}} \left[r'_j + \frac{\alpha_j}{\sum_j} r'_{j+1} + \cdots - \left(\prod_0^{j-1} \alpha_k / \prod_1^j \textstyle\sum_k\right) q\right]^{\alpha_j}$$

$$- g_{\alpha_{j-1}}{}^{\alpha_j}\left(\frac{\alpha_0}{\sum_j} \prod_1^N \frac{\alpha_k}{\sum_k} 2p\cdot q + r'_j \cdot \right.$$

$$\left. \times \left(r'_j + 2\frac{\alpha_j}{\sum_j} r'_{j+1} + \cdots + \prod_j^N \frac{\alpha_k}{\sum_k}\right)p\right). \tag{D1}$$

Symmetrizing and remembering that this will be coupled to similar terms on both sides, we see that everything cancels except the terms quadratic in r'_j, as we asserted in Sec. 7.

We cannot eliminate nearly so many terms from the special segments with fermion rungs and chains linking several such segments are possible. We can, however, still write the contribution of individual terms fairly simply. Clearly the contribution to a particular term from the special segments is proportional to

$$(p\cdot q)^i \prod_k r'^2_k \tag{D2}$$

after symmetrization, where the k's index some of these special segments. Consider the following groupings of the vectors from the special segments. Start with a q factor (omitting the q in the top line). It may come from a D^{β}_{i+1} term in which case we choose for the second vector the vector indexed by α in the term $C_{i\alpha}$ corresponding to the segment above. Alternatively it may come from $C_{i\alpha}$. In this case it must come from the inner product, and for the second vector we choose the vector to which it is coupled. If, in either case, the second vector is a p, then the group is complete and we go on to the next q. If it is r_j', we go on to the partner r_j' and repeat the above procedure. In this way we arrange all the vectors into groups corresponding to particular q's or to individual r_j' pairs from the same segment.

Now, multiplying the coefficients together, we find that we have corresponding to each q a factor

$$\frac{\alpha_0}{\sum_j} \prod_1^N \frac{\alpha_k}{\sum_k}, \tag{D3}$$

to each $r_j'^2$ nothing but a constant factor, and for each three-fermion segment we may or may not have a factor proportional to $\alpha_i/\sum_i$, depending on whether a displacement term is taken from the vector $r_{i+1\,0}'$ in (7.7).

For the case when we have no cross linking except in $C_{i\alpha} D_{i+1}^{\beta}$, we have explicitly written out the coefficients in (7.9).

[1]S. J. Chang and P. M. Fishbane, Phys. Rev. Lett. **24**, 847 (1970) and Phys. Rev. D **2**, 1084 (1970).
[2]T. K. Gaisser and J. C. Polkinghorne, Nuovo Cimento A **1**, 501 (1971), hereafter referred to as GP.
[3]K. G. Wilson, Phys. Rev. **179**, 1499 (1969).
[4]L. D. Landau, A. A. Abrikosov, and I. M. Khalatnikov, Nuovo Cimento Suppl. **3**, 80 (1956).
[5]J. C. Taylor, Proc. R. Soc. A **234**, 296 (1956).
[6]T. Appelquist and J. R. Primack, Phys. Rev. D **1**, 1144 (1970).
[7]R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The Analytic S-Matrix* (Cambridge U.P., Cambridge, 1966).
[8]K. Hepp, Commun. Math. Phys. **2**, 301 (1966).
[9]J. C. Polkinghorne, Nuovo Cimento A **68**, 441 (1970).
[10]J. C. Ward, Phys. Rev. **84**, 897 (1951).
[11]A. Salam, Phys. Rev. **82**, 217 (1951).
[12]A. Salam, Phys. Rev. **79**, 910 (1950).
[13]This came to light upon comparison of our result with that of Gribov and Lipatov, Phys. Lett. B **37**, 78 (1971) and Leningrad preprint.
[14]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.
[15]K. G. Wilson, Trieste Preprint IC/71/47.

Unitary representations of the $SL(2, C)$ group in horospheric basis

M. Huszár

Central Research Institute for Physics, Budapest, 114 P.O. Box 49, Hungary

(Received 10 April 1973)

The matrix elements of unitary representations of the $SL(2, C)$ group are derived in a basis defined by two-dimensional momenta corresponding to the horospheric subgroup $\left(\begin{smallmatrix}1 & \beta\\ 0 & 1\end{smallmatrix}\right)$. A parametrization fitting the above basis well is chosen on the analogy of the rotation group by sandwiching a complex rotation about the y axis with two horospheric translations. In this way the two outer subgroups can be factored out immediately in terms of plane waves which are counterparts to the exponentials formed by the Euler angles φ and ψ occuring in the rotation group. Finally, matrix elements of complex rotations about y axis are found by solving the simultaneous eigenvalue problem of the two Casimir operators. Unitary representations obtained in this way are expressed in a rather simple form in terms of Bessel functions.

INTRODUCTION

After the work of Bargmann,[1] Naimark,[2] and Gel'fand *et al.*[3,4] it appeared that all fundamental questions relating to the representation theory of the $SL(2,C)$ group could be considered as closed. Yet, when investigations concerning relativistic expansions of the scattering amplitude raised the problem of finding an explicit form for unitary representations, considerable difficulties were encountered. Results obtained for the representations in angular momentum basis were expressible only in terms of multiple sums over complicated expressions, and bases corresponding to the reductions $SL(2,C) \supset O(2,1) \supset O(2)$ [or $O(1,1)$] and $SL(2,C) \supset E(2) \supset O(2)$ led to similarly complicated formulas. (A far from complete list of references is given by Refs. 5–14.) It would appear, then, that although the use of the above bases is supported by the interpretation of the scattering amplitude as a function of the little group of the Poincaré group, there is still great importance to be attached to the investigation of the matrix elements of unitary representations in other bases.

Since the finite-dimensional representations are very simple in spinor basis, it is not surprising that a considerable simplification of unitary representations can be achieved by introducing unitary spinors. The structure of matrix elements in this basis is analogous to that of the D functions of the real rotation group: Two of the Euler angles, φ and ψ, and their complex conjugates appear in the form of exponential factors, while the dependence on ϑ and its complex conjugate ϑ^* is contained by $d^j_{mn}(\vartheta.\vartheta^*)$, a complex analog of the familiar $d^j_{mn}(\vartheta)$ functions.[15–17] The deeper reason for this analogy lies in the fact that the proper Lorentz group is isomorphic to $SO(3,C)$, the connected part of the group of motions of the two-dimensional complex sphere $S_1^2 + S_2^2 + S_3^2 = S^2$.[18] The labels of unitary spinors can be obtained as the eigenvalues of Casimir operators of $SO(2,C) = \left(\begin{smallmatrix}\alpha & 0\\ 0 & \alpha^{-1}\end{smallmatrix}\right)$, the little group of a certain fixed point of the complex sphere.

In the present paper it is shown that the so-called horospheric basis offers an opportunity for a still further simplification of matrix elements of unitary representations. To this end the subgroup of the $SL(2,C)$ group of the form $\left(\begin{smallmatrix}1 & \beta\\ 0 & 1\end{smallmatrix}\right)$ will be considered, where β ranges over the whole complex plane. This subgroup, which is isomorphic to $T(2)$, the real translation group in two dimensions, plays an important role in representation theory of the $SL(2,C)$ group.[4] Orbits that are described by it or by any conjugate subgroup in any space homogeneous under the $SL(2,C)$ group, are the horospheres. In particular, in the group space itself the horospheres are straight generators of the surface $\alpha\delta - \gamma\beta = 1$, which explains their central role in representation theory of the $SL(2,C)$ group.

The horospheric subgroup can, in fact, be obtained from the above spinor subgroup by means of a group contraction.[19] It can be shown that the little group of a certain point on the complex sphere $S_1^2 + S_2^2 + S_3^2 = S^2$ is $SO(2,C)$ or $T(2)$, according to whether $S \neq 0$ or $S = 0$.[20] Consequently, if the little group of a point on a complex sphere of radius going to zero is considered, in the limit the little group $SO(2,C)$ contracts into the $T(2)$ group.

That means that the unitary representations of the $SL(2,C)$ group in horospheric basis can be obtained from those in unitary spinor basis by means of a contraction.[19] Actually, the explicit form of representations in horospheric basis has been found in this way, since we could not find the solution of the eigenvalue equations for the representations [Cf. Eq. (2.6)]. Having obtained the representations in this way it was easy to verify that they satisfy the related eigenvalue equations. In order not to encumber the approach we have omitted the description of the contraction procedure, even though it offers some insight into the structure of the matrix elements obtained.

The use of the horospheric group isomorphic to the translation group makes natural the labeling of the basis by means of a two-dimensional momentum in addition to the eigenvalues of the Casimir operators of the $SL(2,C)$ group. This basis is not new in the literature, since it is used as soon as the ordinary Fourier transform of functions on the familiar z space is taken.[2–4,10] Furthermore, in Ref. 9 it was utilized to derive the matrix elements of boosts along the third axis, while certain applications to hadron physics have also been touched on.[21]

In what follows only the principal series of unitary representations will be treated. This is sufficient from the point of view of the harmonic analysis of functions on the group, since any square integrable function can be expanded in terms of representations of the principal series; the supplementary series does not give any contribution.[2–4]

Owing to the good fit of the parametrization to the chosen basis, four of the six parameters of the $SL(2,C)$ group can be factored out immediately in terms of "plane waves." The function containing the two remaining parameters satisfies two formally independent differential equations, which have a regular solution if, and only if, j_0, the label characterizing irreducible representa-

Copyright © 1973 by the American Institute of Physics

tions, takes integer or half-integer values. The solution, finally, can be expressed in a simple form in terms of Bessel functions of imaginary argument. The representations obtained seem to have a simpler form than those in any other basis.

1. PARAMETRIZATION AND BASIS

The spinors $\xi = (u, v)$ form a homogeneous space under the $SL(2,C)$ group, provided the point $(u, v) = (0,0)$ is excluded. The action of $g \in SL(2,C)$ can be given in the form $\xi' = g\xi$, i.e.,

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \quad (\alpha\delta - \gamma\beta = 1).$$

Let us fix now a standard point in the spinor space, say, $\xi_0 = \binom{1}{0}$. The little group of this point [i.e., the subgroup of the $SL(2,C)$ group satisfying the condition $h\xi_0 = \xi_0$] is

$$h = \begin{pmatrix} 1 & -i\psi \\ 0 & 1 \end{pmatrix}, \tag{1.1}$$

where ψ ranges over the whole complex plane. The factor $(-i)$ is introduced only for later convenience. The subgroup (1.1) isomorphic to the real translation group in two dimensions is known as the horospheric subgroup.[4]

The translation property of this group can be demonstrated on the linear fractional mapping

$$z' = (\alpha z + \beta)/(\gamma z + \delta),$$

where $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in SL(2,C)$. In particular, the action of the horospheric subgroup (1.1) is the displacement

$$z' = z - i\psi.$$

Since (1.1) satisfies the condition $h\xi_0 = \xi_0$, each element of $SL(2,C)$ can be decomposed as a horospheric translation followed by a motion in the spinor space.

An appropriate parametrization of spinors for our purposes is

$$u = \cos(\vartheta/2) - i\varphi\sin(\vartheta/2), \qquad v = \sin(\vartheta/2), \tag{1.2}$$

where both φ and $\sin(\vartheta/2)$ range over the whole complex plane. This parametrization applies for each (u, v) except for the singular value $v = 0$. It should be noted that the functions $\varphi(z) = \varphi(u/v)$ used by Gel'fand *et al.* (cf. Ref. 4, p. 142) are not defined for the same value of v either.

Starting from the standard spinor $\xi_0 = \binom{1}{0}$, the above parametrization can be also obtained with the aid of the following $SL(2,C)$ transformations:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & -i\varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\vartheta/2) & -\sin(\vartheta/2) \\ \sin(\vartheta/2) & \cos(\vartheta/2) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{1.3}$$

The subgroup (1.1) parametrizes the little group of ξ_0, while (1.3) parametrizes the left cosets of $SL(2,C)$ with respect to the subgroup (1.1). Every point $\xi = (u, v)$ of the spinor space represents one left coset, while each element of the little group (1.1) characterizes the elements within a coset. This leads finally to the following decomposition of the $SL(2,C)$ group:

$$g = \begin{pmatrix} 1 & -i\varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\vartheta/2) & \sin(\vartheta/2) \\ \sin(\vartheta/2) & \cos(\vartheta/2) \end{pmatrix} \begin{pmatrix} 1 & -i\psi \\ 0 & 1 \end{pmatrix}, \tag{1.4}$$

where the complex parameters

$$\varphi = \varphi_1 + i\varphi_2, \quad \vartheta = \vartheta_1 + i\vartheta_2, \quad \psi = \psi_1 + i\psi_2$$

range over the limits

$$-\infty < \varphi_1, \varphi_2, \psi_1, \psi_2, \vartheta_2 < \infty, \quad -\pi \leq \vartheta_1 < \pi. \tag{1.5}$$

The matrix in the middle of the decomposition (1.4) describes a complex rotation about y axis by an angle $\vartheta = \vartheta_1 + i\vartheta_2$ which can be verified to be equivalent to a rotation by an angle ϑ_1 about y axis followed by a boost by a hyperbolic angle $-\vartheta_2$ along the same axis.

As was mentioned in the Introduction, the horospheric group $\left(\begin{smallmatrix} 1 & -i\psi \\ 0 & 1 \end{smallmatrix}\right)$ can be obtained from the group of complex rotations about the third axis $\left(\begin{smallmatrix} e^{-i\psi/2} & 0 \\ 0 & e^{+i\psi/2} \end{smallmatrix}\right)$ by means of a contraction. This is the reason for keeping the notation usual for the Euler angles, even though neither φ nor ψ is an angular variable, since neither their real nor the imaginary parts are cyclic.

Comparing Eq. (1.4) with the remark on the singular case $v = 0$, we conclude that each $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) \in SL(2,C)$ can be decomposed in the form (1.4), provided $\gamma \neq 0$. For the sake of completeness, a short comment ought to be made on the singular case $\gamma = 0$. Since the little group of the point ξ_0 has been fully parametrized, it is sufficient for the parametrization of the singular case $\gamma = 0$ to parametrize the $\binom{u}{0}$ spinors starting from the standard spinor $\binom{1}{0}$. Clearly, this can be accomplished by means of the subgroup $\left(\begin{smallmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{smallmatrix}\right)$ $(\alpha \neq 0)$ describing dilatations of ξ_0. Thus any $g \in SL(2,C)$ for which $\gamma = 0$ can be parametrized as $g(\alpha, \psi_0) = \left(\begin{smallmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{smallmatrix}\right) \left(\begin{smallmatrix} 1 & -i\psi_0 \\ 0 & 1 \end{smallmatrix}\right)$.

These values of the parameters can be approached by a one-parametric manifold of φ, ϑ, ψ given by (1.4). To this end let

$$\sin(\vartheta/2) = t, \quad i\varphi = (1-\alpha)/t, \quad i\psi = (1-\alpha^{-1})/t + i\psi_0.$$

Then g, as given by (1.4) approaches $g(\alpha, \psi_0)$ as t is going to zero. Therefore, the unitary representations corresponding to $g(\alpha, \psi_0)$ can be evaluated in two ways: either directly, in terms of α, ψ_0, or via φ, ϑ, ψ by approaching $t \to 0$. The consistency is proved by the fact that representations obtained in these two ways can be shown to coincide.

Denoting by M_k and N_k $(k = 1,2,3)$ the generators of rotations about, and boosts along, the kth axis, the linear combinations

$$J_k = \tfrac{1}{2}(M_k + iN_k), \quad K_k = \tfrac{1}{2}(M_k - iN_k)$$

satisfy the Lie algebra of two independent angular momenta

$$\begin{aligned} &[J_k, J_l] = i\epsilon_{klm}J_m, \quad [K_k, K_l] = i\epsilon_{klm}K_m, \\ &[J_k, K_l] = 0 \quad (k, l, m = 1,2,3). \end{aligned} \tag{1.6}$$

In what follows, the usual notation

$$J_\pm = J_1 \pm iJ_2, \quad K_\pm = K_1 \pm iK_2 \tag{1.7}$$

will also be used. Since the elements of a real Lie algebra have been multiplied by complex numbers, the corresponding parameters fail to remain real. The restriction imposed on the parameters can be obtained from the requirement that for any Lie group, the bilinear form $\epsilon^A X_A$ (where ϵ^A are the parameters of the group in question and X_A are the corresponding generators) be invariant under different parametrizations and under different choices of the basis in the algebra. It follows that if the parameters corresponding to J_+ and J_2 are denoted by $J_+ \to \varphi$, $J_2 \to \vartheta$, the parameters corresponding to K_- and K_2 will be their complex conjugates $K_- \to \varphi^*$, $K_2 \to \vartheta^*$.

Since infinitesimal generators of the horospheric subgroup (1.1) can be given as $M_1 - N_2$ and $M_2 + N_1$, or, equivalently, as J_+ and K_-, the unitary representations of the group can be written, in accordance with the decomposition (1.4), in the symbolic form

$$\begin{aligned} T(g) &= e^{-i\varphi J_+} e^{-i\varphi^* K_-} (e^{-i\vartheta J_2} e^{-i\vartheta^* K_2}) e^{-i\psi J_+} e^{-i\psi^* K_-} \\ &= e^{-i\varphi_1(M_1-N_2)} e^{i\varphi_2(M_2+N_1)} (e^{-i\vartheta_1 M_2} e^{i\vartheta_2 N_2}) \\ &\quad \times e^{-i\psi_1(M_1-N_2)} e^{i\psi_2(M_2+N_1)}. \end{aligned} \tag{1.8}$$

It is easy to recognize the analogy to the Euler decomposition of the rotation group, where a rotation about y axis is sandwiched by two rotations about z axis. In the present case the rotation about y axis is a complex one, and the role of the subgroup of rotations about z axis plays the horospheric subgroup.

Unitarity imposes on the generators the condition

$$(J_\pm)^\dagger = K_\mp, \quad J_2^\dagger = K_2.$$

Irreducible representations of the $SL(2,C)$ group are characterized by j_0 taking integer and half-integer values, and by a complex number σ taking real values for the principal series of unitary representations. Eigenvalues of the Casimir operators

$$\mathbf{J}^2 = J_1^2 + J_2^2 + J_3^2, \quad \mathbf{K}^2 = K_1^2 + K_2^2 + K_3^2 \tag{1.9}$$

are conveniently written in the form $\mathbf{J}^2 \to j(j+1)$, $\mathbf{K}^2 \to k(k+1)$, where j and k are related to j_0 and σ for the principal series by

$$j = \tfrac{1}{2}(j_0 - 1 + i\sigma), \quad k = -j^* - 1 = \tfrac{1}{2}(-j_0 - 1 + i\sigma). \tag{1.10}$$

On the basis of the decomposition (1.8) it is natural to reduce irreducible representations, characterized by the above $\chi = (j,k)$, according to representations of the horospheric subgroup. This group is Abelian, so that its representations are exponential functions

$$T^{PP^*}(h) = e^{-i(\psi P + \psi^* P^*)} = e^{-i(P_1\psi_1 - P_2\psi_2)}, \tag{1.11}$$

where P is a complex momentum

$$P = \tfrac{1}{2}(P_1 + iP_2), \quad P^* = \tfrac{1}{2}(P_1 - iP_2) \tag{1.12}$$

with continuous spectrum. In what follows the two-dimensional vector notation $\mathbf{P} = (P_1, P_2)$ will also be used.

Reduction of irreducible representations of the group according to those of (1.11) is an ordinary Fourier expansion. The basis in which the unitary representations of $SL(2,C)$ will be given is labelled by the above momenta:

$$J_+ \phi^{jk}_{PP^*} = P\phi^{jk}_{PP^*}, \quad K_- \phi^{jk}_{PP^*} = P^* \phi^{jk}_{PP^*}. \tag{1.13}$$

Unitary irreducible representations can be realized on functions $\varphi(z,z^*)$ of D^χ, the space of infinitely differentiable functions with infinitely differentiable inversion $\hat{\varphi}(z,z^*) \equiv z^{2j} z^{*2k} \varphi(-1/z, -1/z^*)$.[4] If representations are defined by left displacement, the action of representations on the functions $\varphi(z) \equiv \varphi(z,z^*) \in D^\chi$ will be

$$\begin{aligned} T^\chi(g)\varphi(z) &= (-\gamma z + \alpha)^{2j}(-\gamma^* z^* + \alpha^*)^{2k} \\ &\quad \times \varphi((\delta z - \beta)/(-\gamma z + \alpha)) \end{aligned} \tag{1.14}$$

In this space, the generators J_+, K_- take the form of differential operators $J_+ = -\partial/\partial z$, $K_- = \partial/\partial z^*$. Therefore, the solution of Eq. (1.13) can be written in the form of the "plane wave"

$$\phi_{PP^*}(z) = (1/2\pi)\, e^{-Pz + P^* z^*}. \tag{1.15}$$

This is obviously not an element of the space D^χ, nor even, due to the continuous spectrum of $\mathbf{P}$, of a Hilbert space including the space D^χ. Hence to treat the representations rigorously one has to proceed as follows.[22] We consider the $\phi_{PP^*}(z)$ functions given by (1.15) as functionals on the $\phi(z)$ functions

$$\Psi(P,P^*) = (\phi, \varphi) = \int d^2z\, \phi_{PP^*}(z)\varphi(z) \quad (d^2z = d\,\mathrm{Re}z d\,\mathrm{Im}z) \tag{1.16}$$

The value of the functional depends then on the parameters P, P^*, which results in the above $\Psi(P,P^*)$ function. We define the action of the representations T^χ_g on $\Psi(P,P^*)$ as

$$\begin{aligned} T^\chi_g \Psi(P,P^*) &= \int d^2z\, \phi_{PP^*}(z)(\gamma z + \delta)^{-2j-2} \\ &\quad \times (\gamma^* z^* + \delta^*)^{-2k-2}\varphi((\alpha z + \beta)/(\gamma z + \delta)). \end{aligned} \tag{1.17}$$

Representation given by this formula is the same as that in Eq. (1.14), but expressed in terms of generalized functions. Really, by substituting $(\alpha z + \beta)/(\gamma z + \delta) \to z$ in (1.17) we get

$$\begin{aligned} T^\chi_g \Psi(P,P^*) &= \int d^2z\, (-\gamma z + \alpha)^{2j}(-\gamma^* z^* + \alpha^*)^{2k} \\ &\quad \times \phi_{PP^*}((\delta z - \beta)/(-\gamma z + \alpha))\varphi(z). \end{aligned} \tag{1.18}$$

Expressing now $\varphi(z)$ from (1.16), this equation assumes the form

$$\begin{aligned} T^\chi_g \Psi(P,P^*) &= (2\pi)^{-2} \int d^2z\, e^{-Pz' + P^* z'^*}(-\gamma z + \alpha)^{2j} \\ &\quad \times (-\gamma^* z^* + \alpha^*)^{2k} \int d^2Q\, e^{Qz - Q^* z^*}\Psi(Q,Q^*), \end{aligned} \tag{1.19}$$

where $z' = (\delta z - \beta)/(-\gamma z + \alpha)$ and $d^2Q = dQ_1 dQ_2$, $Q = \frac{1}{2}(Q_1 + iQ_2)$. Thus the "matrix elements" of the representation T^χ_g can be interpreted as the kernel of the integral transformation (1.19), provided the interchange of the order of integrations is legitimate.[23] That is

$$T^\chi_g \Psi(P,P^*) = \int d^2Q\, T^j_{QP}(g)\Psi(Q,Q^*)$$

with

$$\begin{aligned} T^j_{QP}(g) &= (2\pi)^{-2} \int d^2z\, e^{Qz - Q^* z^*}(-\gamma z + \alpha)^{2j} \\ &\quad \times (-\gamma^* z^* + \alpha^*)^{2k} e^{-Pz' + P^* z'^*}, \end{aligned} \tag{1.20}$$

where again $z' = (\delta z - \beta)/(-\gamma z + \alpha)$. The evaluation of this integral is not, however, the best way of deriving the matrix elements, and so we shall proceed in another manner.

2. MATRIX ELEMENTS OF REPRESENTATIONS

By making use of the parameters introduced above, we can represent infinitesimal generators in terms of differential operators on the group and, by solving the eigenvalue equations of the Casimir operators, obtain the matrix elements of the representations. Regularity requirements will yield automatically the quantization of j_0.

Let $f(g)$ be a function on the $SL(2,C)$ group. Then the action of the representation $T(g_0)$ on $f(g)$ can be defined by left displacement:

$$T(g_0)f(g) = f(g_0^{-1}g).$$

From this the following infinitesimal generators are derived:

$$J_+ = \frac{1}{i}\,\frac{\partial}{\partial\varphi},$$

$$J_- = \frac{1}{i}\left((1+\varphi^2+2i\varphi\cot\vartheta)\frac{\partial}{\partial\varphi} - 2i\left(1 - i\varphi\,\frac{1-\cos\vartheta}{\sin\vartheta}\right)\right.$$
$$\left.\times\frac{\partial}{\partial\vartheta} - 2i\varphi\,\frac{1}{\sin\vartheta}\,\frac{\partial}{\partial\psi}\right)$$

$$J_3 = \frac{1}{i}\left((\cot\vartheta - i\varphi)\frac{\partial}{\partial\varphi} + i\,\frac{1-\cos\vartheta}{\sin\vartheta}\,\frac{\partial}{\partial\vartheta} - \frac{1}{\sin\vartheta}\,\frac{\partial}{\partial\psi}\right),$$

$$K_+ = \frac{1}{i}\left((1+\varphi^{*2} - 2i\varphi^*\cot\vartheta^*)\frac{\partial}{\partial\varphi^*}\right.$$
$$\left. + 2i\left(1 + i\varphi^*\,\frac{1-\cos\vartheta^*}{\sin\vartheta^*}\right)\frac{\partial}{\partial\vartheta^*} + 2i\varphi^*\,\frac{1}{\sin\vartheta^*}\,\frac{\partial}{\partial\psi^*}\right),\tag{2.1}$$

$$K_- = \frac{1}{i}\,\frac{\partial}{\partial\varphi^*},$$

$$K_3 = \frac{1}{i}\left((\cot\vartheta^* + i\varphi^*)\frac{\partial}{\partial\varphi^*} - i\,\frac{1-\cos\vartheta^*}{\sin\vartheta^*}\,\frac{\partial}{\partial\vartheta^*} - \frac{1}{\sin\vartheta^*}\,\frac{\partial}{\partial\psi^*}\right)$$

The matrix elements $T^j_{QP}(g)$ of the representations are solutions of the eigenvalue equations

$$[\mathbf{J}^2 - j(j+1)]T^j_{QP}(g) = 0, \qquad [\mathbf{K}^2 - k(k+1)]T^j_{QP}(g) = 0.\tag{2.2}$$

By taking into account Eqs. (1.9) and (2.1), these equations can be written

$$\left[\tan^2\frac{\vartheta}{2}\,\frac{\partial^2}{\partial\vartheta^2} + \frac{1}{2}\tan\frac{\vartheta}{2}\left(3+\tan^2\frac{\vartheta}{2}\right)\frac{\partial}{\partial\vartheta} + \frac{i}{\cos^2(\vartheta/2)}\right.$$
$$\times\left(\frac{\partial}{\partial\varphi} + \frac{\partial}{\partial\psi}\right)\frac{\partial}{\partial\vartheta} - \frac{1}{\sin^2\vartheta}\left(\frac{\partial^2}{\partial\varphi^2} + \frac{\partial^2}{\partial\psi^2} - 2\cos\vartheta\,\frac{\partial^2}{\partial\varphi\partial\psi}\right)$$
$$\left. + \frac{i}{4\sin(\vartheta/2)\cos^3(\vartheta/2)}\left(\frac{\partial}{\partial\varphi} + \frac{\partial}{\partial\psi}\right) - j(j+1)\right]T^j_{QP}(g) = 0,\tag{2.3}$$

$$\left[\tan^2\frac{\vartheta^*}{2}\,\frac{\partial^*}{\partial\vartheta^{*2}} + \frac{1}{2}\tan\frac{\vartheta^*}{2}\left(3+\tan^2\frac{\vartheta^*}{2}\right)\frac{\partial}{\partial\vartheta^*}\right.$$
$$- \frac{i}{\cos^2(\vartheta^*/2)}\left(\frac{\partial}{\partial\varphi^*} + \frac{\partial}{\partial\psi^*}\right)\frac{\partial}{\partial\vartheta^*} - \frac{1}{\sin^2\vartheta^*}$$
$$\times\left(\frac{\partial^2}{\partial\varphi^{*2}} + \frac{\partial^2}{\partial\psi^{*2}} - 2\cos\vartheta^*\,\frac{\partial^2}{\partial\varphi^*\partial\psi^*}\right) - \frac{i}{4\sin(\vartheta^*/2)\cos^3(\vartheta^*/2)}$$
$$\left.\times\left(\frac{\partial}{\partial\varphi^*} + \frac{\partial}{\partial\psi^*}\right) - k(k+1)\right]T^j_{QP}(g) = 0.\tag{2.4}$$

In accordance with Eqs. (1.8), (1.13), the variables separate as follows:

$$T^j_{QP}(g) = e^{-i(Q\varphi + Q^*\varphi^* + P\psi + P^*\psi^*)}\,d^j_{QP}(\vartheta,\vartheta^*).\tag{2.5}$$

Substituting this into (2.3) and (2.4), we get

$$\left\{\tan^2\frac{\vartheta}{2}\,\frac{\partial^2}{\partial\vartheta^2} + \left[\frac{Q+P}{\cos^2(\vartheta/2)} + \frac{1}{2}\tan\frac{\vartheta}{2}\left(3+\tan^2\frac{\vartheta}{2}\right)\right]\frac{\partial}{\partial\vartheta}\right.$$
$$\left. + \frac{Q^2+P^2-2QP\cos\vartheta}{\sin^2\vartheta} + \frac{Q+P}{4\sin(\vartheta/2)\cos^3(\vartheta/2)} - j(j+1)\right\}$$
$$\times\, d^j_{QP}(\vartheta,\vartheta^*) = 0,\tag{2.6}$$

$$\left\{\tan^2\frac{\vartheta^*}{2}\,\frac{\partial^2}{\partial\vartheta^{*2}} + \left[-\frac{Q^*+P^*}{\cos^2\vartheta^*/2} + \frac{1}{2}\tan\frac{\vartheta^*}{2}\left(3+\tan^2\frac{\vartheta^*}{2}\right)\right]\frac{\partial}{\partial\vartheta^*}\right.$$
$$+ \frac{Q^{*2}+P^{*2}-2Q^*P^*\cos\vartheta^*}{\sin^2\vartheta^*} - \frac{Q^*+P^*}{4\sin(\vartheta^*/2)\cos^3(\vartheta^*/2)}$$
$$\left. - k(k+1)\right\}d^j_{QP}(\vartheta,\vartheta^*) = 0.\tag{2.7}$$

If we now introduce the function $I(\vartheta,\vartheta^*)$ by

$$d^j_{QP}(\vartheta,\vartheta^*) = e^{(Q+P)\cot(\vartheta/2) - (Q^*+P^*)\cot(\vartheta^*/2)}$$
$$\times\, 1/[\sin(\vartheta/2)\,\sin(\vartheta^*/2)]\,I(\vartheta,\vartheta^*)\tag{2.8}$$

and the new variables

$$t = 2\sqrt{Q}\sqrt{P}/\sin(\vartheta/2), \qquad t^* = 2\sqrt{Q^*}\sqrt{P^*}/\sin(\vartheta^*/2)$$
$$(-\pi < \arg P, Q, P^*, Q^* < \pi),\tag{2.9}$$

then (2.6) and (2.7) reduce to the equations

$$\frac{d^2I}{dt^2} + \frac{1}{t}\,\frac{dI}{dt} - \left[1 + \frac{(2j+1)^2}{t^2}\right]I = 0,\tag{2.10}$$

$$\frac{d^2I}{dt^{*2}} + \frac{1}{t^*}\,\frac{dI}{dt^*} - \left[1 + \frac{(2k+1)^2}{t^{*2}}\right]I = 0\tag{2.11}$$

both of which are known to be differential equations for the Bessel function of imaginary argument. Two linearly independent solutions of (2.10) are $I_{2j+1}(t)$ and $I_{-2j-1}(t)$, where $I_\nu(t)$ is defined by the series[24]

$$I_\nu(t) = \sum_{m=0}^{\infty}\frac{(t/2)^{\nu+2m}}{m!\,\Gamma(m+\nu+1)}.\tag{2.12}$$

Equation (2.10) has singular points at $t=0$ and $t=\infty$ and accordingly the cut will be directed along the negative imaginary t axis while the phase will be fixed by $-\pi/2 < \arg t < 3\pi/2$. In a similar way, the solutions of (2.11) can be written in the form $I_{2k+1}(t^*)$ and $I_{-2k-1}(t^*)$. Here, it is convenient to cut the t^* plane along the positive imaginary axis and use the convention $\arg t^* = -\arg t$.

It is well known that if $2j+1$ takes integer values, the above solutions fail to remain linearly independent. We shall return to this singular case later on.

Taking into account that differentiation with respect to a complex variable does not act on a function of the complex conjugate variable, we see that the general solution of (2.10) and (2.11) assumes the form

$$I(\vartheta,\vartheta^*) = c_1 I_{2j+1}(t)I_{2k+1}(t^*) + c_2 I_{-2j-1}(t)I_{-2k-1}(t^*)$$
$$+ c_3 I_{2j+1}(t)I_{-2k-1}(t^*) + c_4 I_{-2j-1}(t)I_{2k+1}(t^*).\tag{2.13}$$

It is worth noting that the two eigenvalue equations of Casimir operators of the $SL(2,C)$ group are, generally speaking, intimately linked, and in reality the representations satisfy a rather complicated fourth-order differential equation.[5] In the present case, however, due to the favorable parametrization and basis, the differential equations in question decouple into two formally independent equations. It is this fact which enabled us to construct four linearly independent solutions of (2.3) and (2.4) by means of two linearly independent solutions of a second-order equation.

The constants c_1, c_2, c_3, c_4 should be determined from the regularity requirements imposed on solutions at the singular points $\sin(\vartheta/2) = \infty$ ($t = 0$) and $\sin(\vartheta/2) = 0$ ($t = \infty$). The series (2.12) provides the behavior of the d^j_{QP} functions at $t = 0$. By requiring finiteness of the representation $\chi = (j, k)$ as well as of the equivalent representation $-\chi = (-j-1, -k-1)$, we arrive at the conditions $c_3 = 0, c_4 = 0$.

Restrictions on the remaining two constants are obtained from the behavior at $t = \infty$. To this end we use the following asymptotic form of the $I_\nu(t)$ functions[24]:

$$I_\nu(t) \underset{|t|\to\infty}{\sim} (1/\sqrt{2\pi t})(e^t + ie^{-t+i\pi\nu}),$$

$$I_\nu(t^*) \underset{|t|\to\infty}{\sim} (1/\sqrt{2\pi t^*})(e^{t^*} - ie^{-t^*-i\pi\nu}) \quad (-\pi/2 < \arg t < 3\pi/2). \tag{2.14}$$

On substituting this into (2.13) we get two singular terms, which vanish if

$$c_1 + c_2 = 0, \quad e^{-2\pi i j_0} c_1 + e^{2\pi i j_0} c_2 = 0. \tag{2.15}$$

These equations have a nontrivial solution only if $\sin(2\pi j_0) = 0$ (i.e., $2j_0$ takes integer values).

With an appropriate choice of normalizing factor and phase, the functions d^j_{QP} take the form

$$d^j_{QP}(\vartheta, \vartheta^*) = \frac{1}{4\sin(2\pi j)}\left(\frac{P}{Q}\right)^{j+1/2}\left(\frac{P^*}{Q^*}\right)^{k+1/2}$$
$$\times \frac{e^{(Q+P)\cot(\vartheta/2) - (Q^*+P^*)\cot(\vartheta^*/2)}}{\sin(\vartheta/2)\sin(\vartheta^*/2)}$$
$$\times [I_{2j+1}(t)I_{2k+1}(t^*) - I_{-2j-1}(t)I_{-2k-1}(t^*)], \tag{2.16}$$

where

$$t = 2\sqrt{Q}\sqrt{P}/\sin(\vartheta/2).$$

The final form of representations is obtained by recalling (2.5),

$$T^j_{QP}(g) = e^{-i(Q\varphi + Q^*\varphi^* + P\psi + P^*\psi^*)} d^j_{QP}(\vartheta, \vartheta^*) \tag{2.17}$$

and taking this with (2.16). The normalization assures fulfillment of the condition

$$\lim_{g\to e} T^j_{QP}(g) = \delta(P_1 - Q_1)\delta(P_2 - Q_2), \tag{2.18}$$

where e is the unit element of the group, $e = (\varphi, \vartheta, \psi) = (0, 0, 0)$, and

$$P = \tfrac{1}{2}(P_1 + iP_2), \quad Q = \tfrac{1}{2}(Q_1 + iQ_2).$$

If $2j + 1$ takes integer values, the relation $I_{2j+1}(t) = I_{-2j-1}(t)$ holds and hence the bracket in (2.16) containing I functions vanishes. As, however, the factor $\sin 2\pi j$ becomes zero at the same points and in the same order, we do not have to worry about the linear dependence of solutions; the representation at these singular points should be understood as the limit when $2j + 1$ is going to an integer number.

Let us finally look briefly at some simple properties of representations. First the condition of unitarity,

$$d^j_{QP}(-\vartheta, -\vartheta^*) = d^j_{PQ}(\vartheta, \vartheta^*)^*,$$

can be inferred from (2.16) and (2.12).

Further, representations $\chi = (j, k)$ and $-\chi = (-j-1, -k-1)$ are known to be unitarily equivalent. In the present case they differ merely by a phase

$$d^{-j-1}_{QP}(\vartheta, \vartheta^*) = (Q/P)^{2j+1}(Q^*/P^*)^{2k+1} d^j_{QP}(\vartheta, \vartheta^*).$$

To define the Bessel functions unambiguously, it was necessary to cut the t and t^* planes and fix the arguments of the variables accordingly. However, it can be verified that the discontinuity across the cut vanishes. Let

$$t_+ = |t|e^{-i(\pi/2-0)}, \quad t^*_+ = |t|e^{i(\pi/2-0)},$$
$$t_- = |t|e^{i(3\pi/2-0)}, \quad t^*_- = |t|e^{-i(3\pi/2-0)}.$$

Then, as a consequence of the quantization of j_0, one gets

$$\mathrm{Disc}\, d^j_{QP}(t, t^*) = d^j_{QP}(t_+, t^*_+) - d^j_{QP}(t_-, t^*_-)$$
$$= (1 - e^{4\pi i j_0}) d^j_{QP}(t_+, t^*_+) = 0.$$

Similarly, the discontinuity produced by having to cut the P and Q planes likewise vanishes due to the quantization of j_0.

Finally, the representations obtained form a complete orthonormal set in the space of square integrable functions on the $SL(2, C)$ group, where the Haar measure can be expressed in terms of the parameters given by (1.4) and (1.5) as

$$dg = \sin\vartheta \sin\vartheta^* d\varphi_1 d\varphi_2 d\vartheta_1 d\vartheta_2 d\psi_1 d\psi_2.$$

With this, the orthogonality relation reads

$$\int dg\, T^{j'}_{Q'P'}(g)^* T^j_{QP}(g)$$
$$= \frac{64\pi^4}{(2j+1)(-2k-1)} \delta_{j_0' j_0} \delta(\sigma' - \sigma)\delta(P_1' - P_1)$$
$$\times \delta(P_2' - P_2)\delta(Q_1' - Q_1)\delta(Q_2' - Q_2).$$

The factor $(2j + 1)(-2k - 1) = j_0^2 + \sigma^2$ on the right-hand side is the familiar Plancherel measure.

ACKNOWLEDGMENTS

I would like to thank Dr. K. Szegö and Dr. K. Tóth for valuable remarks.

[1]V. Bargmann, Ann. Math. **48**, 568 (1947).
[2]M. A. Naimark, *Linear Representations of the Lorentz Group* (Macmillan, New York, 1964).
[3]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Oxford, New York, 1963).
[4]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Integral Geometry and Representation Theory, Generalized Functions, Vol. 5* (Academic, New York, 1966).
[5]S. Ström, Ark. Fys. **23**, 467 (1965); Ark. Fys. **33**, 465 (1967); Ark. Fys. **38**, 373 (1969).
[6]A. Sciarrino and M. Toller, J. Math. Phys. **8**, 1252 (1967).
[7]J. A. Verdiev, L. A. Dadashev, Yad. Fiz. **6**, 1094 (1967) [Sov. J. Nucl. Phys. **6**, 795 (1968)].

[8]N. Mukunda, J. Math. Phys. **9,** 50 (1968).
[9]Shau-jin Chang and L. O'Raifeartaigh, J. Math. Phys. **10,** 21 (1969).
[10]G. J. Iverson and G. Mack, J. Math. Phys. **11,** 1581 (1970).
[11]E. G. Kalnins, J. Math. Phys. **13,** 1304 (1972).
[12]K. Szegö and K. Tóth, J. Math. Phys. **12,** 853 (1971).
[13]J. Smorodinsky and G. Shepelev, Yad. Fiz. **13,** 441 (1971) [Sov. J. Nucl. Phys. **13,** 248 (1971)].
[14]N. W. Macfadyen, J. Math. Phys. **12,** 492 (1971).
[15]M. Huszár and J. Smorodinsky, Theor. Math. Phys. **4,** 328 (1970), JINR Preprint E2-4225.
[16]H. Joos and R. Schrader, DESY Preprint 68/40.
[17]G. J. Iverson, preprint, University of Adelaide, 1967, Part III.
[18]A. J. Macfarlane, J. Math. Phys. **3,** 1116 (1962).
[19]M. Huszár, J. Math. Phys. (to be published).
[20]M. Huszár, Commun. Math. Phys. **23,** 132 (1971).
[21]G. Domokos, S. Kövesi-Domokos, and F. Mansouri, Phys. Rev. D **1,** 500 (1970); Phys. Rev. D **1,** 512 (1970).
[22]N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations,* Mathematical Monographs Vol. 22 (Amer. Math. Soc., Providence, R. I., 1968). FT

[23]At first glance one would think that Eq. (1.20) for the matrix elements $T^{j}_{QP}(g)$ can be put down at once by means of the basis functions $(2\pi)^{-1} e^{-Pz + P^* z^*}$, and thus, the way in which Eq. (1.20) has been obtained is superfluously complicated. Actually, in the present case (and in general, in the case of continuous bases) the notion of matrix element of a representation has a slightly restricted meaning. Operators of representations act on $\Psi(P, P^*)$ functions as indicated in Eq. (1.19) and interpretation of the matrix elements $T^{j}_{QP}(g)$ as the kernel of the integral transformation (1.20) is possible only if the interchange of the order of integrations in (1.19) is legitimate. However, a detailed investigation shows that at $P = 0$ integrations in Eq. (1.19) cannot be interchanged. If in the explicit form of $T^{j}_{QP}(g)$ as given by Eq. (2.16) one still tries to take $\lim_{P \to 0} T^{j}_{QP}(g)$ it turns out to be nonexisting (an oscillating undetermined expression). Due to the symmetry of the d^{j}_{QP} functions a similar statement can be made on the limit $Q \to 0$ too.

[24]H. Bateman, A. Erdélyi, *Higher Transcendental Functions, Vol. 2* (McGraw-Hill, New York, 1953).

Bose–Einstein condensation in the presence of impurities

M. Kac

Department of Physics, Rockefeller University, New York, New York 10021*

J. M. Luttinger

Department of Physics, Columbia University†, New York, New York 10027

(Received 14 May 1973)

It is shown that, in three dimensions, a gas of noninteracting bosons in the presence of purely repulsive impurity centers undergoes a "Bose–Einstein" phase transition at sufficiently low temperature. In the course of the proof it is also shown that, as the size of the system approaches infinity, the lowest energy state approaches zero with probability one.

1. INTRODUCTION

In this paper we shall consider a gas of independent (i.e., noninteracting) bosons. This system differs, however, from that of free bosons by the presence of 'impurities.' That is, it is assumed that each boson interacts with a set of fixed centers, whose distribution is given. It is further assumed that the interaction between the bosons and the impurities is purely repulsive and finite-ranged. Under these circumstances we shall show that in three dimensions the Einstein–Bose condensation phenomena occurs.

Let the Hamiltonian h of the individual bosons be given by

$$h = h_0 + U, \tag{1.1}$$

where

$$h_0 \equiv p^2/2m \tag{1.2}$$

and U is the interaction with the impurities.

Write

$$h\psi_i = \epsilon_i \psi_i. \tag{1.3}$$

As usual, the chemical potential μ is determined by the equation

$$N = \sum_i \bar{n}_i = \sum_i \frac{1}{e^{\beta(\epsilon_i - \mu)} - 1}, \tag{1.4}$$

where $\bar{n}_i$ is the mean number of particles in the state i, N is the number of bosons present, and $\beta \equiv 1/kT$. By defining

$$\zeta = e^{\beta\mu} \tag{1.5}$$

the condition that $N \geq \bar{n}_i \geq 0$ implies that

$$e^{\beta\epsilon_i} > \zeta. \tag{1.6}$$

If the lowest energy level of h is ϵ_0, the strongest form of this inequality is

$$e^{\beta\epsilon_0} > \zeta. \tag{1.7}$$

Returning now to (1.4), we may write

$$N = \sum_i \frac{\zeta e^{-\beta\epsilon_i}}{1 - \zeta e^{-\beta\epsilon_i}} = \sum_i \sum_{l=1}^{\infty} e^{-\beta l \epsilon_i} \zeta^l \tag{1.8}$$

by (1.6)

Let us define the 'partition function' $Q(t)$ by

$$Q(t) = \sum_i e^{-t\epsilon_i}. \tag{1.9}$$

Then (1.8) becomes

$$N = \sum_{l=1}^{\infty} \zeta^l Q(l\beta). \tag{1.10}$$

Let us now recall how the *existence* of the Bose–Einstein condensation phenomena is proved for free particles. For free particles (periodic boundary conditions) $\epsilon_0 = 0$. Replacing the sum over states by an integral for large volume,

$$Q(t) = Q_0(t) \to [V/(2\pi)^3] \int d^3k \exp\{-\hbar^2 k^2 t/2m\} = V(2\pi\hbar^2 t/m)^{-3/2}. \tag{1.11}$$

Therefore

$$N = V\left(\frac{2\pi\hbar^2\beta}{m}\right)^{-3/2} \sum_{l=1}^{\infty} \frac{\zeta^l}{l^{3/2}} \tag{1.12}$$

Now define β_0 by

$$N = V\left(\frac{2\pi\hbar^2\beta_0}{m}\right)^{-3/2} \sum_{l=1}^{\infty} \frac{1}{l^{3/2}} \tag{1.13}$$

If $\beta > \beta_0$, we have

$$V\left(\frac{2\pi\hbar^2\beta}{m}\right)^{-3/2} \sum_{l=1}^{\infty} \frac{\zeta^l}{l^{3/2}} < V\left(\frac{2\pi\hbar^2\beta_0}{m}\right)^{-3/2} \sum_{1}^{\infty} \frac{\zeta^l}{l^{3/2}} = N\frac{\sum_1^\infty (\zeta^l/l)^{3/2}}{\sum_1^\infty (1/l)^{3/2}}. \tag{1.14}$$

Since, by (1.7), $\zeta < 1$ and $\sum_1^\infty \zeta^l/l^{3/2}$ is a monotonically increasing function of ζ, we have at once that

$$V\left(\frac{2\pi^2\beta}{m}\right)^{-3/2} \sum_{l=1}^{\infty} \frac{\zeta^l}{l^{3/2}} < N \quad (\beta > \beta_0) \tag{1.15}$$

so that (1.12) has no solution. Therefore a mistake has been made in deriving (1.12). As is well known, the mistake is the replacing of the sum over states by the integral for any large t (i.e., very large l) in (1.11). That is, as soon as β gets the least bit larger than β_0, the terms in (1.10) for l very large contribute comparably to the small l terms. The small l terms may be thought of as the normal contribution, the very large l terms as the contribution of the "condensate." For the free-boson case it can in fact be shown rigorously that the "condensate" is simply the macroscopic occupation of the lowest energy level.

For the impure boson gas we shall show the following: If the $Q(l\beta)$ are evaluated in the large volume limit hold-

Copyright © 1973 by the American Institute of Physics

ing l fixed, we again find that the resulting equation for ζ has no solution for β sufficiently large. This means that for sufficiently large β there is a contribution to (1.10) from a normal part (l of order unity) and a "condensate" (l very large). In other words a Bose-Einstein condensation takes place. However, the method of proof precludes the possibility of investigating the nature of the transition in detail. We have *not* shown, for example, that what happens physically is that a macroscopic number of particles go into the lowest state below the transition temperature. (But we suspect that it is true.)

The proof consists of two steps. In Sec. 2 we shall show that $Q(t) \le Q_0(t)$. Then, for $\beta > \beta_0$, ignoring the error for very large l, we have

$$\sum_{l=1}^{\infty} \zeta^l Q(l\beta) \le \sum_{l=1}^{\infty} \zeta^l Q_0(l\beta) < N \frac{\sum_1^\infty \zeta^l/l^{3/2}}{\sum_1^\infty 1/l^{3/2}}. \qquad (1.16)$$

In Sec. 3 we will prove that (with probability one) $\epsilon_0 - 0$. Therefore, by (1.7) ζ is again less than unity, and from (1.16)

$$\sum_{l=1}^{\infty} \zeta^l Q(l\beta) < N, \qquad (1.17)$$

which contradicts (1.10). Therefore for $\beta > \beta_0$ there is again a contribution from the "condensate" (very large l terms). This establishes the existence of the Bose-Einstein condensation phenomena, but not its exact nature. We also mention that this tells us that the value of β at the condensation temperature is *less* for the impure system than for the free system, or that the transition temperature is higher.

2. THE INEQUALITY $Q(t) \le Q_0(t)$

Consider

$$Q(t,\lambda) \equiv \mathrm{Tr} e^{-t(\mathcal{H}_0+\lambda U)}. \qquad (2.1)$$

We have at once from the definitions

$$Q(t) = Q(t,1), \qquad (2.2)$$

$$Q_0(t) = Q(t,0). \qquad (2.3)$$

Now clearly

$$\frac{\partial Q(t,\lambda)}{\partial\lambda} = -t\,\mathrm{Tr}[e^{-t(\mathcal{H}_0+\lambda U)}U] = -tQ(t,\lambda)U_\lambda, \qquad (2.4)$$

where

$$U_\lambda \equiv \mathrm{Tr}(e^{-t(\mathcal{H}_0+\lambda U)}U)/\mathrm{Tr}(e^{-t(\mathcal{H}_0+\lambda U)}). \qquad (2.5)$$

Integrating (2.4), we obtain

$$Q(t,\lambda) = Q(t,0)\exp(-t\int_0^\lambda d\lambda' U_{\lambda'}). \qquad (2.6)$$

Putting $\lambda = 1$, we have

$$Q(t) = Q_0(t)\exp(-t\int_0^1 d\lambda U_\lambda). \qquad (2.7)$$

Now if U is a nonnegative operator, its expectation value with respect to any density matrix must be nonnegative. From (2.5) we see that U_λ is just such an expectation value. Therefore, for U nonnegative

$$U_\lambda \ge 0, \qquad (2.8)$$

and from (2.7)

$$Q(t) \le Q_0(t).$$

If we apply this result to the case where U is a purely repulsive (positive) potential representing the interaction of the bosons with impurities, we have the desired result.

3. PROOF THAT $\epsilon_0 = 0$ WITH PROBABILITY ONE

We first assume that we have ν impurity centers distributed at random in our container of volume V. We take this container to be a cube of edge L, $V = L^3$. Imagine the container divided into $M = m^3$ smaller cubes (cells), each of edge L/M. Label the smaller cubes by an index $s = 1, 2, \ldots, M$. Call $P(n_1, n_2, \ldots, n_M)$ the probability that there are n_1 impurity centers in cell 1, n_2 impurity centers in cell 2, and so forth. Clearly this is given by

$$P(n_1,n_2,\ldots) = \frac{\nu!}{n_1!n_2!\ldots n_M!}\left(\frac{1}{M}\right)^{n_1}\left(\frac{1}{M}\right)^{n_2}\cdots\left(\frac{1}{M}\right)^{n_M}, \qquad (3.1)$$

since for a random distribution the probability of a center being in a small cell is $1/M$. Since

$$\sum_{s=1}^{M} n_s = \nu, \qquad (3.2)$$

(3.1) may be written

$$P(n_1,n_2,\ldots) = \frac{\nu!}{n_1!\ldots n_M!}\left(\frac{1}{M}\right)^\nu. \qquad (3.3)$$

Let P be the total probability that every small cell have an occupancy of at least one center; then

$$P = \left(\frac{1}{M}\right)^\nu \sum_{\substack{n_s\ge 1\\ (\Sigma n_s=\nu)}} \frac{\nu!}{n_1!\ldots n_M!}. \qquad (3.4)$$

This summation may be carried out in the usual fashion:

$$\sum_{n_s\ge1}\frac{\nu!}{n_1!\ldots n_M!} = \frac{1}{2\pi}\int_0^{2\pi} d\phi \sum_{n_s\ge1}\frac{\nu!}{n_1!\ldots n_M!}e^{i[n_1+\cdots+n_M-\nu]\phi}$$
$$= \frac{\nu!}{2\pi}\int_0^{2\pi} d\phi\, e^{-i\nu\phi}\left(\sum_{n=1}^{\infty}(e^{i\phi})^n\frac{1}{n!}\right)^M$$
$$= \frac{\nu!}{2\pi}\int_0^{2\pi} d\phi\, e^{-i\nu\phi}[\exp(e^{i\phi})-1]^M$$
$$= \frac{\nu!}{2\pi i}\oint_{\text{unit circle}}\frac{dz}{z^{\nu+1}}(e^z-1)^M$$
$$= \frac{d}{dz}^{\nu}(e^z-1)^M\Big|_{z=0} = \sum_{l=0}^{M}\frac{(-1)^l M!}{l!(M-l)!}(M-l)^\nu. \qquad (3.5)$$

Therefore

$$P = \sum_{l=0}^{M}\frac{(-1)^l M!}{l!(M-l)!}\left(1-\frac{l}{M}\right)^\nu \qquad (3.6)$$

We are interested in evaluating this for the case of a very large system ($\nu \gg 1$), where the "small" cells are chosen to be very large but much smaller than the size of the entire system, i.e.,

$$1 \ll M \ll \nu. \qquad (3.7)$$

It is clear that, under these circumstances, as l increases the factor $(1 - l/M)^\nu$ decreases more rapidly than any other factor can increase, and the main contribution comes from small l/M. For small l/M we may approximate

$$(1 - l/M)^\nu = e^{-\nu l/M}$$

so that (3.6) becomes

$$P \cong \sum_{l=0}^{M} \frac{(-)^l M!}{l!(M-l)!} (e^{-\nu/M})^l = (1 - e^{-\nu/M})^M \qquad (3.8)$$

or[1]

$$P \cong \exp(-Me^{-\nu/M}). \qquad (3.9)$$

Now suppose we choose

$$M = \nu/(\log\nu)^{1/2}. \qquad (3.10)$$

Then

$$Me^{-\nu/M} = [\nu/(\log\nu)^{1/2}]e^{-(\log\nu)^{1/2}},$$

which approaches ∞ as ν approaches ∞. Therefore for a "small" cell size determined by (3.10), the probability of every "small" cell being occupied goes to zero as the size of the system approaches ∞. That is, with probability one, at least one small cell is empty. Now the small cells are cubes of edge $a = L/M^{1/3}$. Using (3.10), we have

$$a = [1/\rho^{1/3}](\log\nu)^{1/6}, \qquad (3.11)$$

where

$$\rho \equiv \nu/V, \qquad (3.12)$$

the density of impurities. Therefore the "small" cells get infinitely large as the size of the system approaches infinity.

Now consider a small cell which is free of impurity centers. Since we have assumed that the range of the potential is finite, this means that there is, with probability one, a region R (say a cube of edge a') which goes to ∞ as the system size goes to ∞ in which the impurity potential is zero. Let us take as a normalized trial wave function ψ (place the origin at one corner of cube R)

$$\psi = \left(\frac{2}{a'}\right)^{3/2} \sin\left(\frac{\pi}{a'}x\right) \sin\left(\frac{\pi}{a'}y\right) \sin\left(\frac{\pi}{a'}z\right) \quad \text{in } R \qquad (3.13)$$

$$= 0 \quad \text{otherwise.}$$

The Rayleigh–Ritz variational principle tells us that

$$\begin{aligned} \epsilon_0 &\le \int_V d\mathbf{r}\ \psi h \psi \\ &= +\frac{\hbar^2}{2m} \int_R d\mathbf{r} \left|\left(\frac{\partial\psi}{\partial x}\right)^2 + \left(\frac{\partial\psi}{\partial y}\right)^2 + \left(\frac{\partial\psi}{\partial z}\right)^2\right| \\ &= +\frac{3\hbar^2}{2m}\left(\frac{\pi}{a'}\right)^2. \end{aligned} \qquad (3.14)$$

Since a' approaches infinity as the size of the system approaches infinity, we have the result that, with probability one, an upper bound for ϵ_0 approaches zero. On the other hand, the nonnegativity of h implies that

$$\epsilon_0 \geqslant 0 \qquad (3.15)$$

so that as the size of the system approaches ∞ we have, with probability one, that ϵ_0 approaches zero.

If the impurity centers are not distributed at random, we may proceed as follows. Call $P(\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_\nu)$ $d\mathbf{R}_L d\mathbf{R}_2 \ldots d\mathbf{R}_\nu$ the probability that the first impurity center is in $d\mathbf{R}_1$ around $\mathbf{R}_1$, the second in $d\mathbf{R}_2$ around $\mathbf{R}_2$, and so forth. Again, divide V into M equals "cells" of volume Ω:

$$\Omega = V/M \qquad (3.16)$$

such that (3.7) is satisfied. Now call q the probability that there is no impurity center in one of these cells. Then $1-q$ is the probability that there is at least one center in the cell. Clearly, because of (3.7) the probability that there is at least one center in two of the cells is $(1-q)^2$, since the constraint that at least one center is in one cell is negligible. Continuing in this fashion, we obtain for P (the probability that at least one center is in each cell)

$$P = (1-q)^M. \qquad (3.17)$$

It remains to calculate q. This is given by

$$q = \int_{V-R} d\mathbf{R}_1 \cdots d\mathbf{R}_\nu P(\mathbf{R}_1, \ldots, \mathbf{R}_\nu). \qquad (3.18)$$

Define the l-center distribution function by

$$n_l(R_1, \ldots, R_l) = \frac{\nu!}{(\nu-l)!} \times \int_V d\mathbf{R}_{l+1} d\mathbf{R}_{l+2} \cdots d\mathbf{R}_\nu P(\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_\nu), \qquad (3.19)$$

then one has at once[2]

$$q = 1 - \frac{1}{1!}\int_\Omega d\mathbf{R}_1 n_1(\mathbf{R}_1) + \frac{1}{2!}\int_\Omega d\mathbf{R}_1 d\mathbf{R}_2 n_2(R_1, R_2) + \cdots. \qquad (3.20)$$

In terms the *correlation functions* χ_l defined by

$$\begin{aligned} n_1(\mathbf{R}_1) &= \chi_1(\mathbf{R}_1), \\ n_2(\mathbf{R}_1, \mathbf{R}_2) &= \chi_2(\mathbf{R}_1, \mathbf{R}_2) + \chi_1(\mathbf{R}_1)\chi_1(\mathbf{R}_2), \\ n_3(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) &= \chi_2(\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3) + \chi_1(\mathbf{R}_1)\chi_2(\mathbf{R}_2, \mathbf{R}_3) \\ &\quad + \chi_1(\mathbf{R}_2)\chi_2(\mathbf{R}_2, \mathbf{R}_3) + \chi_1(\mathbf{R}_3)\chi_2(\mathbf{R}_1, \mathbf{R}_2) \\ &\quad + \chi_1(\mathbf{R}_1)\chi_1(\mathbf{R}_2)\chi_1(\mathbf{R}_3) \end{aligned} \qquad (3.21)$$

and so forth; (3.20) becomes

$$q = \exp\left(\sum_{l=1}^{\infty} \frac{(-1)^l}{l!} \int_\Omega d\mathbf{R}_1 \cdots d\mathbf{R}_l \chi_l(\mathbf{R}_1, \ldots, \mathbf{R}_l)\right). \qquad (3.22)$$

When Ω is large (larger than any characteristic length in the probability distribution function of the centers), the factor in the exponent is proportional to Ω for a homogeneous system. For independent centers $\chi_1 = \rho$, $\chi_l = 0$ $(l \ge 2)$ and $q = \exp(-\rho\Omega)$ which at once yields (3.8). If $P(R_1, \ldots, R_N)$ is given by a thermal equilibrium distribution then, under suitable conditions,[1] the bracket is just $-\beta p\Omega$, where p is the thermodynamic pressure of the gas of impurity centers. Writing for the bracket $-\rho\gamma\Omega$, where γ is a dimensionless number independent of Ω, we have

$$q = e^{-\rho\gamma\Omega} = e^{-\rho\gamma\nu/M} = e^{-\gamma(\nu/M)}, \qquad (3.23)$$

$$P = (1 - e^{-\gamma\nu/M})^M = \exp(-Me^{-\gamma\nu/M})^M. \qquad (3.24)$$

This is identical with (3.9) except for the factor γ, and consequently all the reasoning which follows (3.9) still holds. Therefore, we conclude that $\epsilon_0 \to 0$ with probability one as the size of the system goes to ∞ as long as the correlation function series in (3.22) converges.

*Supported in part by AFOSR, Grant 722187.
†Supported in part by the National Science Foundation.

[1]A more precise discussion may be found in W. Feller, *Probability Theory and Its Applications* (Wiley, New York, 1950), Vol. I, pp. 69ff.

[2]The discussion that follows is very closely patterned after a similar but more carefully stated analysis by the authors in a paper "A New Formula for the Pressure in Statistical Mechanics" (to be published). We are using, of course, the inclusion–exclusion lemma of probability theory.

Kinetic theory of a weakly coupled and weakly inhomogeneous plasma in a magnetic field

Jacqueline Naze Tjøtta and Alf H. Øien

Department of Applied Mathematics, University of Bergen, Bergen, Norway
(Received 22 December 1972; revised manuscript received 13 March 1973)

A study of the evolution toward continuum of a weakly coupled, weakly inhomogeneous two-component electron–ion plasma in an external electromagnetic field using the multiple-time-scale method is presented. An electron–ion mass ratio parameter and a weak inhomogeneity parameter are introduced. Kinetic and macroscopic equations for a certain ordering between these parameters are obtained. The equations are solved when using simplified collision terms.

I. INTRODUCTION

We study the evolution toward continuum of weakly coupled, weakly inhomogeneous two-component electron–ion plasma in an electromagnetic field using the multiple-time-scale method. An impetus to this examination has been the theory of Chap. 18 in the book of Chapman and Cowling.[1] For us it seems that the underlying parametrization of the kinetic equations used there is a little arbitrary. The problem is connected with the separation of the two magnetic force terms in each of the kinetic equations and the underlying hypothesis when these two terms are taken to be of different order of magnitude. This can be understood when the ratio between the magnitude of the mass transport velocity and the thermal speed of each particle type is a small quantity. However, this ordering seems to be restricted to only these two terms in the equations.

We have introduced various parameters in the kinetic equations in order to allow the two above-mentioned terms to be treated as of different orders of magnitude and at the same time make it possible to derive a set of macroscopic equations for the plasma model by use of the multiple-time-scale method. Similar derivations have been performed by McCune *et al.*[2] for a one-component Boltzmann gas and by us[3] for a one-component weakly coupled, weakly inhomogeneous gas without fields. It seems as if a similar derivation is possible for the two-component plasma model only if the square root of the mass ratio of electrons and ions is considered as a small parameter α and is related to the weak inhomogenity parameter $\epsilon \sim \bar{c}_1\tau_{20}/L$, where $\bar{c}_1$ is the mean particle speed for electrons, τ_{20} the effective time between electron-electron collisions, and L a characteristic length for inhomogenities. If the square root mass ratio is not introduced, the procedure as far as we can see goes through only under very special conditions. However, we will not discuss this problem here.

In Sec. II we define the plasma model to be studied and write down kinetic and macroscopic equations in parametrized form using the parameters ϵ and α. In Sec. III we assume that ϵ and α are of the same order of magnitude and solve the kinetic and macroscopic equations by the multiple-time-scale method up to second order in the smallness parameter. From the equations correct to zeroth and first orders we get that distribution functions for electrons and ions to lowest order evolve toward functions of the same form as in Ref. 1 but at different times: Evolution of electrons takes place on τ_{20} time scale while evolution of ions takes place on the $\tau_{21} \sim \tau_{20}/\alpha$ time scale. Small corrections to these distribution functions have transient behavior on the τ_{20} time scale, due to the electron evolution; so have the macroscopic quantities. From the equations to second order we obtain, together with transients on the τ_{20} and τ_{21} time scales due to electron and ion evolution, the macroscopic transport equations on the $\tau_{22} \sim \tau_{20}/\alpha^2$ time scale. This scale is also the time scale for energy transfer between electrons and ions. Our equations are similar but not identical to the equations obtained in Ref. 1. In Sec. IV we solve the equations of Sec. III using instead of Fokker–Planck collision terms simple relaxation terms in the kinetic equations. A picture of the evolution into equilibrium for our gas model is obtained.

II. BASIC EQUATIONS AND ASSUMPTIONS

We will study the evolution of a fully ionized two-component inhomogeneous electron–ion plasma placed in an external electromagnetic field. The notations of Ref. 1 are used throughout. Subscript 1 refers to electrons and 2 to ions.

The gas is assumed to be weakly coupled,[4-6] i.e.,

$$n_i r_0^3 \sim 1, \quad \varphi_0/kT_i \ll 1, \quad i = 1,2.$$

Here T_1 and T_2 are temperatures for the gas components, φ_0 is a characteristic potential energy of two interacting particles, and r_0 is the range of this interaction. Then interactions between particles give rise to Fokker–Planck (FP) collision terms in kinetic equations.

We also assume that the kinetic energies of electrons and ions are of equal order of magnitude so that

$$\tfrac{1}{2} m_1\bar{c}_1^2 \sim \tfrac{1}{2} m_2\bar{c}_2^2.$$

Thus

$$\bar{c}_2/\bar{c}_1 \sim (m_1/m_2)^{1/2} = \alpha \ll 1.$$

Having $n_1 \sim n_2$ and $e_1 \sim e_2$ and assuming for the mean particle velocities

$$|\bar{\mathbf{c}}_1| \sim \bar{c}_1, \quad |\bar{\mathbf{c}}_2| \sim \alpha\bar{c}_1,$$

for the mass transport vector $\mathbf{c}_0$, we therefore have

$$|\mathbf{c}_0| = (1/\rho)|n_1m_1\bar{\mathbf{c}}_1 + n_2m_2\bar{\mathbf{c}}_2| \sim \alpha\bar{c}_1.$$

The peculiar velocities $\mathbf{C}_1$ and $\mathbf{C}_2$ are estimated as

$$|\mathbf{C}_1| = |\mathbf{c}_1 - \mathbf{c}_0| \sim \bar{c}_1,$$
$$|\mathbf{C}_2| = |\mathbf{c}_2 - \mathbf{c}_0| \sim \alpha\bar{c}_1.$$

Besides the parameter α we introduce the parameter

$$\epsilon \sim \bar{c}_1\tau_{20}/L \ll 1,$$

where τ_{20} is the effective time[4] between electron–electron collisions and L is a characteristic length for

Copyright © 1973 by the American Institute of Physics

inhomogenities (assumed to be the same for electrons and ions). The external magnetic field is assumed to give rise to gyrofrequencies Ω_1 and Ω_2 such that

$$\tau_{20}\Omega_1 \sim 1$$

(and $\tau_{20}\Omega_2 \sim \alpha^2$). The particle accelerations, $\mathbf{F}_i = e_i/m_i \mathbf{E}$ due to the electric field $\mathbf{E}$, are assumed to be so weak that $\tau_{20}|\mathbf{F}_1|/|\mathbf{C}_1| \sim \epsilon$ (and $\tau_{20}|\mathbf{F}_2|/|\mathbf{C}_2| \sim \alpha\epsilon$). Both $\mathbf{H}$ and $\mathbf{F}_i$ are assumed to be stationary and uniform. This last assumption may be avoided; see Appendix A.

Starting from the BBGKY equations for this two-component gas model the following kinetic equations for electrons and ions are derived by standard procedure:[4-6]

$$\begin{aligned}
&\frac{\partial f_1}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial f_1}{\partial \mathbf{r}} + \epsilon \mathbf{C}_1 \cdot \frac{\partial f_1}{\partial \mathbf{r}} + \epsilon \mathbf{F}_1 \cdot \frac{\partial f_1}{\partial \mathbf{C}_1} \\
&\quad - \alpha\left(\frac{\partial \mathbf{c}_0}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}}\right) \cdot \frac{\partial f_1}{\partial \mathbf{C}_1} + \alpha \frac{e_1}{m_1} \mathbf{c}_0 \; \mathbf{H} \cdot \frac{\partial f_1}{\partial \mathbf{C}_1} \\
&\quad + \frac{e_1}{m_1} \mathbf{C}_1 \times \mathbf{H} \cdot \frac{\partial f_1}{\partial \mathbf{C}_1} - \epsilon\alpha \frac{\partial f_1}{\partial \mathbf{C}_1} \mathbf{C}_1 : \cdot \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}} \\
&\quad = \frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1} \cdot \int d\mathbf{C}_1' \mathbf{\Phi}^{11}(\mathbf{C}_1 - \mathbf{C}_1') \cdot \\
&\quad \times \left(\frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1} - \frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1'}\right) f_1(\mathbf{C}_1) f_1(\mathbf{C}_1') \\
&\quad + \frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1} \cdot \int d\mathbf{C}_2 \mathbf{\Phi}^{12}(\mathbf{C}_1 - \alpha \mathbf{C}_2) \cdot \\
&\quad \times \left(\frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1} - \frac{\alpha}{m_2} \frac{\partial}{\partial \mathbf{C}_2}\right) f_1(\mathbf{C}_1) f_2(\mathbf{C}_2), \qquad (1)
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial f_2}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial f_2}{\partial \mathbf{r}} + \epsilon\alpha \mathbf{C}_2 \cdot \frac{\partial f_2}{\partial \mathbf{r}} + \epsilon\alpha \mathbf{F}_2 \cdot \frac{\partial f_2}{\partial \mathbf{C}_2} \\
&\quad - \left(\frac{\partial \mathbf{c}_0}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}}\right) \cdot \frac{\partial f_2}{\partial \mathbf{C}_2} + \alpha^2 \frac{e_2}{m_2} \mathbf{c}_0 \times \mathbf{H} \cdot \frac{\partial f_2}{\partial \mathbf{C}_2} \\
&\quad + \alpha^2 \frac{e_2}{m_2} \mathbf{C}_2 \times \mathbf{H} \cdot \frac{\partial f_2}{\partial \mathbf{C}_2} - \epsilon\alpha \frac{\partial f_2}{\partial \mathbf{C}_2} \mathbf{C}_2 : \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}} \\
&\quad = \frac{\alpha}{m_2} \frac{\partial}{\partial \mathbf{C}_2} \cdot \int d\mathbf{C}_2' \mathbf{\Phi}^{22}(\mathbf{C}_2 - \mathbf{C}_2') \cdot \left(\frac{1}{m_2} \frac{\partial}{\partial \mathbf{C}_2} - \frac{1}{m_2} \frac{\partial}{\partial \mathbf{C}_2'}\right) \\
&\quad \times f_2(\mathbf{C}_2) f_2(\mathbf{C}_2') + \frac{\alpha}{m_2} \frac{\partial}{\partial \mathbf{C}_2} \cdot \int d\mathbf{C}_1 \mathbf{\Phi}^{21}(\alpha \mathbf{C}_2 - \mathbf{C}_1) \cdot \\
&\quad \times \left(\frac{\alpha}{m_2} \frac{\partial}{\partial \mathbf{C}_2} - \frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1}\right) f_1(\mathbf{C}_1) f_2(\mathbf{C}_2). \qquad (2)
\end{aligned}$$

Here t is a time variable on the τ_{20} time scale. $\mathbf{\Phi}^{ij}(\mathbf{C}_i - \mathbf{C}_j)$ are tensors and given by (Refs. 3-6 for a one-component gas)

$$\mathbf{\Phi}^{ij}(\mathbf{v}) = \int d\mathbf{x}_{ij} \frac{\partial \varphi_{ij}}{\partial \mathbf{x}_{ij}} \int_0^\infty d\tau \frac{\partial \varphi_{ij}}{\partial \mathbf{x}'_{ij}} \quad (\mathbf{x}'_{ij} = \mathbf{x}_{ij} - \mathbf{v}\tau).$$

φ_{ij} is the potential energy and $\mathbf{x}_{ij}$ the relative coordinate of two colliding particles "i" and "j". It is to be noted that $|\mathbf{\Phi}^{22}|/|\mathbf{\Phi}^{11}| \sim 1/\alpha$. This is the reason why α and not α^2 appears in front of the ion-ion collision integral.

Note that the distribution functions f_i are considered as functions of the peculiar velocities $\mathbf{C}_i = \mathbf{c}_i - \mathbf{c}_0$ besides $\mathbf{r}$ and t. We must then be aware of the *a priori* condition

$$\int d\mathbf{C}_1 f_1 m_1 \mathbf{C}_1 + \int d\mathbf{C}_2 f_2 m_2 \mathbf{C}_2 = 0. \qquad (3)$$

When deriving the parametrized moment equations, we assume that relative order of smallness of two averaged quantities are the same as for non average quantities, i.e.,

$$\begin{aligned}
&|\overline{\mathbf{C}}_1| \sim \bar{c}_1, \quad |\overline{\mathbf{C}}_2| \leq \alpha \bar{c}_1, \\
&|m_i \overline{\mathbf{C}_i \mathbf{C}_i}| \sim m_i \overline{C_i^2} \sim m_i \bar{c}_i^2, \\
&|\mathbf{q}_i|/n_i = |\tfrac{1}{2} m_i \overline{C_i^2 \mathbf{C}_i}| \sim |m_i \overline{C}_i^2 \overline{\mathbf{C}}_i|.
\end{aligned}$$

Consequently, we also parametrize the *a priori* condition Eq. (3) in the following way:

$$\alpha \int d\mathbf{C}_1 f_1 m_1 \mathbf{C}_1 + \int d\mathbf{C}_2 f_2 m_2 \mathbf{C}_2 = 0. \qquad (4)$$

It can easily be shown that this parametrization of Eq. (3) is necessary in order to avoid a break down of the multiple time scale procedure when we consider evolution from an initial state which lies far from equilibrium.

Using Eq. (4) and the conservation properties of the collision integrals, we derive

$$\frac{\partial \rho}{\partial t} + \epsilon\alpha \frac{\partial}{\partial \mathbf{r}} \cdot (\rho \mathbf{c}_0) = 0, \qquad (5)$$

$$\begin{aligned}
\rho\left(\frac{\partial \mathbf{c}_0}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}}\right) &= -\epsilon\alpha \frac{\partial}{\partial \mathbf{r}} \cdot \sum_{i=1}^{2} n_i m_i \overline{\mathbf{C}_i \mathbf{C}_i} \\
&+ \alpha^2 \rho_e \mathbf{c}_0 \times \mathbf{H} + \alpha n_1 e_1 \overline{\mathbf{C}}_1 \times \mathbf{H} + \alpha^2 n_2 e_2 \overline{\mathbf{C}}_2 \times \mathbf{H} \\
&+ \epsilon\alpha \sum_{i=1}^{2} \rho_i \mathbf{F}_i, \qquad (6)
\end{aligned}$$

$$\begin{aligned}
\frac{3}{2} nk \left(\frac{\partial T}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial T}{\partial \mathbf{r}}\right) &= \epsilon \frac{3}{2} kT \frac{\partial}{\partial \mathbf{r}} \cdot (n_1 \overline{\mathbf{C}}_1) \\
&+ \epsilon\alpha \frac{3}{2} kT \frac{\partial}{\partial \mathbf{r}} \cdot (n_2 \overline{\mathbf{C}}_2) + \epsilon \rho_1 \mathbf{F}_1 \cdot \overline{\mathbf{C}}_1 + \epsilon\alpha \rho_2 \mathbf{F}_2 \cdot \overline{\mathbf{C}}_2 \\
&+ \alpha n_1 e_1 \overline{\mathbf{C}}_1 \cdot (\mathbf{c}_0 \times \mathbf{H}) + \alpha^2 n_2 e_2 \overline{\mathbf{C}}_2 \cdot (\mathbf{c}_0 \times \mathbf{H}) - \epsilon \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{q}_1 \\
&- \epsilon\alpha \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{q}_2 - \epsilon\alpha \sum_{i=1}^{2} n_i m_i \overline{\mathbf{C}_i \mathbf{C}_i} : \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}}. \qquad (7)
\end{aligned}$$

We also take into account the following moment equations for each gas component:

$$\frac{\partial \rho_1}{\partial t} + \epsilon\alpha \frac{\partial}{\partial \mathbf{r}} \cdot (\rho_1 \mathbf{c}_0) + \epsilon \frac{\partial}{\partial \mathbf{r}} \cdot (\rho_1 \overline{\mathbf{C}}_1) = 0, \qquad (8)$$

$$\frac{\partial \rho_2}{\partial t} + \epsilon\alpha \frac{\partial}{\partial \mathbf{r}} \cdot (\rho_2 \mathbf{c}_0) + \epsilon\alpha \frac{\partial}{\partial \mathbf{r}} \cdot (\rho_2 \overline{\mathbf{C}}_2) = 0, \qquad (9)$$

$$\begin{aligned}
&\frac{\partial}{\partial t} (\rho_1 \overline{\mathbf{C}}_1) + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial}{\partial \mathbf{r}} (\rho_1 \overline{\mathbf{C}}_1) + \epsilon\alpha \rho_1 \overline{\mathbf{C}}_1 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0 \\
&\quad + \epsilon \frac{\partial}{\partial \mathbf{r}} \cdot (\rho_1 \overline{\mathbf{C}_1 \mathbf{C}_1}) - \epsilon \rho_1 \mathbf{F}_1 + \alpha \rho_1 \left(\frac{\partial \mathbf{c}_0}{\partial t} + \epsilon\alpha \mathbf{c}_0 \cdot \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}}\right) \\
&\quad - \alpha n_1 e_1 \mathbf{c}_0 \times \mathbf{H} - n_1 e_1 \overline{\mathbf{C}}_1 \times \mathbf{H} + \epsilon\alpha \rho_1 \overline{\mathbf{C}}_1 \cdot \frac{\partial \mathbf{c}_0}{\partial \mathbf{r}} \\
&\quad = -\int d\mathbf{C}_1 d\mathbf{C}_2 \mathbf{\Phi}^{12}(\mathbf{C}_1 - \alpha \mathbf{C}_2) \cdot \\
&\quad \times \left(\frac{1}{m_1} \frac{\partial}{\partial \mathbf{C}_1} - \frac{\alpha}{m_2} \frac{\partial}{\partial \mathbf{C}_2}\right) f_1(\mathbf{C}_1) f_2(\mathbf{C}_2), \qquad (10)
\end{aligned}$$

$$\frac{\partial}{\partial t}(\rho_2\overline{\mathbf{C}}_2)+\epsilon\alpha\mathbf{c}_0\cdot\frac{\partial}{\partial\mathbf{r}}(\rho_2\overline{\mathbf{C}}_2)+\epsilon\alpha\rho_2\overline{\mathbf{C}}_2\frac{\partial}{\partial\mathbf{r}}\cdot\mathbf{c}_0$$
$$+\epsilon\alpha\frac{\partial}{\partial\mathbf{r}}\cdot(\rho_2\overline{\mathbf{C}_2\mathbf{C}_2})-\epsilon\alpha\rho_2\mathbf{F}_2$$
$$+\rho_2\left(\frac{\partial\mathbf{c}_0}{\partial t}+\epsilon\alpha\mathbf{c}_0\cdot\frac{\partial\mathbf{c}_0}{\partial\mathbf{r}}\right)-\alpha^2n_2e_2\mathbf{c}_0\times\mathbf{H}$$
$$-\alpha^2n_2e_2\overline{\mathbf{C}}_2\times\mathbf{H}+\epsilon\alpha\rho_2\overline{\mathbf{C}}_2\cdot\frac{\partial\mathbf{c}_0}{\partial\mathbf{r}}$$
$$=-\alpha\int d\mathbf{C}_1d\mathbf{C}_2\mathbf{\Phi}^{12}(\mathbf{C}_1-\alpha\mathbf{C}_2)\cdot\left(\frac{\alpha}{m_2}\frac{\partial}{\partial\mathbf{C}_2}-\frac{1}{m_1}\frac{\partial}{\partial\mathbf{C}_1}\right)$$
$$\times f_1(\mathbf{C}_1)f_2(\mathbf{C}_2), \tag{11}$$

$$\frac{3}{2}n_1k\left(\frac{\partial T_1}{\partial t}+\epsilon\alpha\mathbf{c}_0\cdot\frac{\partial T_1}{\partial\mathbf{r}}\right)-\epsilon\frac{3}{2}kT_1\frac{\partial}{\partial\mathbf{r}}\cdot(n_1\overline{\mathbf{C}}_1)+\epsilon\frac{\partial}{\partial\mathbf{r}}\cdot\mathbf{q}_1$$
$$-\epsilon\rho_1\mathbf{F}_1\cdot\overline{\mathbf{C}}_1+\alpha\rho_1\overline{\mathbf{C}}_1\cdot\left(\frac{\partial\mathbf{c}_0}{\partial t}+\epsilon\alpha\mathbf{c}_0\cdot\frac{\partial\mathbf{c}_0}{\partial\mathbf{r}}\right)$$
$$-\alpha n_1e_1\overline{\mathbf{C}}_1\cdot\mathbf{c}_0\times\mathbf{H}+\epsilon\alpha\rho_1\overline{\mathbf{C}_1\mathbf{C}_1}:\frac{\partial\mathbf{c}_0}{\partial\mathbf{r}}$$
$$=-\alpha\int d\mathbf{C}_1d\mathbf{C}_2\mathbf{C}_2\cdot\mathbf{\Phi}^{12}(\mathbf{C}_1-\alpha\mathbf{C}_2)\cdot\left(\frac{1}{m_1}\frac{\partial}{\partial\mathbf{C}_1}\right.$$
$$\left.-\frac{\alpha}{m_2}\frac{\partial}{\partial\mathbf{C}_2}\right)f_1(\mathbf{C}_1)f_2(C_2), \tag{12}$$

$$\frac{3}{2}n_2k\left(\frac{\partial T_2}{\partial t}+\epsilon\alpha\mathbf{c}_0\cdot\frac{\partial T_2}{\partial\mathbf{r}}\right)-\epsilon\alpha\frac{3}{2}kT_2\frac{\partial}{\partial\mathbf{r}}\cdot(n_2\overline{\mathbf{C}}_2)$$
$$+\epsilon\alpha\frac{\partial}{\partial\mathbf{r}}\cdot\mathbf{q}_2-\epsilon\alpha\rho_2\mathbf{F}_2\cdot\overline{\mathbf{C}}_2$$
$$+\rho_2\overline{\mathbf{C}}_2\cdot\left(\frac{\partial\mathbf{c}_0}{\partial t}+\epsilon\alpha\mathbf{c}_0\cdot\frac{\partial\mathbf{c}_0}{\partial\mathbf{r}}\right)$$
$$-\alpha^2n_2e_2\overline{\mathbf{C}}_2\cdot(\mathbf{c}_0\times\mathbf{H})+\epsilon\alpha\rho_2\overline{\mathbf{C}_2\mathbf{C}_2}:\frac{\partial\mathbf{c}_0}{\partial\mathbf{r}}$$
$$=-\alpha\int d\mathbf{C}_1d\mathbf{C}_2\mathbf{C}_2\cdot\mathbf{\Phi}^{12}(\mathbf{C}_1-\alpha\mathbf{C}_2)\cdot\left(\frac{\alpha}{m_2}\frac{\partial}{\partial\mathbf{C}_2}\right.$$
$$\left.-\frac{1}{m_1}\frac{\partial}{\partial\mathbf{C}_1}\right)f_1(\mathbf{C}_1)f_2(\mathbf{C}_2). \tag{13}$$

III. SOLUTION OF KINETIC- AND MOMENT EQUATIONS BY THE MULTIPLE-TIME-SCALE METHOD

We will now solve the set of kinetic and macroscopic equations by successive approximations.

We have two parameters, α and ϵ, which are both small but independent of each other. In order to solve the equations by a one-parameter successive approximations scheme, we have to assume some relation between α and ϵ. In this work we study the model corresponding to $\epsilon\sim\alpha$, $\epsilon^2\sim\epsilon\alpha\sim\alpha^2$ and so on.

We assume that the distribution functions $f_i(\mathbf{C}_i)$ can be found as

$$f_i=\sum_{p=0}^{\infty}\epsilon^pf_i^{(p)}.$$

Using the definitions of ρ and T, we have

$$\rho=\sum_{p=0}^{\infty}\epsilon^p\rho^{(p)},$$

where $\rho^{(p)}=\sum_{i=1}^{2}\int d\mathbf{C}_i f_i^{(p)}(\mathbf{C}_i)m_i$ and similarly for n,n_1,n_2,ρ_1, and ρ_2,

$$nkT=\sum_{p=0}^{\infty}\epsilon^p(nkT)^{(p)},$$

where $(nkT)^{(p)}=\frac{1}{3}\sum_{i=1}^{2}\int d\mathbf{C}_if_i^{(p)}(\mathbf{C}_i)m_iC_i^2$ and similarly for T_1 and T_2 and also for $\rho_1\overline{\mathbf{C}}_1$ and $\rho_2\overline{\mathbf{C}}_2$.

There is no such connection between expansions for $\rho\mathbf{c}_0$ and for $f_i(\mathbf{C}_i)$. However, we seek $\mathbf{c}_0$ as

$$\mathbf{c}_0=\sum_{p=0}^{\infty}\epsilon^p\mathbf{c}_0^{(p)}.$$

Later we shall delete all ( ) in the superscripts.

In the collision integrals we expand the tensor $\mathbf{\Phi}^{12}(\mathbf{C}_1-\alpha\mathbf{C}_2)$ in a Taylor series which is assumed to be convergent in the distributional sense:

$$\mathbf{\Phi}^{12}(\mathbf{C}_1-\alpha\mathbf{C}_2)=\mathbf{\Phi}^{12}(\mathbf{C}_1)-\alpha\mathbf{C}_2\cdot\frac{\partial}{\partial\mathbf{C}_1}\mathbf{\Phi}^{12}(\mathbf{C}_1)$$
$$+\frac{\alpha^2}{2!}\left(\mathbf{C}_2\mathbf{C}_2:\frac{\partial^2}{\partial\mathbf{C}_1^2}\right)\mathbf{\Phi}^{12}(\mathbf{C}_1)-\cdots.$$

Following the multiple-time-scale method,[4-6] the time derivative is expanded as

$$\frac{\partial}{\partial t}\to\frac{\partial}{\partial t_{20}}+\epsilon\frac{\partial}{\partial t_{21}}+\epsilon^2\frac{\partial}{\partial t_{22}}+\epsilon^3\frac{\partial}{\partial t_{23}}+\cdots$$

in all the equations. Here t_{2i}, $i=0,1,2,\cdots$, is the time variable on the $\tau_{2i}\sim\tau_{20}/\epsilon^i$ time scale.

Substituting all these expansions in the equations and in the *a priori* condition, and collecting terms of equal order of magnitude, we derive for each order in the small parameter a set of equations which we next solve step by step.

A. Zeroth-order equations

$$\int d\mathbf{C}_2f_2^0(\mathbf{C}_2)m_2\mathbf{C}_2=0, \tag{14}$$

$$\frac{\partial f_1^0}{\partial t_{20}}+\frac{e_1}{m_1}\mathbf{C}_1\times\mathbf{H}\cdot\frac{\partial f_1^0}{\partial\mathbf{C}_1}=FP_{11}[f_1^0f_1^0]+D_1[f_1^0]. \tag{15}$$

Here

$$\mathrm{FP}_{11}=\frac{1}{m_1}\frac{\partial}{\partial\mathbf{C}_1}\cdot\int d\mathbf{C}_1'\mathbf{\Phi}^{11}(\mathbf{C}_1-\mathbf{C}_1')\cdot\left(\frac{1}{m_1}\frac{\partial}{\partial\mathbf{C}_1}-\frac{1}{m_1}\frac{\partial}{\partial\mathbf{C}_1'}\right),$$

$$D_1=\frac{n_2^0}{m_1^2}\frac{\partial}{\partial\mathbf{C}_1}\cdot\left(\mathbf{\Phi}^{12}(\mathbf{C}_1)\cdot\frac{\partial}{\partial\mathbf{C}_1}\right).$$

Further, $f_2^0,\rho^0,\rho_1^0,\rho_2^0,\mathbf{c}_0^0,T^0,T_1^0,T_2^0$, and $\overline{\mathbf{C}}_2^0$ show no time variation on the τ_{20} time scale, while

$$\frac{\partial}{\partial t_{20}}(\rho_1^0\overline{\mathbf{C}}_1^0)-n_1^0e_1\overline{\mathbf{C}}_1^0\times\mathbf{H}=-\frac{n_2^0}{m_1}\int d\mathbf{C}_1\mathbf{\Phi}^{12}(\mathbf{C}_1)\cdot\frac{\partial f_1^0}{\partial\mathbf{C}_1}. \tag{16}$$

Consequently, on the τ_{20} time scale only the distribution function to zeroth order for the electrons evolves. The collision side is the sum of two parts. The first takes account of the electron–electron collisions, the second expresses the effect to zeroth order of the collisions between heavy ions and electrons and has the effect of diffusing the electrons in velocity space.

The evolution can be studied through the H_1^0 function

$$H_1^0 = \int d\mathbf{C}_1 f_1^0 \ln f_1^0.$$

By standard procedure, using the properties of Φ^{ij^3}, we find that due to the effect of the FP_{11} term f_1^0 tends toward

$$n_1^0\left(\frac{m_1}{2\pi kT_1^0}\right)^{3/2}\exp\left(-\frac{m_1}{2kT_1^0}(\mathbf{C}_1-\overline{\mathbf{C}}_1^0)^2\right)\quad \text{as } t_{20}\to\infty$$

while the D_1 term has the effect of making this isotropic in velocity space. Thus we obtain

$$f_1^0(\mathbf{C}_1,\mathbf{r},t_{20},t_{21},\ldots)\to f_{1M}^0 = n_1^0\left(\frac{m_1}{2\pi kT_1^0}\right)^{3/2}\exp\left(-\frac{m_1C_1^2}{2kT_1^0}\right)\ \text{as } t_{20}\to\infty, \tag{17}$$

which is a t_{20} independent solution of Eq. (15). (For the electron distribution functions and corresponding velocity moments and later for functions describing ion evolution, subscript "M" will indicate these functions in the limits $t_{20}\to\infty$, and $t_{20}\to\infty, t_{21}\to\infty$, respectively. However, the index is omitted when no confusion is possible).

How f_1^0 evolves toward a Maxwellian distribution is an open question which can only be answered by solving Eq. (15). Su[7] and McLeod and Ong[8] have studied a related equation using Landau's form for Φ^{11} when $\mathbf{H}=0$.

From Eqs. (14) and (17) we get

$$\overline{\mathbf{C}}_2^0=0,\quad \overline{\mathbf{C}}_1^0\to 0\quad \text{as } t_{20}\to\infty. \tag{18}$$

This relaxation can be described by solving Eq. (16).

B. First-order equations

Using zeroth-order results, we get to first order in ϵ

$$\int d\mathbf{C}_1 f_1^0(\mathbf{C}_1)m_1\mathbf{C}_1+\int d\mathbf{C}_2 f_2^1(\mathbf{C}_2)m_2\mathbf{C}_2=0, \tag{19}$$

$$\begin{aligned}\frac{\partial f_1^1}{\partial t_{20}}+\frac{\partial f_1^0}{\partial t_{21}}+\mathbf{C}_1\cdot\frac{\partial f_1^0}{\partial \mathbf{r}}+\mathbf{F}_1\cdot\frac{\partial f_1^0}{\partial \mathbf{C}_1}+\frac{e_1}{m_1}\mathbf{c}_0^0\times\mathbf{H}\cdot\frac{\partial f_1^0}{\partial \mathbf{C}_1}&\\ +\frac{e_1}{m_1}\mathbf{C}_1\times\mathbf{H}\cdot\frac{\partial f_1^1}{\partial \mathbf{C}_1}&\\ =\mathrm{FP}_{11}[f_1^0(\mathbf{C}_1)f_1^1(\mathbf{C}_1')+f_1^1(\mathbf{C}_1)f_1^0(\mathbf{C}_1')]&\\ +\frac{1}{m_1^2}\frac{\partial}{\partial \mathbf{C}_1}\cdot\left[\Phi^{12}(\mathbf{C}_1)\cdot\frac{\partial}{\partial \mathbf{C}_1}(n_2^1 f_1^0+n_2^0 f_1^1)\right],&\end{aligned} \tag{20}$$

$$\begin{aligned}\frac{\partial f_2^1}{\partial t_{20}}+\frac{\partial f_2^0}{\partial t_{21}}-\left(\frac{\partial \mathbf{c}_0^1}{\partial t_{20}}+\frac{\partial \mathbf{c}_0^0}{\partial t_{21}}\right)\cdot\frac{\partial f_2^0}{\partial \mathbf{C}_2}=\mathrm{FP}_{22}[f_2^0(\mathbf{C}_2)f_2^0(\mathbf{C}_2')]&\\ -\frac{1}{m_1m_2}\frac{\partial}{\partial \mathbf{C}_2}\cdot\int d\mathbf{C}_1\Phi^{12}(\mathbf{C}_1)\cdot\frac{\partial}{\partial \mathbf{C}_1}f_1^0 f_2^0.&\end{aligned} \tag{21}$$

Here

$$\mathrm{FP}_{22}=\frac{1}{m_2}\frac{\partial}{\partial \mathbf{C}_2}\cdot\int d\mathbf{C}_2'\Phi^{22}(\mathbf{C}_2-\mathbf{C}_2')\cdot\left(\frac{1}{m_2}\frac{\partial}{\partial \mathbf{C}_2}-\frac{1}{m_2}\frac{\partial}{\partial \mathbf{C}_2'}\right).$$

Further, integrating the macroscopic equations to first order and eliminating all secularities on the τ_{20} time scale show that $\rho^0,\rho_1^0,\rho_2^0,\mathbf{c}_0^0,T^0,T_1^0$, and T_2^0 have no variation on the τ_{21} time scale but may vary on the τ_{22} time scale, ρ^1,ρ_2^1 and $\mathbf{c}_{0\parallel}^1$ (where $\parallel$ means the component parallel to the magnetic field) and T_2^1 have no variation on the τ_{20} time scale, but may vary on the τ_{21} time scale. However, $\rho_1^1, T^1, T_1^1, \mathbf{C}_{0\perp}^1$ (where $\perp$ means the component perpendicular to the magnetic field) and $\overline{\mathbf{C}}_1^1$ all have a transient variation on the τ_{20} time scale because of the variation of f_1^0 (and its moments) on that scale before f_{1M}^0 is established, for instance

$$\rho_1^1(t_{20},t_{21},\cdots)=\rho_1^1(t_{20}=0,t_{21},\cdots)-\frac{\partial}{\partial \mathbf{r}}\cdot\int_0^{t_{20}}d\tau\,\rho_1^0\overline{\mathbf{C}}_1^0(\tau).$$

From the ion kinetic equation, Eq. (21), we get

$$\begin{aligned}f_2^1&=f_2^1(t_{20}=0,t_{21},\cdots)-t_{20}\left(\frac{\partial f_2^0}{\partial t_{21}}-\mathrm{FP}_{22}[f_2^0 f_2^0]\right)\\ &+\int_0^{t_{20}}d\tau\left(\frac{n_1^0e_1}{\rho^0}\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H}\cdot\frac{\partial f_2^0}{\partial \mathbf{C}_2}-\frac{1}{m_1m_2}\right.\\ &\left.\times\frac{\partial f_2^0}{\partial \mathbf{C}_2}\cdot\int d\mathbf{C}_1\Phi^{12}\cdot\frac{\partial f_1^0}{\partial \mathbf{C}_1}\right).\end{aligned}$$

The time integral may be assumed finite. We eliminate the secularity and get

$$\frac{\partial f_2^0}{\partial t_{21}}=\mathrm{FP}_{22}[f_2^0 f_2^0], \tag{22}$$

$$\begin{aligned}f_2^1&=f_2^1(t_{20}=0,t_{21},\cdots)\\ &+\frac{\partial f_2^0}{\partial \mathbf{C}_2}\cdot\int_0^{t_{20}}d\tau\left(\frac{n_1^0e_1}{\rho^0}\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H}\right.\\ &\left.-\frac{1}{m_1m_2}\int d\mathbf{C}_1\Phi^{12}\cdot\frac{\partial f_1^0(\mathbf{C}_1,\tau)}{\partial \mathbf{C}_1}\right).\end{aligned} \tag{23}$$

Equation (22) is the kinetic equation for the ion distribution to zeroth order on the τ_{21} time scale. Using a H theorem and also the *a priori* condition to zeroth order, Eq. (14), we see that

$$f_2^0\to f_{2M}^0=n_2^0\left(\frac{m_2}{2\pi kT_2^0}\right)^{3/2}\exp\left(-\frac{m_2C_2^2}{2kT_2^0}\right)\quad\text{as } t_{21}\to\infty. \tag{24}$$

In this evolution the density n_2^0 and temperature T_2^0 are stationary.

From Eq. (23) we derive

$$\begin{aligned}\rho_2^0\overline{\mathbf{C}}_2^1(t_{20},\cdots)&=\int d\mathbf{C}_2 f_2^1 m_2\mathbf{C}_2\\ &=\int d\mathbf{C}_2 f_2^1(t_{20}=0,t_{21},\cdots)m_2\mathbf{C}_2\\ &-\rho_2^0\int_0^{t_{20}}d\tau\left(\frac{n_1^0e_1}{\rho^0}\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H}\right.\\ &\left.-\frac{1}{m_1m_2}\int d\mathbf{C}_1\Phi^{12}(\mathbf{C}_1)\cdot\frac{\partial f_1^0(\mathbf{C}_1,\tau)}{\partial \mathbf{C}_1}\right).\end{aligned}$$

Now, the *a priori* condition Eq. (19) together with Eq. (16) give

$$\int d\mathbf{C}_1 f_1^0(t_{20}=0,t_{21},\cdots)m_1\mathbf{C}_1+\int d\mathbf{C}_2 f_2^1(t_{20}=0,t_{21},\cdots)m_2\mathbf{C}_2=0$$

and, since $\rho_1^0\overline{\mathbf{C}}_1^0(t_{20}) \to 0$ when $t_{20} \to \infty$, we have

$$\rho_1^0\overline{\mathbf{C}}_1^0(t_{20}=0,t_{21},\cdots) = -\rho_2^0\overline{\mathbf{C}}_2^1(t_{20}=0,t_{21},\cdots) = -\int_0^\infty d\tau\left(n_1^0 e_1\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H} - \frac{n_2^0}{m_1}\int d\mathbf{C}_1\mathbf{\Phi}^{12}(\mathbf{C}_1)\cdot\frac{\partial f_1^0(\mathbf{C}_1,\tau)}{\partial\mathbf{C}_1}\right).$$

Thus

$$\rho_2^0\overline{\mathbf{C}}_2^1(t_{20},\cdots) = \rho_2^0\int_{t_{20}}^\infty d\tau\left(\frac{n_1^0e_1}{\rho^0}\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H} - \frac{1}{m_1m_2}\int d\mathbf{C}_1\mathbf{\Phi}^{12}(\mathbf{C}_1)\cdot\frac{\partial f_1^0(\mathbf{C}_1,\tau)}{\partial\mathbf{C}_1}\right). \quad (25)$$

f_2^1 itself therefore has the following form:

$$f_2^1(t_{20},t_{21},\cdots) = \bar f_2^1(t_{21},\cdots) - \frac{\partial f_2^0}{\partial\mathbf{C}_2}\cdot\int_{t_{20}}^\infty d\tau\left(\frac{n_1^0e_1}{\rho^0}\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H} - \frac{1}{m_1m_2}\int d\mathbf{C}_1\mathbf{\Phi}^{12}(\mathbf{C}_1)\cdot\frac{\partial f_1^0(\mathbf{C}_1,\tau)}{\partial\mathbf{C}_1}\right), \quad (26)$$

where $\bar f_2^1(t_{21},\cdots)$ obeys the condition

$$\int d\mathbf{C}_2\bar f_2^1(t_{21},\cdots)m_2\mathbf{C}_2 = 0. \quad (27)$$

To solve the electron kinetic equation (20) in the limit $t_{20}\to\infty$ we assume that $\partial f_1^1/\partial t_{20} \to 0$ when $t_{20}\to\infty$ (see Appendix B). Assuming further that lim as $t_{20}\to\infty$ commutes with differential and integral operations and using the properties of $\mathbf{\Phi}^{ij}$ and Eq. (17) and the fact that ρ_1^0 and T_1^0 are independent of t_{20} and t_{21}, we get the following linear integro-differential equation for f_{1M}^1:

$$\mathrm{FP}_{11}[f_{1M}^0(\mathbf{C}_1)f_{1M}^1(\mathbf{C}_1') + f_{1M}^1(\mathbf{C}_1)f_{1M}^0(\mathbf{C}_1')] - \frac{e_1}{m_1}\mathbf{C}_1\times\mathbf{H}\cdot\frac{\partial f_{1M}^1}{\partial\mathbf{C}_1} + D_1[f_{1M}^1]$$
$$= \mathbf{C}_1\cdot\frac{\partial f_{1M}^0}{\partial\mathbf{r}} + \mathbf{F}_1\cdot\frac{\partial f_{1M}^0}{\partial\mathbf{C}_1} + \frac{e_1}{m_1}\mathbf{c}_0^0\times\mathbf{H}\cdot\frac{\partial f_{1M}^0}{\partial\mathbf{C}_1}.$$

When we use the explicit expression for f_{1M}^0 and set $f_{1M}^1 = f_{1M}^0\Phi_1^1$, this equation can be written as

$$\mathrm{FP}_{11}[f_{1M}^0(\mathbf{C}_1)f_{1M}^0(\mathbf{C}_1')(\Phi_1^1(\mathbf{C}_1)+\Phi_1^1(\mathbf{C}_1'))] - f_{1M}^0\frac{e_1}{m_1}\mathbf{C}_1\times\mathbf{H}\cdot\frac{\partial\Phi_1^1}{\partial\mathbf{C}_1} + D_1[f_{1M}^0\Phi_1^1]$$
$$= f_{1M}^0\left[\left(\frac{m_1C_1^2}{2kT_1^0}-\frac{5}{2}\right)\frac{1}{T_1^0}\frac{\partial T_1^0}{\partial\mathbf{r}} - \frac{e_1}{kT_1^0}\left(\frac{m_1}{e_1}\mathbf{F}_1 + \mathbf{c}_0^0\times\mathbf{H} - \frac{kT_1^0}{e_1}\frac{\partial}{\partial\mathbf{r}}\ln p_1^0\right)\right]\cdot\mathbf{C}_1. \quad (28)$$

This is similar to the electron kinetic equation used by Robinson and Bernstein.[9]

A solution of Eq. (28) can now be sought in the form

$$\Phi_1^1 = \alpha_1^1 + \gamma_1^1\frac{1}{2}m_1C_1^2 + \mathbf{A}_1(\mathbf{C}_1,\mathbf{H})\cdot\frac{1}{T_1^0}\frac{\partial T_1^0}{\partial\mathbf{r}} + \mathbf{D}_1(\mathbf{C}_1,\mathbf{H})\cdot\left(\frac{m_1}{e_1}\mathbf{F}_1 + \mathbf{c}_0^0\times\mathbf{H} - \frac{kT_1^0}{e_1}\frac{\partial}{\partial\mathbf{r}}\ln p_1^0\right) \quad (29)$$

as in Ref. 1. The constants α_1^1 and γ_1^1 are chosen equal to zero,[1,2] so that

$$\int d\mathbf{C}_1 f_{1M}^1 = n_{1M}^1 = 0, \quad (30)$$

$$\int d\mathbf{C}_1 f_{1M}^1\frac{1}{2}m_1C_1^2 - \frac{3}{2}(n_1kT_1)_m^1 = 0. \quad (31)$$

This implies that we choose properly the *a priori* undetermined initial functions $n_1^1(t_{20}=0,t_{21},\cdots)$ and $T_1^1(t_{20}=0,t_{21},\cdots)$ which show up in the expressions for n_1^1 and T_1^1.

C. Second-order equations

This set of equations are integrated on the τ_{20} as well as on the τ_{21} time scale, and secular terms in each corresponding variable are eliminated. Hence, as nonsecularity conditions when integrating the mass-conservation equations, we obtain, besides the transient terms, the following continuity equations:

$$\frac{\partial\rho^0}{\partial t_{22}} + \frac{\partial}{\partial\mathbf{r}}\cdot(\rho^0\mathbf{c}_0^0) = 0, \quad (32)$$

$$\frac{\partial\rho_2^0}{\partial t_{22}} + \frac{\partial}{\partial\mathbf{r}}\cdot(\rho_2^0\mathbf{c}_0^0) = 0, \quad (33)$$

$$\frac{\partial\rho_1^0}{\partial t_{22}} + \frac{\partial}{\partial\mathbf{r}}\cdot(\rho_1^0\mathbf{c}_0^0) + \frac{\partial}{\partial\mathbf{r}}\cdot(\rho_1^0\overline{\mathbf{C}}_{1M}^1) = 0. \quad (34)$$

As nonsecularity conditions for the mass transport equation we obtain

$$\mathbf{c}_0^2(t_{20},t_{21},\cdots) = \mathbf{c}_0^2(t_{20}=0,t_{21},\cdots) + \int_0^{t_{20}} d\tau\left(\frac{1}{\rho^0}\frac{\partial}{\partial\mathbf{r}}\cdot[\,\mathsf{I}p_1^0 - n_1^0m_1\cdot\overline{\mathbf{C}\mathbf{C}}_1^0(\tau)] + \frac{n_1^0e_1}{\rho^0}(\overline{\mathbf{C}}_1^1(\tau) - \overline{\mathbf{C}}_{1M}^1)\times\mathbf{H} + \frac{n_1^1(\tau)e_1}{\rho^0}\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H}\right), \quad (35)$$

$$\mathbf{c}_0^1(t_{20},t_{21},\cdots) = \mathbf{c}_0^1(t_{20}=0,t_{21}=0,t_{22},\cdots) + \frac{n_1^0e_1}{\rho^0}\int_0^{t_{20}} d\tau\,\overline{\mathbf{C}}_1^0(\tau)\times\mathbf{H} + \int_0^{t_{21}} d\lambda\,\frac{1}{\rho^0}\frac{\partial}{\partial\mathbf{r}}\cdot[\,\mathsf{I}p_2^0 - n_2^0m_2\cdot\overline{\mathbf{C}_2\mathbf{C}_2}^0(\lambda)] \quad (36)$$

$$\rho^0\left(\frac{\partial\mathbf{c}_0^0}{\partial t_{22}} + \mathbf{c}_0^0\cdot\frac{\partial\mathbf{c}_0^0}{\partial\mathbf{r}}\right) = -\sum_{i=1}^2\frac{\partial p_i^0}{\partial\mathbf{r}} + \rho_e^0\mathbf{c}_0^0\times\mathbf{H} + n_1^0e_1\overline{\mathbf{C}}_{1M}^1\times\mathbf{H} + \sum_{i=1}^2\rho_i^0\mathbf{F}_i. \quad (37)$$

Equation (37) should be compared with the corresponding equation in Ref. 1. Equations (35) and (36) represent the two steps of relaxation on the τ_{20} and τ_{21} time scales due to electron and ion evolution, respectively, before the establishment of Eq. (37).

In a similar way the equation for the total temperature shows that T^2 has a variation on the τ_{20} time scale due to electron evolution, T^1 has a variation on the τ_{21} time scale because of ion evolution, and

$$\frac{3}{2}n^0k\left(\frac{\partial T^0}{\partial t_{22}} + \mathbf{c}_0^0 \cdot \frac{\partial T^0}{\partial \mathbf{r}}\right) = \frac{3}{2}kT^0 \frac{\partial}{\partial \mathbf{r}} \cdot (n_1^0\overline{\mathbf{C}}_{1M}^1) + \rho_1^0\mathbf{F}_1 \cdot \overline{\mathbf{C}}_{1M}^1 + n_1^0 e_1 \overline{\mathbf{C}}_{1M}^1 \cdot (\mathbf{c}_0^0 \times \mathbf{H}) - \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{q}_{1M}^1 - p^0 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0^0. \tag{38}$$

The first four terms on the right-hand side of Eq. (38) are due to the electron evolution. This feature is also observed from the equations for the electron and ion temperatures. Corresponding to Eq. (38), we get

$$\frac{3}{2}n_1^0k\left(\frac{\partial T_1^0}{\partial t_{22}} + \mathbf{c}_0^0 \cdot \frac{\partial T_1^0}{\partial \mathbf{r}}\right) = \frac{3}{2}kT_1^0 \frac{\partial}{\partial \mathbf{r}} \cdot (n_1^0\overline{\mathbf{C}}_{1M}^1) + \rho_1^0\mathbf{F}_1 \cdot \overline{\mathbf{C}}_{1M}^1 + n_1^0 e_1 \overline{\mathbf{C}}_{1M}^1 \cdot (\mathbf{c}_0^0 \times \mathbf{H}) - \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{q}_{1M}^1 - p_1^0 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0^0 - \frac{n_1^0 n_2^0}{m_2}\left(\frac{m_1}{2\pi k}\right)^{3/2} \times \left(\frac{T_1^0 - T_2^0}{T_1^{0\;5/2}}\right) \int d\mathbf{C}_1 \mathbf{\Phi}^{12}(\mathbf{C}_1) : \mathsf{I} \exp\left(-\frac{m_1 C_1^2}{2kT_1^0}\right) \tag{39}$$

$$\frac{3}{2}n_2^0k\left(\frac{\partial T_2^0}{\partial t_{22}} + \mathbf{c}_0^0 \frac{\partial T_2^0}{\partial \mathbf{r}}\right) = - p_2^0 \frac{\partial}{\partial \mathbf{r}} \cdot \mathbf{c}_0^0 + \frac{n_1^0 n_2^0}{m_2}\left(\frac{m_1}{2\pi k}\right)^{3/2} \times \left(\frac{T_1^0 - T_2^0}{T_1^{0\;5/2}}\right) \int d\mathbf{C}_1 \mathbf{\Phi}^{12}(\mathbf{C}_1) : \mathsf{I} \exp\left(-\frac{m_1 C_1^2}{2kT_1^0}\right). \tag{40}$$

Here we have rewritten the electron–ion energy exchange term in the limit when t_{20}, t_{21} both are infinite. When using Landau's form for the tensor $\mathbf{\Phi}^{12}$,[10,11] the energy exchange term takes the same form as in Ref. 11.

When we solve the ion kinetic equation to this order we first find that f_2^2 has a transient on the τ_{20} time scale. Let us assume that $\partial f_2^0/\partial t_{22}$ tends toward $\partial f_{2M}^0/\partial t_{22}$ as $t_{21} \to \infty$. In that limit it is also plausible to assume that $\partial f_2^1/\partial t_{21}$ tends toward zero (cf. the Appendix in Ref. 3). Comparison with the expression for $\partial f_2^1/\partial t_{21}$ obtained from Eq. (26) then shows in turn that $\partial \bar{f}_2^1/\partial t_{21}$ tends toward zero. The kinetic equation then reduces to

$$\frac{\partial f_{2M}^0}{\partial t_{22}} + \mathbf{c}_0^0 \cdot \frac{\partial f_{2M}^0}{\partial \mathbf{r}} + \mathbf{C}_2 \cdot \frac{\partial f_{2M}^0}{\partial \mathbf{r}} + \mathbf{F}_2 \cdot \frac{\partial f_{2M}^0}{\partial \mathbf{C}_2} - \left(\frac{\partial \mathbf{c}_0^0}{\partial t_{22}} + \mathbf{c}_0^0 \cdot \frac{\partial \mathbf{c}_0^0}{\partial \mathbf{r}}\right) \cdot \frac{\partial f_{2M}^0}{\partial \mathbf{C}_2} + \frac{e_2}{m_2}\mathbf{c}_0^0 \times \mathbf{H} \cdot \frac{\partial f_{2M}^0}{\partial \mathbf{C}_2} - \frac{\partial f_{2M}^0}{\partial \mathbf{C}_2}\mathbf{C}_2 : \frac{\partial \mathbf{c}_0^0}{\partial \mathbf{r}}$$
$$= \mathrm{FP}_{22}[f_{2M}^0(\mathbf{C}_2)\bar{f}_{2M}^1(\mathbf{C}_2') + \bar{f}_{2M}^1(\mathbf{C}_2)f_{2M}^0(\mathbf{C}_2')] - \frac{1}{m_1 m_2}\frac{\partial f_{2M}^0}{\partial \mathbf{C}_2} \cdot \int d\mathbf{C}_1 \mathbf{\Phi}^{12}(\mathbf{C}_1) \cdot \frac{\partial}{\partial \mathbf{C}_1} f_{1M}^1 + \frac{1}{m_2^2}\frac{\partial^2}{\partial \mathbf{C}_2^2} f_{2M}^0 : \int d\mathbf{C}_1 \mathbf{\Phi}^{12}(\mathbf{C}_1) f_{1M}^0 + \frac{1}{m_1 m_2} \times \frac{\partial}{\partial \mathbf{C}_2} \cdot \left(\mathbf{C}_2 f_{2M}^0 \cdot \int d\mathbf{C}_1 \frac{\partial}{\partial \mathbf{C}_1} \mathbf{\Phi}^{12}(\mathbf{C}_1) \cdot \frac{\partial f_{1M}^0}{\partial \mathbf{C}_1}\right). \tag{41}$$

Taking into account the explicit expression for f_{2M}^0, performing the various differentiations and substituting from Eqs. (33), (40) and the equation similar to Eq. (37) with ρ_2^0 on the left [obtainable from Eq. (11)], some straightforward calculations give

$$\mathrm{FP}_{22}[f_{2M}^0(\mathbf{C}_2)\bar{f}_{2M}^1(\mathbf{C}_2') + \bar{f}_{2M}^1(\mathbf{C}_2)f_{2M}^0(\mathbf{C}_2')] = f_{2M}^0\mathbf{C}_2 \cdot \left(\frac{m_2 C_2^2}{2kT_2^0} - \frac{5}{2}\right)\frac{1}{T_2^0}\frac{\partial T_2^0}{\partial \mathbf{r}} + \frac{m_2}{kT_2^0} f_{2M}^0 \mathring{\mathbf{C}}_2\mathbf{C}_2 : \left(\frac{\partial \mathbf{c}_0^0}{\partial \mathbf{r}} + \frac{1}{m_2 k} \times \frac{T_2^0 - T_1^0}{T_2^0 T_1^0} \int d\mathbf{C}_1 \mathbf{\Phi}^{12}(\mathbf{C}_1) f_{1M}^0\right),$$

where $\mathring{\mathbf{C}}_2\mathbf{C}_2$ is the traceless tensor $\mathbf{C}_2\mathbf{C}_2 - \frac{1}{3}C_2^2\,\mathsf{I}$. It shows that when using for $\mathbf{\Phi}^{12}(\mathbf{C}_1)$ either Landau's form or Balescu–Lénard's form,[10] the velocity integral $\int d\mathbf{C}_1 \mathbf{\Phi}^{12}(\mathbf{C}_1) f_{1M}^0$ is proportional to the unit tensor. Since $\mathring{\mathbf{C}}_2\mathbf{C}_2 : \mathsf{I} = 0$, the kinetic equation further simplifies to

$$\mathrm{FP}_{22}[f_{2M}^0(\mathbf{C}_2)\bar{f}_{2M}^1(\mathbf{C}_2') + f_{2M}^0(\mathbf{C}_2')\bar{f}_{2M}^1(\mathbf{C}_2)] = f_{2M}^0\mathbf{C}_2 \cdot \left(\frac{m_2 C_2^2}{2kT_2^0} - \frac{5}{2}\right)\frac{1}{T_2^0}\frac{\partial T_2^0}{\partial \mathbf{r}} + \frac{m_2}{kT_2^0} f_{2M}^0 \mathring{\mathbf{C}}_2\mathbf{C}_2 : \frac{\partial \mathbf{c}_0^0}{\partial \mathbf{r}}. \tag{42}$$

This equation has to be solved for $\bar{f}_{2M}^1$ subject to the condition Eq. (27). The solution has the following form[1]:

$$\bar{f}_{2M}^1 = f_{2M}^0\left(\alpha_2^1 + \boldsymbol{\beta}_2^1 \cdot m_2\mathbf{C}_2 + \gamma_2^1 \frac{1}{2}m_2 C_2^2 + \mathbf{A}_2 \cdot \frac{1}{T_2^0}\frac{\partial T_2^0}{\partial \mathbf{r}} + \mathsf{B}_2 : \frac{\partial \mathbf{c}_0^0}{\partial \mathbf{r}}\right), \tag{43}$$

where $\mathbf{A}_2 = \mathbf{C}_2\mathcal{A}_2(C_2)$, $\mathsf{B}_2 = \mathring{\mathbf{C}}_2\mathbf{C}_2\mathcal{B}_2(C_2)$, and $\alpha_2^1, \boldsymbol{\beta}_2^1$, and γ_2^1 are constant. $\boldsymbol{\beta}_2^1$ is determined by Eq. (27) while α_2^1 and γ_2^1 are chosen equal to zero,[1,2], so that

$$\int d\mathbf{C}_2 f_{2M}^1 = n_{2M}^1 = 0, \tag{44}$$

$$\int d\mathbf{C}_2 f_{2M}^1 \frac{1}{2} m_2 C_2^2 = \frac{3}{2}(n_2 kT_2)_M^1 = 0, \tag{45}$$

which implies that the *a priori* undetermined initial functions in n_2^1 and T_2^1 at $t_{21} = 0$ are properly chosen.

D. Higher-order equations

It can easily be shown that the electron kinetic equation to second order in the limit $t_{20} \to \infty$ gives an integro-differential equation for f_{1M}^2 of the same form as Eq. (28) for f_{1M}^1, the operators in both equations being the same. To third order in ϵ the ion kinetic equation in the limit $t_{20} \to \infty$ and $t_{21} \to \infty$ reduces to an equation for f_{2M}^2 of the same form as Eq. (42) for f_{2M}^1. The equation for f_{2M}^2 has to be solved subject to the *a priori* condition to second order,

$$(\rho_1\overline{\mathbf{C}}_1)^1 + (\rho_2\overline{\mathbf{C}}_2)^2 = 0. \tag{46}$$

This is possible since the general solution for f_{2M}^2 contains five arbitrary constants. Two of these can be chosen equal to zero as in C. The remaining three are chosen to fulfil the *a priori* condition Eq. (46).

In principle this procedure can be followed to even higher orders, the electron kinetic equations in the limit $t_{20} \to \infty$ and the ion kinetic equations in the limits $t_{20} \to \infty$ and $t_{21} \to \infty$ having the same operators as before.

IV. SOME QUALITATIVE SOLUTIONS

To obtain qualitative solutions of Eqs. (28) and (42), we use the relaxation terms

$$\nu_1(f_{1M}^0 - f_1), \qquad \alpha\nu_2(f_{2M}^0 - f_2)$$

in the electron and ion kinetic equations respectively instead of the Fokker–Planck and diffusion collision terms. Here ν_1 is a measure of the electron–electron and electron–ion collision frequency and ν_2 is a measure of the ion–ion collision frequency. They are both assumed velocity, time and space independent. In the limit $t_{20} \to \infty$ and in the limit $t_{21} \to \infty$ we then get

$$f_{1M}^1 = f_{1M}^0\left[\frac{1}{\nu_1}\mathbf{C}_{1\parallel} + \frac{1/\nu_1}{1+(\Omega_1/\nu_1)^2}\times\left(\mathbf{C}_{1\perp} - \frac{\Omega_1}{\nu_1 H}\mathbf{C}_1\times\mathbf{H}\right)\right]\cdot\left[-\left(\frac{m_1C_1^2}{2kT_1} - \frac{5}{2}\right)\frac{1}{T_1}\frac{\partial T_1}{\partial \mathbf{r}} + \frac{e_1}{kT_1}\left(\frac{m_1}{e_1}\mathbf{F}_1 + \mathbf{c}_0^0\times\mathbf{H} - \frac{kT_1}{e_1}\frac{\partial}{\partial\mathbf{r}}\ln p_1\right)\right], \tag{47}$$

$$f_{2M}^1(t_{22}, \mathbf{C}_2) = \frac{1}{\nu_2}f_{2M}^0\left[-\mathbf{C}_2\cdot\left(\frac{m_2C_2^2}{2kT_2} - \frac{5}{2}\right)\frac{1}{T_2}\frac{\partial T_2}{\partial\mathbf{r}} - \frac{m_2}{kT_2}\mathbf{C}_2^\circ\mathbf{C}_2 : \frac{\partial \mathbf{c}_0^0}{\partial\mathbf{r}}\right]. \tag{48}$$

These expressions are of the same form as in Ref. 1. From the expression for f_{1M}^1 we find $\overline{\mathbf{C}}_{1M}^1$. We obtain, parallel and perpendicular to the magnetic field,

$$n_1\overline{\mathbf{C}}_{1\parallel}^1 = \frac{n_1e_1}{m_1\nu_1}\left(\frac{m_1}{e_1}\mathbf{F}_1 - \frac{kT_1}{e_1}\frac{\partial}{\partial\mathbf{r}}\ln p_1\right)_\parallel \tag{49}$$

$$n_1\overline{\mathbf{C}}_{1\perp}^1 = \frac{1/\nu_1}{1+(\Omega_1/\nu_1)^2}\left[\frac{n_1e_1}{m_1}\left(\frac{m_1}{e_1}\mathbf{F}_1 + \mathbf{c}_0^0\times\mathbf{H} - \frac{kT_1}{e_1}\frac{\partial}{\partial\mathbf{r}}\ln p_1\right)_\perp - \frac{n_1e_1}{m_1}\times\frac{\Omega_1}{\nu_1}\frac{\mathbf{H}}{H}\times\left(\frac{m_1}{e_1}\mathbf{F}_1 + \mathbf{c}_0^0\times\mathbf{H} - \frac{kT_1}{e_1}\frac{\partial}{\partial\mathbf{r}}\ln p_1\right)\right]. \tag{50}$$

Setting now $\partial/\partial\mathbf{r} = 0$, we substitute Eqs. (49) and (50) into Eqs. (37)–(40). The mass-transport equations parallel and perpendicular to the magnetic field then become

$$\rho\frac{\partial \mathbf{c}_{0\parallel}^0}{\partial t_{22}} = \sum_{i=1}^2 \rho_i\mathbf{F}_{i\parallel}, \tag{51}$$

$$\rho\frac{\partial\mathbf{c}_{0\perp}^0}{\partial t_{22}} = \left(\rho_e - \delta\frac{\Omega_1 H}{\nu_1}\right)\mathbf{c}_0^0\times\mathbf{H} - \delta H^2\mathbf{c}_{0\perp}^0 + \delta\frac{m_1}{e_1}\mathbf{F}_1\times\mathbf{H} - \delta\frac{\Omega_1 H}{\nu_1}\frac{m_1}{e_1}\mathbf{F}_{1\perp} + \sum_{i=1}^2\rho_i\mathbf{F}_{i\perp}, \tag{52}$$

Here

$$\delta = \frac{n_1e_1^2}{m_1}\frac{1/\nu_1}{1+(\Omega_1/\nu_1)^2} > 0.$$

Equation (38) gives in the same way the following equation for temperature:

$$\frac{3}{2}nk\frac{\partial T}{\partial t_{22}} = \frac{n_1e_1^2}{m_1}\frac{1/\nu_1}{1+(\Omega_1/\nu_1)^2}\left(\frac{m_1\mathbf{F}_{1\perp}}{e_1} + \mathbf{c}_0^0\times\mathbf{H}\right)^2 + \frac{n_1m_1}{\nu_1}\mathbf{F}_{1\parallel}^2. \tag{53}$$

As for a neutral gas the mass transport parallel to the magnetic field grows linearly with time. Equation (52) is a linear inhomogeneous equation for $\mathbf{c}_{0\perp}^0$. A simple calculation shows that the corresponding homogeneous equation has solutions proportional to $e^{\pm i\omega_\pm t}$, where $\omega_\pm = \eta \pm i\,\delta H^2/\rho$.

Here

$$\eta = (H/\rho)(\rho_e - \delta\Omega_1 H/\nu_1).$$

These solutions are damped away in a time of order $\rho/\delta H^2$, i.e., on the τ_{22} time scale. The steady state solution of Eq. (52) can be found by setting the left-hand side of Eq. (52) equal to zero.

Equation (53) shows that after a time of order $\rho/\delta H^2$ the temperature grows linearly with time. Equations (39) and (40) when $\partial/\partial\mathbf{r} = 0$ show how energy produced by forces is absorbed by electrons and via collisions partly released to ions provided that $T_1 > T_2$.

Equation (47) for the homogeneous case shows how f_{1M}^1 tends toward a steady value after a time of order $\rho/\delta H^2$.

In absence of forces $\mathbf{F}_i$ also, however, $\mathbf{c}_{0\perp}^0$ is damped away after a time of order $\rho/\delta H^2$, while T reaches a stationary value and $f_{1M}^1(t_{22}, \cdots)$ dies away; that is, a state of equilibrium is established.

ACKNOWLEDGMENT

We thank the referee for suggesting us to comment on the generation of fields in the plasma. (See Appendix A.)

APPENDIX A

Taking into account also an electromagnetic field generated by the evolution of the plasma, we have to consider the kinetic and moment equations coupled with the Maxwell equations. Assuming almost charge neutrality an electric field $\mathbf{E}$ of the same order of magnitude as in Sec. II may be generated as well as a weak magnetic field $\mathbf{B}$. Appropriate zeroth- and first-order Maxwell equations may be written down and solved together with the equations of Sec. III. It shows that the zeroth-order kinetic and moment equations are unaltered. From the first-order kinetic and moment equations and zeroth-order Maxwell equations we see new transients. New transients also appear at the next order of approximation. However, in the limits $t_{20} = \infty$ and $t_{21} = \infty$, i.e., on the τ_{22} time scale, the kinetic and moment equations are the same as in Sec. III. The electric field obeys Poisson equation, the source of which varies on the τ_{22} time scale.

APPENDIX B

The assumption that $\partial f_1^1/\partial t_{20} \to 0$ as $t_{20} \to \infty$ in Eq. (20) can be made plausible in the same way as in the

Appendix of Ref. 3. By defining the operators $\mathfrak{F}$, $\mathfrak{M}$, and $\mathfrak{g}$ by

$$\mathfrak{F}\varphi_1 = \mathrm{FP}_{L11}\varphi_1 + \frac{n_2^0}{m_1}\frac{\partial}{\partial \mathbf{C}_1}\cdot\left(\Phi^{12}\cdot\frac{\partial}{\partial \mathbf{C}_1}\varphi_1\right),$$

$$\mathfrak{M}\varphi_1 = -\frac{e_1}{m_1}\mathbf{C}_1 \times \mathbf{H}\cdot\frac{\partial}{\partial \mathbf{C}_1}\varphi_1,$$

$$\mathfrak{g} = \mathfrak{F} + \mathfrak{M},$$

where FP_{L11} is the linearized electron–electron Fokker–Planck operator, Eq. (20) can be written as

$$\frac{\partial \varphi_1}{\partial t_{20}} = \mathfrak{g}\varphi_1 + G_1, \tag{B1}$$

where

$$f_1^1 = f_{1M}^0(\varphi_{1M} + \varphi_1).$$

φ_{1M} is a t_{20}-independent solution of

$$f_{1M}^0\,\mathfrak{g}\varphi_{1M} = \left[\mathbf{C}_1\cdot\frac{\partial}{\partial \mathbf{r}} + \mathbf{F}_1\cdot\frac{\partial}{\partial \mathbf{C}_1} + \frac{e_1}{m_1}\mathbf{c}_0^0 \times \mathbf{H}\cdot\frac{\partial}{\partial \mathbf{C}_1}\right]f_{1M}^0$$

and G_1 a quantity which tends toward zero as $t_{20}\to\infty$. $\mathfrak{F}$ is a symmetric negative operator on $L^2(f_{1M}^0)$ space, while $\mathfrak{M}$ is antisymmetric. [$(\cdot,\cdot)$ and $\|\ \|$ denote scalar product and norm in this space.] It follows that $(\mathfrak{g}\varphi_1, \varphi_1)$ equals zero if and only if $(\mathfrak{F}\varphi_1, \varphi_1)$ equals zero, i.e., $\varphi_1 = \alpha + \gamma C_1^2$, where α and γ are $\mathbf{C}_1$ independent. Since $\mathfrak{M}$ is an infinitesimal rotation operator, it can be shown as in Ref. 12 that $\mathfrak{F}\mathfrak{M} = \mathfrak{M}\mathfrak{F}$, so that $(\mathfrak{F}\varphi_1, \mathfrak{M}\varphi_1) = 0$. It follows then from Eq. (B1) that

$$\frac{\partial}{\partial t_{20}}\frac{1}{2}(\mathfrak{g}\varphi_1, \varphi_1) = \frac{\partial}{\partial t_{20}}\frac{1}{2}(\mathfrak{F}\varphi_1, \varphi_1) = \|\mathfrak{F}\varphi_1\|^2 + (\mathbf{G}_1, \mathfrak{F}\varphi_1)$$

and the proof is completed in the same way as in Ref. 3.

[1]S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge U. P., Cambridge, 1958), Chap. 18.

[2]J. E. McCune, T. F. Morse, and G. Sandri, in *Rarefied Gas Dynamics*, edited by J. A. Laurman (Academic, New York, 1963), Vol. 1, p. 115.

[3]A. H. Øien and J. Naze Tjøtta, Phys. Fluids **11**, 2373 (1971).

[4]E. A. Frieman, J. Math. Phys. **4**, 410 (1963).

[5]G. Sandri, Ann. Phys. (N.Y.) **24**, 332 (1963); Ann. Phys. (N.Y.) **24**, 380 (1963).

[6]C. H. Su, J. Math. Phys. **5**, 1273 (1964).

[7]C. H. Su, J. Math. Phys. **8**, 248 (1967).

[8]J. B. McLeod and R. S. B. Ong, J. Math. Phys. **8**, 240 (1967).

[9]B. B. Robinson and I. B. Bernstein, Ann. Phys. (N.Y.) **18**, 110 (1962).

[10]D. C. Montgomery and D. A. Tidman, *Plasma Kinetic Theory* (McGraw-Hill, New York, 1964).

[11]L. Landau, Phys. Zh. Sovjetunion **10**, 154 (1936).

[12]J. Naze, C.R. Acad. Sci. (Paris) **251**, 651 (1960); see also *Publications Scientifiques et Techniques du Ministère de l'Air, n° 387*, Paris 1962.

Coupled channel operator approach to e–H scattering

Michael Baer*†‡ and Donald J. Kouri

Department of Chemistry, University of Houston, Houston, Texas 77004
(Received 29 July 1971; revised manuscript received 27 January 1972)

A coupled channel T operator formalism is applied to the problem of e–H collisions. Due to the electron–electron repulsion interaction, coupled integral equations for the direct and exchange amplitude density functions are obtained in contrast to the purely algebraic equations encountered in previous studies of model systems. These integral equations may be solved in a number of ways. One procedure involves the use of the Sams–Kouri homogeneous integral solution formalism to convert the integral equations into Volterra equations of the second kind. The kernals for these equations are such that a very rapid numerical solution may be obtained. Numerical results are presented for the special case of s-wave elastic scattering.

I. INTRODUCTION

In the years since the development of quantum mechanical scattering theory, there has been great interest in collisions in which identical particles are present.[1] The simplest example of such a system is, of course, the e-H atom system.[2] Because of the requirements of the Pauli exclusion principle, it is immediately necessary to treat rearrangements whenever one deals with scattering of systems containing identical Fermions.[1a] By far, the most widely studied procedure for such problems is the adaptation of the Hartree-Fock procedure to collisions.[2] In addition, more direct applications of the variation principle to such collisions have resulted in very accurate values of phase shifts for the elastic scattering of electrons by H atoms at energies below threshold for excitation [2] Another procedure is generally discussed in various reference texts but has been applied in practice only in perturbation calculations (i.e., the Born approximation[1(a)(b)(d)]).

This is the procedure based on computing the direct and exchange T matrix elements and then computing the singlet and triplet amplitudes by linear combinations of the exchange and direct amplitudes. If the variational and T matrix calculations both are done exactly, one should get identical results. However, approximate calculations based on the various different formalisms may be expected to lead to different results. In the present paper, we present a discussion and application of nonperturbative scheme for computing the direct and exchange T matrices for e-H collisions. The procedure involves the use of coupled operator equations for the channel operators, and it has been previously used to study models for three body rearrangements.[3] In those studies, the interactions are such that the final equations which must be solved are purely algebraic equations. In contrast, e-H scattering involves the electron-electron repulsion, and it is shown that this leads to somewhat different integral equations. However, it is still possible to obtain numerical solutions to the coupled integral equations by application of the homogeneous integral solution procedure of Sams and Kouri.[3] In order to illustrate the approach, numerical calculations are carried out for the s-wave component of elastic scattering. Although the present application is restricted to energies below threshold for excitation, the formalism is general (if ionization processes are neglected).

In Section II of this paper we present the formalism for the e-H system using channel operators. In Section III the specific case of s-wave elastic scattering is dealt with and numerical results given.

II. COUPLED CHANNEL OPERATOR FORMALISM FOR e-H COLLISIONS

The channel operators which are employed in this discussion are given by[1a,c,3]

$$\tau_{\gamma\alpha} = V_\alpha + V_\gamma (E - H + i\epsilon)^{-1} V_\alpha \tag{1}$$

where α is an initial configuration channel index and γ any possible final configuration channel index. For the e-H problem, $\tau_{\alpha\alpha}$ is the direct scattering operator and $\tau_{\beta\alpha}$ is the exchange scattering operator. The interesting feature of these operators is that they permit one to treat the electrons as distinguishable during the computation and antisymmetrization is then performed after the calculation of the T matrix elements. If we therefore label one electron as 1 and the other as 2, and assume electron 1 is initially bound and 2 is free, then

$$V_\alpha = -2/r_2 + 2/r_{12} \tag{2}$$

and

$$V_\beta = -2/r_1 + 2/r_{12}. \tag{3}$$

Here we measure the energy in units of Rydbergs. Equation (1) for $\tau_{\gamma\alpha}$ is not the most convenient form since the full Green's operator $(E - H + i\epsilon)^{-1}$ cannot be known. We therefore introduce quantities $W_{\gamma\gamma'}$ such that

$$\sum_{\gamma'} W_{\gamma\gamma'} = 1 \tag{4}$$

and employ the well known identity[1]

$$(E - H + i\epsilon)^{-1} = (E - K_{\gamma'} + i\epsilon)^{-1}[1 + V_{\gamma'}(E - H + i\epsilon)^{-1}]. \tag{5}$$

It follows that Eq. (1) may be written as

$$\tau_{\gamma\alpha} = V_\alpha + \sum_{\gamma'} V_\gamma W_{\gamma\gamma'}(E - K_{\gamma'} + i\epsilon)^{-1}\tau_{\gamma'\alpha}. \tag{6}$$

Obviously there are infinitely many choices for the $W_{\gamma\gamma'}$.

1. $W_{\gamma\gamma'} = \delta_{\gamma\gamma'}$, where $\delta_{\gamma\gamma'}$, is the usual Kronnecker delta for γ, γ'.

This choice leads to uncoupled channel operator equations given by

$$\tau_{\gamma\alpha} = V_\alpha + V_\gamma (E - K_\gamma + i\epsilon)^{-1}\tau_{\gamma\alpha}. \tag{7}$$

These equations are very attractive since they enable one to avoid difficulties associated with nonlocal "potentials" arising due to exchange. That is, one can specify the scattering coordinate in each channel separately as that most suited to the particular final configuration channel. This equation has been employed by Sams and Kouri[3] to treat e-H collisions. However the results obtained do not agree at all with calculations using other formalisms.[2] Subsequently, the present authors have applied these equations to an analytically soluable model for a three body rearrangement and found the results in complete disagreement with the well-known correct results.[4] Essentially, the continuum contributions to the

Copyright © 1973 by the American Institute of Physics

Green's operator $(E - K_\gamma + i\epsilon)^{-1}$ play a fundamental role in these uncoupled channel operator equations so that incorrect results are obtained if these contributions are ignored or approximated by a quadrature scheme. Since in the present study we have chosen to neglect the continuum contributions associated with these Green's operators, we have not utilized this choice for $W_{\gamma\gamma'}$.

2. $W_{\gamma\gamma'} = (1 - \delta_{\gamma\gamma'})/(N-1)$ where N is the number of channels. For example in the present two channel problem (neglecting ionization channels) we have

$$W_{\gamma\gamma'} = 1 - \delta_{\gamma\gamma'}. \tag{8}$$

Combining Eqs. (6) and (8), we obtain the coupled channel operator equations

$$\begin{aligned}\tau_{\alpha\alpha} &= V_\alpha + V_\alpha(E - K_\beta + i\epsilon)^{-1}\tau_{\beta\alpha}\\ \tau_{\beta\alpha} &= V_\alpha + V_\beta(E - K_\alpha + i\epsilon)^{-1}\tau_{\alpha\alpha}.\end{aligned} \tag{9}$$

In our previous study of the soluable model for three body rearrangements, these equations were found to yield correct results even though contributions from the continuum portions of the Green's operators were neglected.[4] This is probably the result of the direct coupling between the channels so that the effect of flux entering the α channel is directly reflected on the flux entering the β channel. In the case of Eq. (7), the information about flux in other channels enters solely through the continuum portion of $(E - K_\gamma + i\epsilon)^{-1}$. Consequently, it was decided to base our calculations for e-H scattering on the coupled channel operator equations given by Eq. (9).

In order to develop the equations which are used in actual calculations, it is convenient to introduce the so-called "total amplitude density" functions $\zeta_{\gamma\alpha}(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2)$ defined by[5]

$$\zeta_{\gamma\alpha}(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2) = \tau_{\gamma\alpha}\varphi(nJl_1l_2|\mathbf{r}_1,\mathbf{r}_2). \tag{10}$$

Here $\varphi(nJl_1l_2|\mathbf{r}_1,\mathbf{r}_2)$ is an initial state wave function in the coupled angular momentum representation. Thus, n is the bound electron principle quantum number, J is the total angular momentum quantum number, l_1 is the bound electron orbital angular momentum quantum number and l_2 is the relative orbital angular momentum quantum number for the free electron. In this representation the total J noninteracting Green's function $G_\alpha(\mathbf{r}_1,\mathbf{r}_2|\mathbf{r}_1',\mathbf{r}_2')$ is given by[1]

$$G_\alpha(\mathbf{r}_1,\mathbf{r}_2|\mathbf{r}_1',\mathbf{r}_2') = -i\sum_n\sum_{l_1}\sum_{l_2}\frac{1}{k_n}\frac{j(l_2|k_nr_{2<})h^1(l_2|k_nr_{2>})}{r_{2<}r_{2>}}$$
$$\varphi_{nl_1}(r_1)\varphi_{nl_1}(r_1')Y_{Jl_1l_2}(\hat{r}_1,\hat{r}_2)Y^*_{Jl_1l_2}(\hat{r}_1',\hat{r}_2') \tag{11}$$

where $j(l_2|k_nr)$ is the Ricatti-Bessel function and $h^1(l_2|k_nr)$ is the Ricatti-Hankel function of the first kind,[1] $\varphi_{nl_1}(r)$ is the radial portion of the bound electron wave function and $Y_{Jl_1l_2}(\hat{r}_1,\hat{r}_2)$ is the coupled angular state where since the scattering is independent of the projection of total angular momentum along the Z axis, we suppress the dependence on a magnetic quantum number M). The wave number k_n for radial motion is given by

$$k_n^2 = E + 1/n^2 \tag{12}$$

where E is the total energy expressed in Rydbergs.

Using Equations (8)-(11) we may write

$$\begin{aligned}\zeta_{\alpha\alpha}&(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2)\\ &= V_\alpha(\mathbf{r}_1,\mathbf{r}_2)\varphi(nJl_1l_2|\mathbf{r}_1,\mathbf{r}_2) - iV_\alpha(\mathbf{r}_1,\mathbf{r}_2)\\ &\times\sum_{n'}\sum_{l_1'}\sum_{l_2'}\frac{1}{k_{n'}}\int d\mathbf{r}_1'\int d\mathbf{r}_2'\frac{j(l_2'|k_{n'}r_{1<})h^1(l_2'|k_{n'}r_{1>})}{r_{1<}r_{1>}}\\ &\times\varphi_{n'l_1'}(r_2)\varphi_{n'l_1'}(r_2')Y_{Jl_1'l_2'}(\hat{r}_2,\hat{r}_1)\\ &\times Y^*_{Jl_1'l_2'}(\hat{r}_2',\hat{r}_1')\zeta_{\beta\alpha}(J|nl_1l_2|\mathbf{r}_1',\mathbf{r}_2')\end{aligned} \tag{13}$$

and

$$\begin{aligned}\zeta_{\beta\alpha}&(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2) = V_\alpha(\mathbf{r}_1,\mathbf{r}_2)\varphi(nJl_1l_2|\mathbf{r}_1,\mathbf{r}_2)\\ &- iV_\beta(\mathbf{r}_1,\mathbf{r}_2)\sum_{n'}\sum_{l_1'}\sum_{l_2'}\frac{1}{k_{n'}}\\ &\times\int d\mathbf{r}_1'\int d\mathbf{r}_2'\frac{j(l_2'|k_{n'}r_{2<})h^1(l_2'|k_{n'}r_{2>})}{r_{2<}r_{2>}}\varphi_{n'l_1'}(r_1)\varphi_{n'l_1'}(r_1')\\ &\times Y_{Jl_1'l_2'}(\hat{r}_1,\hat{r}_2)Y^*_{Jl_1'l_2'}(\hat{r}_1',\hat{r}_2')\zeta_{\alpha\alpha}(J|nl_1l_2|\mathbf{r}_1',\mathbf{r}_2').\end{aligned} \tag{14}$$

(Here explicit use is made of the well known fact that no coupling in J occurs).

We now expand the functions $\zeta_{\gamma\alpha}(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2)$ in the basis set $\varphi_{nl_1}(r_\gamma)Y_{Jl_1l_2}(\hat{r}_\gamma,\hat{r}_{\gamma'})$ and also expand $1/r_{12}$ in the usual multipole series so that

$$\begin{aligned}\zeta_{\alpha\alpha}(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2) &= \sum_{n'}\sum_{l_1'}\sum_{l_2'}\varphi_{n'l_1'}(r_1)Y_{Jl_1'l_2'}(\hat{r}_1,\hat{r}_2)\\ &\times\zeta_{\alpha\alpha}(J|n'l_1'l_2'|nl_1l_2|r_2)/k_nr_2,\end{aligned} \tag{15}$$

$$\begin{aligned}\zeta_{\beta\alpha}(J|nl_1l_2|\mathbf{r}_1,\mathbf{r}_2) &= \sum_{n'}\sum_{l_1'}\sum_{l_2'}\varphi_{n'l_1'}(r_2)Y_{Jl_1'l_2'}(\hat{r}_2,\hat{r}_1)\\ &\times\zeta_{\beta\alpha}(J|n'l_1'l_2'|nl_1l_2|r_1)/k_nr_1\end{aligned} \tag{16}$$

and

$$1/r_{12} = \sum_l\frac{r_<^l}{r_>^{l+1}}P_l(\cos\theta_{12}). \tag{17}$$

If we substitute Eqs. (15)-(17) into Eqs. (13)-(14) and multiply each by the proper functions $\varphi_{nl_1}Y_{Jl_1l_2}$ and integrate, we obtain

$$\begin{aligned}\zeta_{\alpha\alpha}(J|\lambda'|\lambda|r_2) &= \left[\int d\mathbf{r}_1\int d\hat{r}_2Y^*_{Jl_1'l_2'}(\hat{r}_1,\hat{r}_2)\varphi_{n'l_1'}(r_1)\right.\\ &\times \left.V_\alpha(\mathbf{r}_1,\mathbf{r}_2)Y_{Jl_1l_2}(\hat{r}_1,\hat{r}_2)\varphi_{nl_1}(r_1)\right]j(l_2|k_nr_2)\\ &- i\sum_{\substack{n''l_1''\\l_2''}}\frac{r_2}{k_{n''}}\left[\int d\mathbf{r}_1\int d\hat{r}_2Y^*_{Jl_1'l_2'}(\hat{r}_1,\hat{r}_2)\frac{\varphi_{n'l_1'}(r_1)V_\alpha(\mathbf{r}_1,\mathbf{r}_2)}{r_1}\right.\\ &\times Y_{Jl_1''l_2''}(\hat{r}_1\hat{r}_2)\int_0^\infty dr_1'j(l_2''|k_{n''}r_{1<})h^1(l_2''|k_{n''}r_{1>})\\ &\times\left.\zeta_{\beta\alpha}(J|\lambda''|\lambda|r_1')\right]\varphi_{n''l_1''}(r_2)\end{aligned} \tag{18}$$

and

$$\begin{aligned}\zeta_{\beta\alpha}(J|\lambda'|r_1) &= \left[\int d\mathbf{r}_2\int d\hat{r}_1Y^*_{Jl_1'l_2'}(\hat{r}_2,\hat{r}_1)\varphi_{n'l_1'}(r_2)\right.\\ &\times\left.V_\alpha(\mathbf{r}_1,\mathbf{r}_2)Y_{Jl_1l_2}(\hat{r}_1,\hat{r}_2)j\frac{(l_2|k_nr_2)}{r_2}\right]r_1\varphi_{nl_1}(r_1)\\ &- i\sum_{\substack{n''l_1''\\l_2''}}\frac{r_1}{k_{n''}}\left[\int d\mathbf{r}_2\int d\hat{r}_1Y^*_{Jl_1'l_2'}(\hat{r}_2,\hat{r}_1)\frac{\varphi_{n'l_1'}(r_2)}{r_2}\right.\\ &\times\varphi_{n'l_1'}(r_2)V_\beta(\mathbf{r}_1,\mathbf{r}_2)Y_{Jl_1''l_2''}(\hat{r}_1,\hat{r}_2)\end{aligned}$$

$$\times \int_0^\infty dr_2' j(l_2''|k_{n''}r_{2<})h^1(l_2''|k_{n''}r_{2>})$$
$$\times \zeta_{\alpha\alpha}(J|\lambda''|\lambda|r_2') \quad \varphi_{n''l_1''}(r_1)\Big], \tag{19}$$

where λ represents the set of quantum numbers $(nl_1 l_2)$. The angular integrals appearing in Eqs. (18) and (19) have been dealt with by Percival and Seaton[6] in their discussion of the Hartree-Fock equations for e-H collisions and are shown to be

$$\int d\hat{r}_1 \int d\hat{r}_2 Y^*_{Jl_a l_b}(\hat{r}_1,\hat{r}_2) P_\gamma(\hat{r}_1,\hat{r}_2) Y_{Jl_c l_d}(\hat{r}_1,\hat{r}_2)$$
$$\equiv \langle l_a l_b J | P_\gamma(\hat{r}_1\cdot\hat{r}_2)| l_c l_d J\rangle$$
$$= \frac{(-1)^{l_a+l_c-J}}{(2\gamma+1)} C(l_a l_c \gamma|000) C(l_b l_d \gamma|000)[(2l_a+1)$$
$$\times (2l_b+1)(2l_c+1)(2l_d+1)]^{1/2} W(l_a l_b l_c l_d; J\gamma) \tag{20}$$

where $C(ll'\gamma|000)$ are Clebsch-Gordan coefficients and $W(l_a l_b l_c l_d; J\gamma)$ is the Racah W coefficient. We follow Percival and Seaton[6] in defining

$$f_\gamma(l_1'l_2', l_1''l_2''; J) = \langle l_1'l_2'J|P_\gamma(\hat{r}_1\cdot\hat{r}_2)|l_1''l_2''J\rangle, \tag{21}$$

and

$$g_\gamma(l_1'l_2', l_1''l_2''; J) = (-1)^{l_1'+l_2'}\langle l_1'l_2'J|P_\gamma(\hat{r}_1\cdot\hat{r}_2)|l_2''l_1''J\rangle \tag{22}$$

and we note that

$$f_\gamma(l_1'l_2', l_1''l_2''; J) = f_\gamma(l_1''l_2'', l_1'l_2'; J) \tag{23}$$

and

$$g_\gamma(l_1'l_2', l_1''l_2''; J) = g_\gamma(l_1''l_2'', l_1'l_2'; J). \tag{24}$$

It follows that we may write Eqs. (18)-(19) as

$$\zeta_{\alpha\alpha}(J|\lambda'|\lambda|r_2) = j(l_2|k_n r_2)[-2\delta_{\lambda\lambda'}/r_2$$
$$+ 2\sum_l f_l(l_1'l_2', l_1 l_2; J)$$
$$\times \int_0^\infty dr_1 \varphi_{n'l'}(r_1)(r_<^l/r_>^{l+1})\varphi_{nl}(r_1) r_1^2]$$
$$- i\sum_{\substack{n''l_1''\\ l_2''}} \frac{r_2}{k_{n''}}\Big[\varphi_{n''l_1''}(r_2) - \frac{2\delta_{l_1'l_2''}\delta_{l_1''l_2'}}{r_2}\int_0^\infty dr_1 r_1 \varphi_{n'l_1'}(r_1)$$
$$\times \int_0^\infty dr_1' j(l_2''|k_{n''}r_{1<})\, h^1(l_2''|k_{n''}r_{1>})\zeta_{\beta\alpha}(J|\lambda''|\lambda|r_1')$$
$$+ 2\sum_l (-1)^{l_1'+l_2'} g_l(l_1'l_2', l_1''l_2''; J)$$
$$\times \int_0^\infty dr_1 r_1 \varphi_{n'l_1'}(r_1)(r_<^l/r_>^{l+1})\int dr_1' j(l_2''|k_{n''}r_{1<})$$
$$\times h^1(l_2''|k_{n''}r_{1>})\zeta_{\beta\alpha}(J|\lambda''|\lambda|r_1')\Big] \tag{25}$$

and

$$\zeta_{\beta\alpha}(J|\lambda'|\lambda|r_1) = r_1\varphi_{nl_1}(r_1)[-2\delta_{l_1'l_2}\delta_{l_2'l_1}$$
$$\times\int_0^\infty dr_2 \varphi_{n'l_1'}(r_1) j(l_2|k_n r_2) + 2\sum_l (-1)^{l_1'+l_2'} g_l(l_1'l_2', l_1 l_2; J)$$
$$\times \int_0^\infty dr_2 r_2 \varphi_{n'l_1'}(r_2)(r_<^l/r_>^{l+1}) j(l_2|k_n r_2)]$$
$$- i\sum_{\substack{n''l_1''\\ l_2''}} \frac{r_1}{k_{n''}}\varphi_{n''l_1''}(r_1)\Big[-\frac{2\delta_{l_1'l_2''}\delta_{l_2'l_1''}}{r_1}\int_0^\infty dr_2 r_2 \varphi_{n'l_1'}(r_2)$$
$$\times \int_0^\infty dr_2' j(l_2''|k_{n''}r_{2<})$$
$$\times h^1(l_2''|k_{n''}r_{2>})\zeta_{\alpha\alpha}(J|\lambda''|\lambda|r_2') + 2\sum_l (-1)^{l_1'+l_2'}$$
$$\times g_l(l_1'l_2', l_1''l_2''; J)\int_0^\infty dr_2 r_2 \varphi_{n'l_1'}(r_2)(r_<^l/r_>^{l+1})$$
$$\times \int_0^\infty dr_2' j(l_2''|k_{n''}r_{2<})h^1(l_2''|k_{n''}r_{2>})\zeta_{\alpha\alpha}(J|\lambda''|\lambda|r_2')\Big]. \tag{26}$$

These types of equations may be solved in several ways. One simple procedure is to recognize that aside from the inhomogeneity occurring in Eq. (25) for $\zeta_{\alpha\alpha}(J|\lambda'|\lambda|r_2)$, both $\zeta_{\beta\alpha}$ and $\zeta_{\alpha\alpha}$ decay exponentially to zero as r_1 and r_2, respectively, become large. Thus one may write

$$\zeta_{\alpha\alpha}(J|\lambda'|\lambda|r_2)$$
$$= j(l_2|k_n r_2)[-2\delta_{\lambda\lambda'}/r_2 + 2\sum_l f_l(l_1'l_2', l_1 l_2; J)$$
$$\times \int_0^\infty dr_1 \varphi_{n'l_1'}(r_1) r_1^2 (r_<^l/r_>^{l+1})\varphi_{nl_1}(r_1)]$$
$$+ \sum_{p=1}^{M\alpha} B_\alpha(\lambda'|p)\psi_p(r_2) \tag{27}$$

and

$$\zeta_{\beta\alpha}(J|\lambda'|\lambda|r_1)$$
$$= r_1\varphi_{nl_1}(r_1)[-2\delta_{l_1'l_2}\delta_{l_2'l_1}\int_0^\infty dr_2 \varphi_{n'l_1'}(r_2) j(l_2|k_n r_2)$$
$$+ 2\sum_l (-1)^{l_1'+l_2'} g_l(l_1'l_2', l_1 l_2; J)\int_0^\infty dr_2 r_2 \varphi_{n'l_1'}(r_2)$$
$$\times (r_<^l/r_>^{l+1}) j(l_2/k_n r_2)\Big] + \sum_{p=1}^{M\beta} B_\beta(\lambda'|p)\psi_p(r_1) \tag{28}$$

where the $\psi_p(r_1)$ belong to a basis set spanning the Hilbert space of square integrable functions. When Eqs. (27)-(28) are substituted into Eqs. (25)-(26), one obtains equations from which it is trivial to derive simultaneous algebraic equations for the numbers $B_\alpha(\lambda|p)$ and $B_\beta(\lambda|p)$. Thus, these equations can form the basis of close coupled calculations of the direct and exchange T matrix elements since they are given in terms of $\zeta_{\alpha\alpha}$ and $\zeta_{\beta\alpha}$ by

$$T_d^J(\lambda'|\lambda) = \int_0^\infty j(l'|k_{n'}r)\zeta_{\alpha\alpha}(J|\lambda'|\lambda|r)dr \tag{29}$$

and

$$T_{ex}^J(\lambda'|\lambda) = \int_0^\infty j(l'|k_{n'}r)\zeta_{\beta\alpha}(J|\lambda'|\lambda|r)dr. \tag{30}$$

An alternative procedure which in some respects is simpler is to uncouple the equations in the channel index by introducing the singlet and triplet amplitude density expansion coefficients $\zeta^{(\pm)}(J|\lambda'|\lambda|r)$. These are defined by

$$\zeta^{(\pm)}(J|\lambda'|\lambda|r) = \zeta_{\alpha\alpha}(J|\lambda'|\lambda|r) \pm \zeta_{\beta\alpha}(J|\lambda'|\lambda|r), \tag{31}$$

where the + yields the singlet and − yields the triplet. Since there is no spin-spin interaction, it follows that equations for $\zeta^{(+)}(J|\lambda'|\lambda|r)$ and $\zeta^{(-)}(J|\lambda'|\lambda|r)$ must uncouple from each other.[1(a)] These equations are given by

$$\zeta^{(\pm)}(J|\lambda'|\lambda|r) = g^{(\pm)}(J|\lambda'|\lambda|r) \mp i\sum_{n''}\sum_{l_1''}\sum_{l_2''}\frac{r}{k_{n''}}\varphi_{n''l''}(r)$$
$$\times\Big[-\frac{2}{r}\delta_{l_1'l_2''}\delta_{l_2'l_1''}\int_0^\infty dr' r' \varphi_{n'l_1'}(r')\int_0^\infty dr'' j(l_2''|k_{n''}r_<')$$
$$\times h^1(l_2''|k_{n''}r_>')\zeta^{(\pm)}(J|\lambda''|\lambda|r'')$$
$$+ 2\sum_l (-1)^{l_1'+l_2'} g_l(l_1'l_2', l_1''l_2''; J)\int_0^\infty dr' r' \varphi_{n'l_1'}(r_<^l/r_>^{l+1})$$
$$\int_0^\infty dr'' j(l_2''|k_{n''}r_<')h^1(l_2''|k_{n''}r_>')\zeta^{(\pm)}(J|\lambda''|\lambda|r'')\Big], \tag{32}$$

where

$$g^{(\pm)}(J|\lambda'|\lambda|r) = j(l_2|k_n r)\Big[-\frac{2}{r}\delta_{\lambda\lambda'} + 2\sum_l f_l(l_1'l_2', l_1 l_2; J)$$
$$\times\int_0^\infty dr' \varphi_{n'l_1'}(r')(r_<^l/r_>^{l+1})\varphi_{nl_1}(r')(r')^2\Big]$$
$$\pm r\varphi_{nl_1}(r)[-2\delta_{l_1'l_2}\delta_{l_2'l_1}\int_0^\infty dr' \varphi_{n'l_1'}(r') j(l_2|k_n r')$$

$$+ 2\sum_l (-1)^{l_1'+l_2'} g_l(l_1'l_2', l_1l_2; J) \times \int_0^\infty dr' r' \varphi_{n'l_1'}(r')(r_<^l/r_>^{l+1}) j(l_2|k_n r')\Big]. \quad (33)$$

These equations are essentially of the form

$$\zeta(\lambda'|\lambda|r) = I(\lambda'|\lambda|r) + \sum_{\lambda'' \atop l} F_1(\lambda''|r)\int_0^\infty dr' F_2^l(\lambda'|\lambda''|r')(r_<^l/r_>^{l+1}) \times \int_0^\infty dr'' g^1(\lambda''|r_<')\, g^2(\lambda''|r_>')\zeta(\lambda''|\lambda|r'') \quad (34)$$

which we shall employ for convenience sake in discussing the second solution procedure. It is convenient to introduce a matrix notation so Eq. (34) becomes

$$\boldsymbol{\zeta}(r) = \mathsf{I}(r) + \sum_l \mathsf{F}_1(r)\cdot\int_0^\infty dr' \mathsf{F}_2^l(r')(r_<^l/r_>^{l+1})\cdot \times \int_0^\infty dr'' \mathsf{g}^1(r_<')\cdot \mathsf{g}^2(r_>')\cdot \boldsymbol{\zeta}(r'') \quad (35)$$

where comparison of Eqs. (34)-(35) makes the definitions of the matrices $\boldsymbol{\zeta}$, I, F_1, F_2^l, g^1 and g^2 are apparent.

We now eliminate the lesser and greater variables to obtain

$$\boldsymbol{\zeta}(r) = \mathsf{I}(r) + \sum_l \Big[\frac{1}{r^{l+1}}\mathsf{F}_1(r)\cdot\int_0^r dr' \mathsf{F}_2^l(r')(r')^l\cdot(\mathsf{g}^2(r')\cdot \times \int_0^{r'} dr'' \mathsf{g}^1(r'')\cdot\boldsymbol{\zeta}(r'') + \mathsf{g}^1(r')\cdot\int_{r'}^\infty dr'' \mathsf{g}^2(r'')\cdot\boldsymbol{\zeta}(r'')) + r^l\mathsf{F}_1(r)\cdot\int_r^\infty dr' \mathsf{F}_2^l(r') \times \frac{1}{(r')^{l+1}}\cdot(\mathsf{g}^2(r')\cdot\int_0^{r'} dr'' \mathsf{g}^1(r'')\cdot\boldsymbol{\zeta}(r'') + \mathsf{g}'(r')\cdot \times\int_{r'}^\infty dr'' \mathsf{g}^2(r'')\cdot\boldsymbol{\zeta}(r'')\Big]. \quad (36)$$

This equation can now be written as

$$\boldsymbol{\zeta}(r) = \mathsf{I}(r) + \sum_l\Big[\frac{1}{r^{l+1}}\mathsf{F}_1(r)\cdot\int_0^r dr' \mathsf{F}_2^l(r')(r')^l\cdot \times \Big\{\mathsf{g}^2(r')\cdot\int_0^{r'} dr'' \mathsf{g}^1(r'')\cdot\boldsymbol{\zeta}(r'') - \mathsf{g}^1(r')\cdot \times \int_0^{r'} dr'' \mathsf{g}^2(r'')\cdot\boldsymbol{\zeta}(r'') + \mathsf{g}^1(r')\cdot\mathsf{C}\Big\} - r^l\mathsf{F}_1(r)\cdot\int_0^r dr' \mathsf{F}_2^l(r')\frac{1}{(r')^{l+1}}\cdot \times \Big\{\mathsf{g}^2(r')\cdot\int_0^{r'} dr'' \mathsf{g}^1(r'')\cdot\boldsymbol{\zeta}(r'') - \mathsf{g}'(r')\cdot \times \int_0^{r'} dr'' \mathsf{g}^2(r'')\cdot\boldsymbol{\zeta}(r'') + \mathsf{g}^1(r')\cdot\mathsf{C} + r^l\mathsf{F}_1(r)\cdot\mathsf{D}^l\Big\}\Big], \quad (37)$$

where

$$\mathsf{C} = \int_0^\infty dr\, \mathsf{g}^2(r)\cdot\boldsymbol{\zeta}(r) \quad (38)$$

and

$$\mathsf{D}^l = \int_0^\infty dr \mathsf{F}_2^l(r)\frac{1}{r^{l+1}}\cdot\Big\{\mathsf{g}^2(r)\cdot\int_0^r dr' \mathsf{g}^1(r')\cdot\boldsymbol{\zeta}(r') + \mathsf{g}^1(r)\cdot\int_r^\infty dr' \mathsf{g}^2(r')\cdot\boldsymbol{\zeta}(r')\Big\}. \quad (39)$$

This equation is now in a convenient form to apply the Sams-Kouri homogeneous integral solution procedure.[3] The function $\boldsymbol{\zeta}(r)$ is thus written as

$$\boldsymbol{\zeta}(n) = \boldsymbol{\zeta}^{(0)}(r) + \boldsymbol{\zeta}^{(1)}(r)\cdot\mathsf{C} + \sum_l \boldsymbol{\zeta}_l^{(2)}(r)\cdot\mathsf{D}^l, \quad (40)$$

where

$$\boldsymbol{\zeta}^{(0)}(r) = \mathsf{I}(r) + \sum_l\Big[\frac{1}{r^{l+1}}\mathsf{F}_1(r)\cdot\int_0^r dr' \mathsf{F}_2^l(r')(r')^l\cdot \times\Big\{\mathsf{g}^2(r')\cdot\int_0^{r'} dr'' \mathsf{g}^1(r'')\cdot\boldsymbol{\zeta}^{(0)}(r'') - \mathsf{g}^1(r')\cdot \times\int_0^{r'} dr'' \mathsf{g}^2(r'')\cdot\boldsymbol{\zeta}^{(0)}(r'')\Big\} - r^l\mathsf{F}_1(r)\cdot\int_0^r dr' \mathsf{F}_2^l(r') \times\frac{1}{(r')^{l+1}}\cdot\Big\{\mathsf{g}^2(r')\cdot\int_0^{r'} dr'' \mathsf{g}^1(r'')\cdot\boldsymbol{\zeta}^{(0)}(r'') - \mathsf{g}^1(r')\cdot\int_0^{r'} dr'' \mathsf{g}^2(r'')\cdot\boldsymbol{\zeta}^{(0)}(r'')\Big\}\Big] \quad (41)$$

and the equations for $\boldsymbol{\zeta}^{(1)}$ and $\boldsymbol{\zeta}^{(2)}$ are the same except that their inhomogenieties are respectively $\mathsf{g}^1(r)$ and $r^l\mathsf{F}_1(r)$. These equations are recognized as Volterra equations of the second kind and a convenient algorithm for their solution has been reported by Sams and Kouri.[3]

In the next section we report the results of numerical calculations for the purpose of illustrating the present approach.

III. NUMERICAL RESULTS FOR s-WAVE ELASTIC SCATTERING

The coupled channel operator formalism has been used in preliminary calculations of phase shifts for the s-wave singlet and triplet elastic scattering of an electron and hydrogen atom. Since we are directly interested only in the singlet and triplet phase shifts, it is convenient to employ the solution procedure based on the homogeneous integral solution method. Therefore, the direct and exchange amplitude densities are added and subtracted as discussed in the preceding section and the resulting equations are

$$\zeta^{(\pm)}(R) = g^{(\pm)}(R) \mp \frac{8i}{k}\int_R^\infty dr(R-r) \times \exp(-r-R)[h(r)\int_0^r dr'\zeta^{(\pm)}(r')j(r') + j(r)\int_r^\infty dr'\zeta^{(\pm)}(r')h(r')], \quad (42)$$

$$g^{(\pm)}(R) = 8[j(R)\int_R^\infty dr(r-r^2/R)\exp(-2r) \pm \int_0^R dr(r-R)\exp(-r-R)j(r)], \quad (43)$$

where $j(R)$ is the s-wave regular Ricatti-Bessel function and $h(R)$ is the s-wave Ricatti-Hankel function of the first kind. We next interchange the order of integration in Eq. (42). Defining

$$C^{(\pm)} = \int_0^\infty dR\exp(ikR)\zeta^{(\pm)}(R), \quad (44)$$

$$T^{(\pm)} = \int_0^\infty dR\,\sin(kR)\zeta^{(\pm)}(R), \quad (45)$$

$$F_C(R) = \int_R^\infty dr\,\sin(kr)\exp(-r)(R-r),$$

$$F_T(R) = \int_0^\infty dr \exp(-r)(R-r)\exp(ikr), \quad (47)$$

$$a_1^{(\pm)} = \int_0^\infty dR'\zeta^{(\pm)}(R')\int_0^{R'} dr \sin[k(r-R')]\exp(-r), \quad (48)$$

$$a_2^{(\pm)} = \int_0^\infty dR'\zeta^{(\pm)}(R')\int_0^{R'} drr \sin[k(r-R')]\exp(-r), \quad (49)$$

$$b_1^{(\pm)} = -\frac{8}{k}[a_1^{(\pm)} + T^{(\pm)}/(1-ik)], \quad (50)$$

$$b_2^{(\pm)} = -\frac{8}{k}[-a_2^{(\pm)} - T^{(\pm)}/(1-ik)^2], \quad (51)$$

and

$$K(R,R') = \int_{R'}^{R} dr(R-r)\exp(-r)\sin[k(r-R')], \quad (52)$$

we have

$$\zeta^{(\pm)}(R) = g^{(\pm)}(R) \pm \frac{8}{k}C^{(\pm)}F_C(R)\exp(-R) \pm b_1^{(\pm)}R\exp(-R) \pm b_2^{(\pm)}\exp(-R) \mp \frac{8}{k}\exp(-R)\int_0^R dR'\zeta^{(\pm)}(R')K(R,R'). \quad (53)$$

This equation is then readily solved by writing

$$\zeta^{(\pm)}(R) = \zeta_0^{(\pm)}(R) \pm C^{(\pm)}\zeta_1^{(\pm)}(R) \pm b_1^{(\pm)}\zeta_2^{(\pm)}(R) \pm b_2^{(\pm)}\zeta_3^{(\pm)}(R) \quad (54)$$

where the $\zeta_i^{(\pm)}(R)$ are easily seen to satisfy

$$\zeta_0^{(\pm)}(R) = g^{(\pm)}(R) \mp \frac{8}{k}\exp(-R)\int_0^R dR'\zeta_0^{(\pm)}(R')K(R,R'), \quad (55)$$

$$\zeta_1^{(\pm)}(R) = \frac{8}{k}F_C(R)\exp(-R) \mp \frac{8}{k}\exp(-R)\int_0^R dR'\zeta_1^{(\pm)}(R')K(R,R'), \quad (56)$$

$$\zeta_2^{(\pm)}(R) = R\exp(-R) \mp \frac{8}{k}\exp(-R)\int_0^R dR'\zeta_2^{(\pm)}(R')K(R,R'), \quad (57)$$

and

$$\zeta_3^{(\pm)}(R) = \exp(-R) \mp \frac{8}{k}\exp(-R)\int_0^R dR'\zeta_3^{(\pm)}(R')K(R,R'). \quad (58)$$

The kernal $K(R,R')$ has the very important property

$$K(R,R) = 0 \quad (59)$$

so that insertion of a quadrature approximation into Eqs. (55)-(58) leads to simple expressions[3] for the functions $\zeta_i^{(\pm)}$, $i = 0,1,2,3$ at the leading quadrature point R_l.

In Tables I and II the present results are summarized and compared with previous ones. Our results differ significantly from the Hartree-Fock results but agree quite nicely with those obtained by Schwartz[2] in the low energy region. It is generally agreed that the numbers derived by Schwartz are exact to all the reported figures. We note that our results are, in general, closer to these exact values than are the Hartree-Fock results (this is true for the singlet S-wave results and, in the low energy region, the triplet S-wave results).

TABLE I. Singlet S-wave phase shifts for elastic scattering of electrons by hydrogen atoms.[7]

Energy (Ryd)	Hartree-Fock	Massey-Moiseiwitsch	Schwartz	Present Results
0.01	2.396	2.484	2.553	2.541
0.04	1.871	2.003	2.067	2.039
0.09	1.508	1.649	1.696	1.746
0.25	1.031	1.250	1.202	1.080
0.36	0.869	—	1.041	0.871

TABLE II. Triplet S-wave phase shifts for elastic scattering of electrons by hydrogen atoms[7]

Energy (Ryd)	Hartree-Fock	Massey-Moiseiwitsch	Schwartz	Present Results
0.01	2.908	2.909	2.939	2.935
0.04	2.679	2.680	2.717	2.737
0.09	2.461	2.447	2.500	2.564
0.25	2.070	2 029	2.105	2.264
0.36	1.901	1.909	1.933	2.170

It seems that although the approximations involved in the Hartree-Fock and present calculations are similar, i.e., using only the "one state exchange approximation", the efficiency of treating the information available is higher in the present method. (At least this is the case in the low energy region where the approximation mentioned is most valid).

As to the comparison between the present results and those obtained by Schwartz it is very encouraging to see the present results come so close to those of Schwartz in the low energy region where they are expected to be more reliable. Of course, the fact that the results agree less well as the energy increases has to do with the nature of the approximations involved in the present treatment.

1. We used only the first term (the S term) in the expansion of $1/r_{12}$ whereas Schwartz employs the whole function.

2. As mentioned already above we did only a "one state exchange approximation" so that the expansion basis was not augmented with any additional set of functions of any kind. On the other hand, Schwartz included in his calculations a large set of Hylleraas type basis functions.

It is believed that, just as in other studies where rearrangement collisions were treated,[8] the addition of virtual states should improve the results significantly. Although the number of equations which must be integrated will then be increased, the studies of Sams and Kouri suggest that the present solution method can be readily employed.[3]

ACKNOWLEDGMENTS

This research was supported by the Robert A. Welch Foundation, Houston, Texas, under Grant E-395 and by the National Science Foundation. The author, D. J. Kouri, gratefully acknowledges the hospitality of the Theoretical Chemistry Institute of the University of Wisconsin for the month of July, 1971, during which time this manuscript was completed. The authors are grateful to the referee for valuable suggestions.

*Permanent address: Department of Nuclear Chemistry, Soreq Research Center, Yavne, Israel.
†R. A. Welch Foundation postdoctoral fellowship, 1970-71.

‡Present address: Department of Chemistry, California Institute of Technology, Pasadena, California.

[1]See, the reference works by (a) M. L. Goldberger and K. M. Watson, *Collision Theory* (John Wiley, New York, 1964); (b) N. F. Mott and H. S. W. Massey, *Theory of Atomic Collisions*, 3rd Ed. (Oxford Press, London, 1965); (c) R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966); (d) T. Y. Wu and T. Ohmura, *Quantum Theory of Scattering* (Prentice-Hall, Englewood Cliffs, New Jersey, 1962); (e) R. D. Levine, *Quantum Mechanics of Molecular Rate Processes* (Oxford Press, London, 1969); (f) G. F. Drukarev, *Theory of Electron-Atom Collisions* (Academic Press, London, 1965).

[2]A few of the many references include P. M. Morse and W. Allis, Phys. Rev. **44**, 269 (1933); V. Ochkur, Vestn. Leningr. Univ. Ser. Fiz. Khim. **4**, 53 (1958); T. L. John, Proc. Phys. Soc. Lond. **76**, 532 (1960); R. Marriott, Proc. Phys. Soc. Lond. **72**, 121 (1958); K. Smith, Phys. Rev. **120**, 845 (1960); K. Smith and P. G. Burke, Phys. Rev. **123**, 174 (1961); P. G. Burke and H. M. Schey, Phys. Rev. **126**, 147 (1962); H. S. W. Massey and B. L. Moiseiwitsch, Proc. Phys. Soc. Lond. A **205**, 483 (1950); C. Schwartz, Phys. Rev. **124**, 1468 (1961); A. Temkin, Phys. Rev. Lett. **4**, 566 (1960); Phys. Rev. Lett. **6**, 354 (1961); Phys. Rev. **126**, 130 (1962); A. Temkin and J. C. Lamkin, Phys. Rev. **121**, 788 (1961). See also the bibliography in reference 1(f).

[3]W. N. Sams and D. J. Kouri, J. Chem. Phys. **51**, 4809 (1969), J. Chem. Phys. **51**, 4815 (1969), and unpublished research.

[4]M. Baer and D. J. Kouri, Phys. Rev. (to be published).

[5]B. R. Johnson and D. Secrest, J. Math. Phys. **7**, 2187 (1966); D. J. Kouri, J. Chem. Phys. **51**, 5204 (1969).

[6]I. C. Percival and M. J. Seaton, Proc. Camb. Philos. Soc. **53**, 654 (1957).

[7]The results in Tables I-II (other than those of the coupled channel operator method) are taken from a table in the reference work by N. F. Mott and H. S. W. Massey (see p. 530, ref. 1(b) above).

[8]H. Hulburt and J. O. Hirschfelder, J. Chem. Phys. **11**, 273 (1944), K. T. Tang, B. Kleinman, and M. Karplus, J. Chem. Phys. **50**, 119 (1969), D. G. Truhlar and A. Kuppermann, J. Chem. Phys. 52, 3841 (1970) and Phys. Rev. A 4, 1924 (1971), M. Baer, J. Chem. Phys. 54, 3070 (1971), M. Baer and D. J. Kouri, J. Chem. Phys. 56, 1758 (1972), N. S. Evers and D. J. Kouri, J. Chem. Phys. 58, 1955 (1973).

The discrete version of the Marchenko equations in the inverse scattering problem*

K. M. Case and S. C. Chiu

The Rockefeller University, New York, New York 10021
(Received 27 April 1973)

For s-wave and one-dimensional scattering, the discrete Marchenko equations with respect to nontrivial comparison systems have been derived. A relation with the Gel'fand–Levitan equation is obtained. The continuum versions are obtained by a limiting process.

I. INTRODUCTION

The inverse scattering problem in its discrete form was introduced recently,[1] and applied to transport theory.[2] The merit of this discrete approach is in the simplicity of mathematical derivations, while at each stage, one may pass to the continuum limit.

In Refs. 1, 2, the problem was solved by the Gel'fand-Levitan approach where the kernel of the equation was given in terms of the spectral function $\rho(\lambda)$, which is indeed the appropriate given data in the transport problem. In scattering theory, on the other hand, the scattering matrix $S(\lambda)$ is given, and $\rho(\lambda)$ is obtained from $S(\lambda)$ by the Wiener-Hopf technique. In this respect, the Marchenko equation seems to be more suited for the scattering problem, since its kernel is expressed directly in terms of $S(\lambda)$.[3] Furthermore, in the one-dimensional case,[4] there does not seem to be an obvious Gel'fand-Levitan equation.

In this paper, we consider the discrete version of the Marchenko equation for the one-dimensional case and s-wave scattering. The formulation is slightly generalized as follows: Given the solutions and the scattering datum, namely the phase shifts, positions, and normalizations of bound states for a known system characterized by $a^0(n)$, and only the scattering datum for an unknown system $a(n)$, the problem is to solve for the ratio $a(n)/a^0(n)$. In the continuum limit, this gives the difference of the potentials, $q(x) - q^0(x)$.

This generalization has been done for the one-dimensional Marchenko equations and for the Gel'fand-Levitan equation in the continuum case.[5] However, to our knowledge, the generalization for the s-wave Marchenko equation has not appeared in the literature. The relation between the Marchenko solution and the Gel'fand-Levitan solution is obtained. Lastly, the continuum limit is obtained.

II. DISCRETE MARCHENKO EQUATION FOR S-WAVE

The discrete radial Schrödinger equation is[6]

$$a(n+1)\psi(\lambda,n+1) + a(n)\psi(\lambda,n-1) = \lambda\psi(\lambda,n), \tag{1}$$

where $a(n)$ and $v(n)$, the "potential," are related by[1]

$$a(n) = \tfrac{1}{2}\exp[-\tfrac{1}{2}(v(n)+v(n-1))], \tag{1a}$$

$$v(n) = \Delta^2 q(n\Delta), \tag{1b}$$

where Δ is the increment in spatial variable so that $n\Delta \to x$. The solutions of Eq. (1) are defined by their boundary conditions. Specifically, we are interested in the following sets:

(i) $\phi(\lambda,n)$ defined by

$$\phi(\lambda,0)=0, \quad \phi(\lambda,1)=1. \tag{2}$$

(ii) $\phi_\pm(\lambda,n)$ defined by

$$\lim_{n\to\infty} e^{\pm in\theta}\phi_\pm(\lambda,n) = 1, \tag{3}$$

where we define the following:

$$\lambda \equiv \cos\theta \equiv \tfrac{1}{2}(z+z^{-1}). \tag{4}$$

We assume that $\lim_{n\to\infty} a(n) \to \frac{1}{2}$ sufficiently fast. [In fact, it is necessary and sufficient to assume that $\Pi_{n=1}^{\infty} 2a(n) < \infty$.] It can be shown that

$$\phi(\lambda,n) = [2a(1)/(z-z^{-1})][\phi_-(\lambda,0)\phi_+(\lambda,n) - \phi_+(\lambda,0)\phi_-(\lambda,n)]. \tag{5}$$

The phase shift $\delta(\lambda)$ is given by

$$S(\lambda) = e^{2i\delta(\lambda)} = \phi_-(\lambda,0)/\phi_+(\lambda,0). \tag{6}$$

Let λ_i, $i=1,\dots,p$, be the positions of the bound states, i.e., where $\phi_+(\lambda_i,0)$ vanish. It can be shown that

$$\phi(\lambda_i,n) = \phi_+(\lambda_i,n)/\phi_+(\lambda_i,1) \tag{7}$$

and

$$\sum_{n=1}^{\infty}\phi(\lambda_i,n)\phi(\lambda_j,n) = \delta_{ij}a(1)\left[\frac{d\phi_+(\lambda,0)}{d\lambda}\Big/\phi_+(\lambda,1)\right]_{\lambda=\lambda_i} \equiv \delta_{ij}/N_i^2 \tag{8}$$

We also have the completeness relation

$$\delta(n,m) = [2a^2(1)\pi]^{-1}\int_{-1}^{+1} d\lambda\,[\sin\theta\,\phi(\lambda,n)\phi(\lambda,m)/|\phi_+(\lambda,0)|^2] + \sum_i N_i^2\phi(\lambda_i,n)\phi(\lambda_i,m) \tag{9}$$

or, equivalently,

$$\delta(n,m) = -(2\pi i)^{-1}\oint dz[(1-1/z^2)\phi(\lambda,n)\phi_+(\lambda,m)/2a(1)\phi_+(\lambda,0) + \sum_i N_i^2\phi(\lambda_i,n)\phi(\lambda_i,m). \tag{9'}$$

The symmetry relation, $\phi_+(\lambda,n) = \phi_-^*(\lambda,n)$, has been used in arriving at Eq. (9′), and z is defined in Eq. (4).

Let us suppose that the system described by the equation

$$a^0(n+1)\psi^0(\lambda,n+1) + a^0(n)\psi^0(\lambda,n-1) = \lambda\psi^0(\lambda,n) \tag{10}$$

is completely known and distinguish all quantities relevant to this system by the superscript$^{(0)}$. Also, let us suppose that for an arbitrary system described by Eq. (1), we are given the phase shifts $\delta(\lambda)$, the position of the bound states λ_i, and their normalizations N_i^2. The problem is to find $a(n)$. It is of practical advantage to use an arbitrary known system, Eq. (10), for comparison, instead of fixing Eq. (10) to the case $a^0(n) = \frac{1}{2}$.

Copyright © 1973 by the American Institute of Physics

The following representations for $\phi(\lambda,n)$ and $\phi_+(\lambda,n)$ can be rigorously justified:

$$\phi_\pm(\lambda,n) = \sum_{m=n}^{\infty} A(n,m)\phi^0_\pm(\lambda,m), \tag{11}$$

and

$$\phi(\lambda,n) = \sum_{m=0}^{n} K(n,m)\phi^0(\lambda,m). \tag{12}$$

In Eqs. (11) and (12), $A(n,m)$ and $K(n,m)$ are independent of λ.

Let us rewrite Eq. (1) for $\phi_+(\lambda,n)$:

$$a(n+1)\phi_+(\lambda,n+1) + a(n)\phi_+(\lambda,n-1) = \lambda\phi_+(\lambda,n). \tag{13}$$

We note from Eq. (11) that

$$\frac{1}{2\pi i}\oint dz\, z^{-(n+1)}\phi_+(\lambda,n) = A(n,n)\oint \frac{dz}{2\pi i} z^{-(n+1)}\phi_+^0(\lambda,n). \tag{14}$$

Multiplying Eq. (13) by $1/2\pi i z^n$ and integrating over the unit circle, we get

$$a(n)A(n-1,n-1)\oint \frac{dz}{2\pi i z^n}\phi^0_+(\lambda,n-1) = A(n,n)\oint \frac{dz}{2\pi i z^{n+1}}\phi^0_+(\lambda,n). \tag{15}$$

From a similar operation on the equation for $\phi^0_+(\lambda,n)$, we have

$$a(n) = a^0(n)A(n,n)/A(n-1,n-1). \tag{16}$$

Next, let us evaluate the integral

$$I(n,l) \equiv (2\pi i)^{-1}\oint (dz/z)[2i\sin\theta\;\phi(\lambda,n)\phi^0_+(\lambda,l)/\phi_+(\lambda,0)]. \tag{17}$$

We first evaluate by residues at the poles. The residue at the bound state λ_i is

$$R_i(n,l) = 2a(1)N_i^2\phi_+(\lambda_i,n)\phi^0_+(\lambda_i,l)/\phi_+^2(\lambda_i,1), \tag{18}$$

N_i^2 being defined by Eq. (8). The residue at $z=0$ is[7]

$$R_0(n,l) = \begin{cases} 0, & l>n, \\ -2a(1)/A(n,n), & l=n. \end{cases} \tag{19}$$

We are not interested in $R_0(n,l)$ for $l<n$.

$I(n,l)$ can also be evaluated by substituting Eqs. (5) and (11) into the right-hand side of Eq. (17), and we obtain

$$\begin{aligned} I(n,l) &= 2a(1)\Big(-\sum_{m=n}^{\infty} A(n,m)\oint \frac{dz}{2\pi i z}\phi^0_-(\lambda,m)\phi^0_+(\lambda,l) \\ &\quad + \sum_{m=n}^{\infty} A(n,m)\oint \frac{dz}{2\pi i z} S(\lambda)\phi^0_+(\lambda,m)\phi^{(0)}_+(\lambda,l)\Big) \\ &= \begin{cases} \sum_i R_i(n,l), & l>n, \\ R_0(n,l) + \sum_i R_i(n,l), & l=n. \end{cases} \end{aligned} \tag{20}$$

From the completeness relation (9′) for $\phi^0(\lambda,n)$, and the expression for $\phi^0(\lambda,n)$ corresponding to Eq. (5), we have

$$\begin{aligned} I^0(m,l) &\equiv \oint (dz/2\pi i z)[2i\sin\theta\;\phi^0(\lambda,m)\phi^0_+(\lambda,l)/\phi^0_+(\lambda,0)] \\ &= 2a^0(1)\{-\delta(m,l) + \sum_j [N_j^{(0)}/\phi^0_+(\lambda_j^{(0)},1)]^2 \\ &\quad \times \phi^0_+(\lambda_j^{(0)},m)\phi^0_+(\lambda_j^{(0)},l)\} \\ &= 2a^0(1)\Big[-\oint (dz/2\pi i z)\phi^0_-(\lambda,m)\phi^0_+(\lambda,l) \\ &\quad + \oint (dz/2\pi i z)S^0(\lambda)\phi^0_+(\lambda,m)\phi^0_+(\lambda,l)\Big]. \end{aligned} \tag{21}$$

Equation (21) gives us an expression for $-\oint(dz/2\pi i z)\phi^0_-(\lambda,m)\phi^0_+(\lambda,l)$ which, when substituted into Eq. (20), gives us the following equations:

$$A(n,l) + \sum_{m=n}^{\infty} A(n,m)\omega(m,l) = 0 \quad \text{for } l>n \tag{22}$$

and

$$A(n,n) + \sum_{m=n}^{\infty} A(n,m)\omega(m,n) = \frac{1}{A(n,n)}, \tag{23}$$

where

$$\begin{aligned} \omega(m,l) &= \oint \frac{dz}{2\pi i z}\{[S^0(\lambda) - S(\lambda)]\phi^0_+(\lambda,m)\phi^0_+(\lambda,l)\} \\ &\quad + \sum_i \left(\frac{N_i}{\phi_+(\lambda_i,1)}\right)^2 \phi^0_+(\lambda_i,m)\phi^0_+(\lambda_i,l) \\ &\quad - \sum_j \left(\frac{N_j^{(0)}}{\phi^0_+(\lambda_j^{(0)},1)}\right)^2 \phi^0_+(\lambda_j^{(0)},m)\phi^0_+(\lambda_j^{(0)},l). \end{aligned} \tag{24}$$

Let us define

$$\alpha(n,l) \equiv A(n,l)/A(n,n). \tag{25}$$

Then Eqs. (22) and (23) become

$$\alpha(n,l) + \omega(n,l) + \sum_{m=n+1}^{\infty} \alpha(n,m)\omega(m,l) = 0, \qquad l>n, \tag{26}$$

$$\frac{1}{A^2(n,n)} = 1 + \omega(n,n) + \sum_{m=n+1}^{\infty} \alpha(n,m)\omega(m,n). \tag{27}$$

Equation (26) defines $\alpha(n,l)$ for $l>n$, which in turn gives us $A(n,n)$ through Eq. (27). $a(n)$ is then obtained from Eq. (16), i.e.,

$$a(n) = a^0(n)A(n,n)/A(n-1,n-1). \tag{16}$$

Equations (26) and (27) are the obvious analog of the Marchenko equation. For the case when the two systems differ only by the normalization of the ith bound state,

$$[N_i/\phi_+(\lambda_i,1)]^2 - [N_i^{(0)}/\phi^0_+(\lambda_i,1)]^2 = \Delta N_i,$$

we obtain the following expression for $A(n,n)$:

$$A^2(n,n) = 1 + [\Delta N_i\,\phi^0_+(\lambda_i,n)\phi^0_+(\lambda_i,n)]/\Big[1 + \Delta N_i\sum_{l=n}^{\infty}\phi^0_+(\lambda_i,l)\phi^0_+(\lambda_i,l)\Big]. \tag{28}$$

This is the discrete analog of the expression obtained by Jost and Kohn[8] for phase equivalent potentials.

Let us conclude this section by deriving a relation between $A(n,m)$ and $\tilde K(n,m)$ [which is the inverse of $K(n,m)$ defined in Eq. (12), i.e.,

$$\phi^0(\lambda,n) = \sum_{m=0}^{n}\tilde K(n,m)\phi(\lambda,m)\Big]. \tag{12′}$$

From the completeness relation (9), we easily obtain

$$\begin{aligned} \tilde K(n,l) &= (1/\pi)\int_{-1}^{+1} d\lambda\,[\sin\theta\;\phi^0(\lambda,n)\phi(\lambda,l)/2a^2(1)|\phi_+(\lambda,0)|^2] \\ &\quad + \sum_i N_i^2\phi^0(\lambda_i,n)\phi(\lambda_i,l). \end{aligned} \tag{29}$$

Substituting Eq. (5) for $\phi(\lambda, l)$ in the integral, we obtain

$$\tilde{K}(n,l) = \sum_{m=l}^{\infty} A(l,m)\{-\oint \frac{dz}{2\pi i}[(1-z^{-2})\phi^0(\lambda,n) \phi_+^0(\lambda,m)/2a(1)\phi_+(\lambda,0)] + \sum_i N_i^2 \phi^0(\lambda_i,n)\phi_+^0(\lambda_i,m)/\phi_+(\lambda_i,1)\} \equiv \sum_{m=l}^{\infty} A(l,m)p(n,m). \quad (30)$$

The contributions of poles at the bound states cancel in $p(n,m)$. Hence, $p(n,m) = 0$ for $m > n$, and

$$\tilde{K}(n,l) = \sum_{m=l}^{n} A(l,m)p(n,m). \quad (30')$$

In particular, when $l = n$, evaluating the residue at $z = 0$ in Eq. (30) gives

$$\tilde{K}(n,n) = [a^0(1)\phi_+^0(z=0,0)/a(1)\phi_+(z=0,0)]A(n,n) = 1/K(n,n). \quad (31)$$

In Eq. (31), we have used the identity

$$\phi_+^0(z=0,0)\prod_{l=1}^{\infty} 2a^0(l) = 1. \quad (32)$$

Thus,

$$a(n)/a^0(n) = A(n,n)/A(n-1,n-1) = K(n-1,n-1)/K(n,n), \quad (33)$$

as is expected.

III. DISCRETE ONE-DIMENSIONAL MARCHENKO EQUATIONS

The basic equation for the one-dimensional case is still Eq. (1).[9] The following three sets of solutions are separately complete with the inclusion of bound states:

(i) $\psi_1(\lambda,n)$ and $\psi_2(\lambda,n)$ defined by

$$\lim_{n\to+\infty}\psi_1(\lambda,n) \to S_{11}(\lambda)z^n, \quad \psi_2(\lambda,n) \to z^{-n} + S_{21}(\lambda)z^n,$$
$$\lim_{n\to-\infty}\psi_1(\lambda,n) \to z^n + S_{12}(\lambda)z^{-n}, \quad \psi_2(\lambda,n) \to S_{22}(\lambda)z^{-n}, \quad (34)$$

$S_{ij}(\lambda)$ is the scattering matrix.

(ii) $f_1(\lambda,n)$ and $f_1^*(\lambda,n)$ defined by

$$\lim_{n\to\infty} f_1(\lambda,n) \to z^n. \quad (35)$$

(iii) $f_2(\lambda,n)$ and $f_2^*(\lambda,n)$ defined by

$$\lim_{n\to-\infty} f_2(\lambda,n) \to z^{-n}. \quad (36)$$

The S matrix is unitary, and "symmetric", i.e.,

$$|S_{11}(\lambda)|^2 + |S_{12}(\lambda)|^2 = |S_{22}(\lambda)|^2 + |S_{21}(\lambda)|^2 = 1,$$
$$S_{11}(\lambda)S_{21}^*(\lambda) + S_{12}(\lambda)S_{22}^*(\lambda) = 0 \quad (37a)$$

ahd

$$S_{11}(\lambda) = S_{22}(\lambda). \quad (37b)$$

Furthermore, $S_{11}(\lambda)$ and $S_{22}(\lambda)$ are analytic in the unit circle $|z| \le 0$, except for simple poles at the bound states λ_i, $i = 1, \ldots, p$. Either $S_{12}(\lambda)$ or $S_{21}(\lambda)$ is sufficient to determine the whole S matrix. On the unit circle, the following relations are satisfied:

$$\psi_1(\lambda,n) = S_{11}(\lambda)f_1(\lambda,n) = f_2^*(\lambda,n) + S_{12}(\lambda)f_2(\lambda,n), \quad (38)$$

$$\psi_2(\lambda,n) = S_{22}(\lambda)f_2(\lambda,n) = f_1^*(\lambda,n) + S_{21}(\lambda)f_1(\lambda,n). \quad (39)$$

At the bound states, $\lambda = \lambda_i$, $i = 1, \ldots, p$, $f_1(\lambda_i,n)$ and $f_2(\lambda_i,n)$ are proportional, i.e.,

$$f_1(\lambda_i,n) = C_i f_2(\lambda_i,n). \quad (40)$$

Furthermore, the normalizations are given by

$$\sum_{n=-\infty} f_1^2(\lambda_i,n) = -C_i z_i\left[\frac{d}{dz}(S_{11}(\lambda))^{-1}\right]_{z=z_i} \equiv \frac{1}{N_1^2(i)} = C_i^2\sum_{n=-\infty}^{\infty} f_2^2(\lambda_i,n) \equiv \left(\frac{C_i}{N_2(i)}\right)^2. \quad (41)$$

All these properties are in direct analogy to the continuum case, for which the readers are referred to the second paper of Ref. 4.

From the Green's function

$$G(\lambda;n,m) = -2\psi_1(\lambda,n)\psi_2(\lambda,m)/S_{11}(\lambda)(z-z^{-1}), \quad n \ge m,$$
$$= -2\psi_1(\lambda,m)\psi_2(\lambda,n)/S_{11}(\lambda)(z-z^{-1}), \quad n \ge m, \quad (42)$$

we may derive the following completeness relations (see Appendix B):

$$\frac{1}{2\pi i}\oint \frac{dz}{z} f_1(\lambda,n)[f_1^*(\lambda,m) + S_{21}(\lambda)f_1(\lambda,m)] + \sum_i N_1^2(i) f_1(\lambda_i,n)f_1(\lambda_i,m) = \delta(n,m), \quad (43)$$

$$\frac{1}{2\pi i}\oint \frac{dz}{z} f_2(\lambda,m)[f_2^*(\lambda,n) + S_{12}(\lambda)f_2(\lambda,n)] + \sum_i N_2^2(i) f_2(\lambda_i,n)f_2(\lambda_i,m) = \delta(n,m), \quad (44)$$

and

$$\frac{1}{2\pi}\int_{-1}^{+1}\frac{d\lambda}{\sin\theta}[\psi_1(\lambda,n)\psi_1^*(\lambda,m) + \psi_2(\lambda,n)\psi_2^*(\lambda,m)] + \sum_i N_i^2 \psi(\lambda_i,n)\psi(\lambda_i,m) = \delta(n,m). \quad (45)$$

The known system for comparison is again denoted by the superscript$^{(0)}$, and it has all the properties mentioned previously.

Again, one may justify the representations

$$f_1(\lambda,n) = \sum_{m=n}^{\infty} A_1(n,m)f_1^0(\lambda,m), \quad (46)$$

$$f_2(\lambda,n) = \sum_{m=-\infty}^{n} A_2(n,m)f_2^0(\lambda,m). \quad (47)$$

Similar to the S-wave case, one obtains

$$a(n)/a^0(n) = A_1(n,n)/A_1(n-1,n-1) = A_2(n-1,n-1)/A_2(n,n) \quad (48)$$

In analogy to Eq. (30), $\tilde{A}_2(n,l)$, defined by

$$f_2^0(\lambda,n) = \sum_{m=-\infty}^{n} \tilde{A}_2(n,m)f_2(\lambda,m), \quad (49)$$

is related to $A_1(n,m)$ by

$$\tilde{A}_2(n,l) = \sum_{m=l}^{n} A_1(l,m)p_1(n,m), \quad (50)$$

where

$$p_1(n,m) = \oint \frac{dz}{2\pi iz} [f_2^0(\lambda,n)S_{11}(\lambda)f_1^0(\lambda,m)] + \sum_i \frac{N_2^2(i)}{C_i} f_2^0(\lambda_i,n)f_1^0(\lambda_i,m). \quad (50')$$

For $l = n$,

$$\tilde{A}_2(n,n) = 1/A_2(n,n) = S_{11}(z=0)A_1(n,n)/S_{11}^0(z=0). \quad (51)$$

Furthermore, on evaluating the integral

$$I_1(n,l) = \oint \frac{dz}{2\pi iz} f_1^0(\lambda,n)S_{22}(\lambda)f_{22}(\lambda)f_2(\lambda,l)$$

by residues and through Eq. (39), in a manner similar to $I(n,l)$ for s wave, we obtain the Marchenko equations for $A_1(n,m)$:

$$\alpha_1(n,l) \equiv A_1(n,l)/A_1(n,n), \quad (52)$$

$$\alpha_1(n,l) + \Omega_1(n,l) + \sum_{m=n+1}^{\infty} \alpha_1(n,m)\Omega_1(m,l) = 0, \quad l > n, \quad (53)$$

$$\frac{1}{[A_1(n,n)]}^{-2} = 1 + \Omega_1(n,n) + \sum_{m=n+1}^{\infty} \alpha_1(n,m)\Omega_1(m,n), \quad (54)$$

where

$$\Omega_1(m,l) \equiv \oint \frac{dz}{2\pi iz} f_1^0(\lambda,m)f_1^0(\lambda,l)(S_{21}(\lambda) - S_{21}^0(\lambda)) + \sum_i N_1^2(i)f_1^0(\lambda_i,m)f_1^0(\lambda_i,l) - \sum_j N_1^{(0)2}(j)f_1^0(\lambda_j^0,m)f_1^0(\lambda_j^0,l). \quad (55)$$

A similar set of equations to Eqs. (52)-(55) can obviously be derived for $f_2(\lambda,n)$.

IV. CONTINUUM LIMIT

Finally, let us consider the continuum limit of the results obtained. In close analogy to Ref. 1, in the continuum limit,

$$\Delta \to 0, \quad \lim_{\Delta\to 0} n\Delta \to x, \quad m\Delta \to x_1, \quad l\Delta \to x_2$$

$$\cos\theta = 1 - (\sqrt{2E}\,\Delta)^2/2 = 1 - (k\Delta)^2/2.$$

The following limits are valid:

$$\lim_{\Delta\to 0} \Delta\sum \to \int dx, \quad \lim_{\Delta\to 0} \oint \frac{dz}{2\pi iz} \to -\Delta\int_{-\infty}^{\infty} \frac{dk}{2\pi},$$

$$\lim_{\Delta\to 0} z^n \to e^{ikx}, \quad \lim_{\Delta\to 0} z_i \to e^{-K_i\Delta},$$

$$\lim_{\Delta\to 0} \frac{N_i^2}{\phi_+^2(\lambda_i,1)} \to \frac{2_i K_{i\Delta}}{\phi_+'(K_i,0)\phi_+(K_i,0)} = \left(\frac{\mathfrak{N}_i}{\phi_+'(K_i,0)}\right)^2 \Delta,$$

$$\phi_+'(K_i,0) = \left.\frac{d\phi_+(K_i,x)}{dx}\right|_{x\to 0}$$

$$\omega(m,l) \to \Delta\left[\int_{-\infty}^{\infty} \frac{dk}{2}[S(k) - S^0(k)]\phi_+^0(k,x_1)\phi_+^0(k,x_2) + \sum_i \left(\frac{\mathfrak{N}_i}{\phi_+'(k_i,0)}\right)^2 \phi_+^0(K_i,x_1)\phi_+^0(K_i,x_2) - \sum_j \left(\frac{\mathfrak{N}_j^0}{\phi_+^{0\prime}(K_j^0,0)}\right)^2 \phi_+^0(K_j^0,x_1)\phi_+^0(K_j^0,x_2)\right] = \Delta\Omega(x_1,x_2), \quad (56)$$

$$\alpha(n,l) \to \Delta A(x,x_2). \quad (57)$$

Then, we have the Marchenko equation from Eq. (26):

$$A(x,x_2) + \Omega(x,x_2) + \int_x^{\infty} dx_1 A(x,x_1)\Omega(x_1,x_2) = 0, \quad x < x_2. \quad (58)$$

For $A(n,n)$, Eq. (27) implies

$$1/A^2(n,n) \to 1/\Delta^2U^2(x,x) = 1 - 2\Delta A(x,x),$$
$$A(n,n) \to \Delta U(x,x) = 1 + \Delta A(x,x). \quad (59)$$

Remembering that $a(n) = \frac{1}{2}\exp\{-\frac{1}{2}\Delta^2[q(n\Delta - \Delta) + q(n\Delta)]\}$, and with the assumption for convergence, we have

$$\ln\frac{a(n)}{a^0(n)} = \ln A(n,n) - \ln A(n-1,n-1)$$

$$\to -\Delta^2(q(x) - q^0(x)) = \Delta[A(x,x) - A(x-\Delta,x-\Delta)]$$

or

$$q(x) - q^0(x) = -\frac{dA(x,x)}{dx} \quad (60)$$

The limit for the wavefunction $\phi_+(\lambda,n)$ in Eq. (11) is

$$\phi_+(\lambda,n) \to \phi_+(k,x),$$

$$\sum_{m=n}^{\infty} A(n,m)\phi_+^0(\lambda,m) \to \phi_+^0(k,x) + \int_x^{\infty} dx_1 A(x,x_1)\phi_+^0(k,x_1);$$

thus

$$\phi_+(k,x) = \phi_+^0(k,x) + \int_x^{\infty} dx_1 A(x,x_1)\phi_+^0(k,x_1).$$

Equations (56)-(61) completes the limiting procedure.

Obviously, the same limiting process holds for the one-dimensional case. We merely state the results.

Corresponding to Eqs. (52), (53), (54), (55), (46), and (48), we have

$$\alpha_1(n,l) \to \Delta A_1(x,x_1) \quad (52')$$

for $x < x_1$,

$$A_1(x,x_1) + \Omega_1(x,x_1) + \int_x^{\infty} dx_2 A_1(x,x_2)\Omega_1(X_2,x_1) = 0, \quad (53')$$

$$A_1(n,n) \to \Delta U_1(x,x) = 1 + \Delta A_1(x,x), \quad (54')$$

$$\Omega_1(x_2,x_1) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} f_1^0(k,x_2)f_1^0(k,x_1)[S_{21}(k) - S_{21}^0(k)] + \sum_i \mathfrak{N}_1^2(i)f_1^0(k_i,x_2)f_1^0(k_i,x_1) - \sum_j \mathfrak{N}_1^{(0)2}(j)f_1^0(k_j^0,x_2)f_1^0(k_j^0,x_1), \quad (55')$$

where

$$\frac{1}{\mathfrak{N}_1^2(i)} = iC_i\left[\frac{d}{dk}[S_{11}(k)]^{-1}\right]_{k=iK_i},$$

$$f_1(k,x) = f_1^0(k,x) + \int_x^{\infty} dx_1 A_1(x,x_1)f_1^0(k,x_1). \quad (46')$$

Potential $q(x)$ is given by

$$q(x) - q^0(x) = -\frac{dA_1(x,x)}{dx}. \quad (48')$$

V. SUMMARY

For S-wave scattering, we have obtained the discrete analog of the Marchenko equation, Eqs. (24)-(27) and (16), with a nontrivial comparison system. The discrete version for the expression for phase equivalent potentials, Eq. (28), is deduced. The relation between the Gel'fand-

Levitan method and the Marchenko method is summarized by Eqs. (30), (30′), and (31).

For the one-dimensional Schrödinger equation, two equivalent discrete Marchenko equations are obtained: Eqs. (52)-(55) and their corresponding equations for $f_2(\lambda,n)$. The relations between the two sets of equations are given by Eqs. (50), (50′), and (51).

Finally, it is shown how the continuum limit can be obtained.

APPENDIX A

As an example of contrast, the inverse problem for the transport theory[2] is discussed here in the Marchenko scheme. Here the given information is the spectral function

$$d\rho(\lambda)=\begin{cases}\dfrac{\sin\theta\, d\lambda}{\pi|\phi_+(\lambda,0)|^2}, & -1\le\lambda\le 1,\\ \sum_i N_i^2\delta(\lambda-\lambda_i)d\lambda, & \text{otherwise,}\end{cases}\tag{A1}$$

or equivalently we are given $|\phi_+(\lambda,0)|$ on the unit circle, and the positions and normalizations of the bound states. Thus, in order to apply Eqs. (26) and (27), we must find $S(\lambda)=\phi_-(\lambda,0)/\phi_+(\lambda,0)\equiv e^{2i\delta(\lambda)}$.

Consider the function defined by

$$\Psi(z)=\begin{cases}-\ln(\phi_+(\lambda,0)\prod_i[z_i/(z_i-z)])=\psi_+(z), & |z|<1,\\ \ln(\phi_-(\lambda,0)\prod_i[z/(z-1/z_i)])=\psi_-(z), & |z|>1.\end{cases}$$

On $|z|=1$,

$$\psi_+(z)-\psi_-(z)=-\ln|\phi_+(\lambda,0)|^2+\sum_i\ln[(z_i-z)(z-1/z_i)/z_iz],$$

and

$$\psi_+(z)+\psi_-(z)=2i\delta(\lambda)+\ln(\prod_i[(z_i-z)z/(z-1/z_i)z_i]).$$

By the Plemelj formula,

$$2i\delta(\lambda)=\left[\ln\left(\prod_i\frac{(z-1/z_i)z_i}{(z_i-z)z}\right)\right]\frac{1}{\pi i}\oint\frac{f(z')dz'}{z'-z},\tag{A2}$$

where

$$f(z')=-\ln|\phi_+(\lambda',0)|^2+\sum_i\ln[(1-z'/z_i)(1-1/z'z_i)].\tag{A3}$$

Therefore,

$$S(\lambda)=\prod_i\frac{(z_i-1/z)}{(z_i-z)}\exp\left[\frac{1}{\pi i}\oint\frac{f(z')dz'}{z'-z}\right].\tag{A4}$$

This extra step necessary to apply the Marchenko equation to the inverse transport problem is analogous to the inconvenience introduced by using Gel'fand-Levitan equation for inverse scattering problem. Thus, while either the Gel'fand-Levitan or Marchenko approach can be applied to both the inverse transport and inverse scattering problems, the Gel'fand-Levitan method is peculiarly well suited to the transport problem while that of Marchenko has the similar advantages for inverse scattering.

APPENDIX B

The completeness relations for the one-dimensional solutions can be found in a similar fashion as for S-wave case.[1] Here, we consider the integral

$$\tilde I(n,l)=-\frac{1}{2\pi i}\int_c d\lambda\sum_{m=-\infty}^{\infty}G(\lambda;n,m)\delta(m,l),\tag{B1}$$

where the Green's function $G(\lambda;n,m)$ is defined by

$$\begin{aligned}G(\lambda;n,m)&=-2\psi_1(\lambda,n)\psi_2(\lambda,m)/S_{11}(\lambda)(z-z^{-1}), & n\ge m,\\ &=-2\psi_1(\lambda,m)\psi_2(\lambda,n)/S_{11}(\lambda)(z-z^{-1}), & n<m,\end{aligned}\tag{B2}$$

and c is the unit circle.

For $l\le n$,

$$\begin{aligned}\tilde I(n,l)&=\frac{1}{2\pi i}\oint\frac{dz}{z}\frac{\psi_1(\lambda,n)\psi_2(\lambda,l)}{S_{11}(\lambda)}\\ &=\frac{1}{2\pi i}\oint\frac{dz}{z}f_1(\lambda,n)[f_1^*(\lambda,l)+S_{21}(\lambda)f_1(\lambda,l)]\\ &=\frac{1}{2\pi i}\oint\frac{dz}{z}[f_2^*(\lambda,n)+S_{12}(\lambda)f_2(\lambda,n)]f_2(\lambda,l)\\ &=\frac{1}{2\pi}\int_{-1}^{+1}\frac{d\lambda}{\sin\theta}[\psi_1(\lambda,n)\psi_1^*(\lambda,l)+\psi_2(\lambda,n)\psi_2^*(\lambda,l)]\end{aligned}\tag{B2}$$

For $l>n$, the role of n and l is interchanged in Eq. (B2) so that we have the complex conjugate. But, $\tilde I(n,l)$ will eventually be real; so Eqs. (B2) actually hold for all n and l.

Next we evaluate Eq. (B1) by residues inside the unit circle. The contributions from the bound states give

$$\begin{aligned}\sum_i\tilde R_i(n,l)&=-\sum_i N_1^2(i)f_1(\lambda_i,n)f_1(\lambda_i,l)\\ &=-\sum_i N_2^2(i)f_2(\lambda_i,n)f_2(\lambda_i,l).\end{aligned}\tag{B3}$$

The contribution from $z=0$ is zero except for $n=l$, and

$$\tilde R_0(n,l)=\delta(n,l).\tag{B4}$$

Therefore, the completeness relations (43)-(45) follow.

*This work was supported in part by the Air Force Office of Scientific Research Grant 722187

[1]K. M. Case and M. Kac, J. Math. Phys. **14**, 594 (1973); K. M. Case, J. Math. Phys. **14**, 916 (1973).

[2]K. M. Case, Phys. Fluids (to be published).

[3]As a contrast, in Appendix A, we indicate how the transport problem can in principle be treated by the Marchenko method.

[4]I. Kay and H. E. Moses, Nuovo Cimento **3**, 276 (1956). L. D. Faddeev, Trudy Mat. Inst. Steklov **73**, 314 (1964) [Am. Math. Soc. Transl. Ser. 2 **65**, 139 (1964)]. The discrete version with the free wavefunction as the comparison system has been worked out in an unpublished work by K. M. Case.

[5]R. G. Newton, Phys. Rev. **100**, 412 (1955); here the generalized Gel'fand–Levitan equation for all l values is analyzed. R. Jost and W. Kohn, K. Dan. Vidensk. Selsk. Mat.-Fys. Medd. 27 (9) (1953); V. E. Zakhavov and L. D. Faddeev, Funkt. Anal. Ego. Pril. **5** (4), 18 (1971).

[6]For the general theory of Eq. (1), we refer the readers to the Appendix in Ref. 1.

[7]Strictly speaking, the residue is $R_0(n,n)=-2a(1)A(\infty,\infty)/A(n,n)$, but $A(\infty,\infty)=1$.

[8]R. Jost and W. Kohn, Phys. Rev. **88**, 383 (1952); and Ref. 5.

[9]We again assume that $lim_{n\to\pm\infty}a(n)\wedge\to\frac{1}{2}$, and $\Pi_{n=-\infty}^{\infty}2a(n)<\infty$.

Asymptotic behavior of spacing distributions for the eigenvalues of random matrices

J. des Cloizeaux and M. L. Mehta

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, BP. N° 2-91-Gif-sur-Yvette, France
(Received 22 May 1973)

It is known that the probability $E_\beta(0, S)$ that an arbitrary interval of length S contains none of the eigenvalues of a random matrix chosen from the orthogonal ($\beta = 1$), unitary ($\beta = 2$) or symplectic ($\beta = 4$) ensemble can be expressed in terms if infinite products $\Pi_{n=0}^{\infty}[1 - \lambda_{2n}(S)]$ and $\Pi_{n=0}^{\infty}[1 = \lambda_{2n+1}(S)]$, where $\lambda_n(S)$ is an eigenvalue of a certain integral equation. Using values of $\lambda_n(S)$, valid for S large, obtained in connection with a recent study of spheroidal functions, we derive asymptotic expressions ($S \gg 1$) for $E_\beta(0, S)$.

I. INTRODUCTION

In the last decade many authors have studied the distribution of eigenvalues of matrices taken from the so called orthogonal, unitary or symplectic ensembles.[1] Let D be the average distance between successive eigenvalues of these matrices and let $E_\beta(0, S)$ be the probability that a randomly chosen interval of length SD does not contain any of the eigenvalues; the parameter β taking the values 1, 2, or 4 respectively as the matrices are taken from the orthogonal, unitary or the symplectic ensemble. Then we have[2]

$$E_1(0, S) = \prod_{n=0}^{\infty} [1 - \lambda_{2n}(C)], \tag{1}$$

$$E_2(0, S) = \prod_{n=0}^{\infty} [1 - \lambda_n(C)], \tag{2}$$

$$E_4(0, \tfrac{1}{2}S) = \frac{1}{2}\left(\prod_{n=0}^{\infty} [1 - \lambda_{2n}(C)] + \prod_{n=0}^{\infty} [1 - \lambda_{2n+1}(C)]\right), \tag{3}$$

where $C = \pi S/2$ and where the λ_n are the eigenvalues of the integral equation

$$\lambda f(x) = \int_{-1}^{1} \frac{\sin[C(x - y)]}{\pi(x - y)} f(y)\, dy \tag{4}$$

arranged in decreasing order of magnitude,

$$1 \geq \lambda_0 \geq \lambda_1 \geq \cdots \geq 0. \tag{5}$$

Though Eqs. (1)–(5), as they stand, are complete, it is very helpful to have some further information on the eigenfunctions $f_n(x)$; namely, that they also satisfy the differential equation

$$\frac{d}{dx} (1 - x^2) \frac{df}{dx} + (\chi - C^2x^2)f(x) = 0. \tag{6}$$

They are called spheroidal functions, they have been studied extensively in the literature,[3] they have been tabulated,[4] and the tables have been used[5] to calculate numerically the $E_\beta(0, S)$ for not too large S.

Thus the $E_\beta(0, S)$ are quite accurately known for values of $S \lesssim 5$. Our knowledge of them for $S \gg 1$, however, is not so certain. Dyson[6] used a thermodynamic argument to derive their asymptotic behaviors as

$$\ln E_1(0, S) \approx -\frac{\pi^2}{16} S^2 - \frac{\pi}{4} S + \tfrac{1}{8} \ln S + O(1), \tag{7}$$

$$\ln E_2(0, S) \approx -\frac{\pi^2}{8} S^2 + O(1), \tag{8}$$

$$\ln E_4(0, S) \approx -\frac{\pi^2}{4} S^2 + \frac{\pi}{2} S + \tfrac{1}{8} \ln S + O(1), \tag{9}$$

A comparison of these equations with Eqs. (1), (2), and (3) shows that at least the terms in $\ln S$ are certainly wrong.

It is also known[7] that $E_2(0, S) \exp[(\pi^2/8) S^2]$ goes to zero as $S \to \infty$, contradicting Eq. (8). Thus one can trust Eqs. (7)–(9) at most up to terms in S.

Lately, some authors tried to obtain asymptotic formulas for $f_n(x)$ and λ_n themselves. Slepian[8] gave such formulas valid in two regions: (i) n finite, $C \gg 1$, and (ii) $n \sim C \gg 1$. However, they are not sufficient to ascertain the asymptotic behavior of the infinite products in Eqs. (1)–(3). A knowledge of λ_n for intermediate values of n is needed as well. Recently des Cloizeaux and Mehta[9] bridged this gap, giving asymptotic formulas for λ_n valid for any n. They read as follows.

First determine b_n by the implicit relation

$$(n + \tfrac{1}{2}) \frac{\pi}{2} = C \int_0^{\min(1, \sqrt{\epsilon})} \left(\frac{\epsilon - y^2}{1 - y^2}\right)^{1/2} dy + \eta(b), \tag{10}$$

where

$$\epsilon = 1 - \frac{2b}{C}, \tag{11}$$

$$\eta(b) = \varphi(b) - \frac{b}{2}\left(\ln \left| \frac{b}{2} \right| - 1\right), \tag{12}$$

and $\varphi(b)$ is the phase of $\Gamma[\frac{1}{2} + i(b/2)]$,

$$\Gamma[\tfrac{1}{2} + i(b/2)] = \left(\frac{\pi}{\operatorname{ch}(\pi b/2)}\right)^{1/2} e^{i\varphi(b)}. \tag{13}$$

Then

$$1 - \lambda_n = (2\pi)^{1/2}(u/2e)^{u/2}\left[\Gamma\left(\frac{u + 1}{2}\right)\right]^{-1} e^{-2C\delta(\epsilon)} [1 + e^{\pi b}]^{-1}, \tag{14}$$

where

$$u = C\epsilon = C - 2b \tag{15}$$

and the function $\delta(\epsilon)$ is defined by

$$\delta(\epsilon) = (1 - \epsilon) \left(\int_0^{\pi/2} \frac{\sin^2\alpha}{(\sin^2\alpha + \epsilon \cos^2\alpha)^{1/2}}\, d\alpha - \frac{\pi}{4}\right),$$

$$\text{if } \epsilon \leq 1 = 0,$$

$$\text{if } \epsilon \geq 1. \tag{16}$$

We shall use Eqs. (10)–(16) to derive the asymptotic results

$$E_1(0, S) \simeq g\, 2^{1/2} \exp\left(-\frac{\pi^2S^2}{16} - \frac{\pi S}{4} - \frac{1}{8} \log S\right), \tag{17}$$

$$E_2(0, S) \simeq g^2\, 2^{1/2} \exp\left(-\frac{\pi^2S^2}{8} - \frac{1}{4} \log S\right), \tag{18}$$

Copyright © 1973 by the American Institute of Physics

$$E_n(0,S) \simeq g\, 2^{-9/8} \exp\left(-\frac{\pi S^2}{4} + \frac{\pi S}{2} - \frac{1}{8}\log S\right), \quad (19)$$

where g is an unknown constant.

II. CALCULATION

Let us derive asymptotic expressions of

$$F(C) = \prod_{n=0}^{\infty} [1-\lambda_{2n}(C)], \quad (20)$$

$$G(C) = \prod_{n=0}^{\infty} [1-\lambda_{2n+1}(C)], \quad (21)$$

and

$$H(C) \equiv F(C)G(C) = \prod_{n=0}^{\infty} [1-\lambda_n(C)] \quad (22)$$

for large values of C.

First, it is convenient to express $F(C)$ and $G(C)$ in terms of $H(C)$. As was shown in Ref. 9, when n is finite, $u_n \simeq 2n+1$, $\epsilon \ll 1$, and Eq. (14) becomes

$$1-\lambda_n(C) \simeq \sqrt{\pi}\, 2^{3n+2} e^{-2C} C^{n+1/2}/n!. \quad (23)$$

Thus for fixed n and $C \to \infty$

$$\frac{[1-\lambda_{2n}(C)]^{1/4}[1-\lambda_{2n+2}(C)]^{3/4}}{[1-\lambda_{2n+1}(C)]^{3/4}[1-\lambda_{2n+3}(C)]^{1/4}} \simeq \left(1-\frac{1}{(2n+2)^2}\right)^{1/4} \quad (24)$$

and this relation may also be considered as valid when n and C are both large, since in this case the ratio on the right side of Eq. (24) is nearly equal to one.

Thus, we may write

$$\frac{F(C)}{G(C)} = \frac{(1-\lambda_0)^{3/4}}{(1-\lambda_1)^{1/4}} \prod_{n=0}^{\infty} \frac{(1-\lambda_{2n})^{1/4}(1-\lambda_{2n+2})^{3/4}}{(1-\lambda_{2n+1})^{3/4}(1-\lambda_{2n+3})^{1/4}}$$
$$= \frac{(1-\lambda_0)^{3/4}}{(1-\lambda_1)^{1/4}} \prod_{n=0}^{\infty}\left(1-\frac{1}{(2n+2)^2}\right)^{1/4}$$
$$\simeq (2\pi)^{1/4} e^{-C}(2/\pi)^{1/4} = \sqrt{2}e^{-C} \quad (25)$$

[since $\sin\theta/\theta \equiv \prod_{n=1}^{\infty}(1-\theta^2/n^2\pi^2)$].

Thus

$$F(C) \simeq 2^{1/4} e^{-C/2}[H(C)]^{1/2},$$
$$G(C) \simeq 2^{-1/4} e^{C/2}[H(C)]^{1/2}. \quad (26)$$

Now, we have to calculate $\ln H(C)$, which according to Eq. (14) is given by

$$\ln H(C) = \sum_{n=0}^{\infty} \ln(1-\lambda_n)$$
$$= \sum_{n=0}^{\infty} \ln\left(\frac{(2\pi)^{1/2}(u_n/2e)^{u_n/2}}{\Gamma[(u_n+1)/2]}\right) - 2C\sum_{n=0}^{\infty}\delta(\epsilon_n)$$
$$- \sum_{n=0}^{\infty} \ln(1+e^{\pi b_n}). \quad (27)$$

In Eq. (10), the condition $\epsilon < 1$ (or $b > 0$) implies that $n < 2C/\pi - \frac{1}{2}$. Thus the preceding equation can also be written

$$\ln H(C) = \mathfrak{D}_1 + \mathfrak{D}_2 + \mathfrak{D}_3, \quad (28)$$

where

$$\mathfrak{D}_1 = \sum_{n=0}^{\infty} \ln\left(\frac{(2\pi)^{1/2}(u_n/2e)^{u_n/2}}{\Gamma[(u_n+1)/2]}\right) \quad (29)$$

$$\mathfrak{D}_2 = -2C \sum_{0<n<2C/\pi-1/2} [\delta(\epsilon_n) + (\pi/4)(1-\epsilon_n)], \quad (30)$$

$$\mathfrak{D}_3 = -\sum_{n=0}^{\infty} \ln[1+e^{-\pi b_n}]. \quad (31)$$

For large values of u, Stirling formula gives

$$\ln\Gamma\left(\frac{u+1}{2}\right) = \ln\left(\frac{u}{2}\right) - \frac{u}{2} + \frac{1}{2}\ln(2\pi) - \frac{1}{12u},$$

$$\mathfrak{D}_1 \simeq \sum_{n=0}^{\infty} \frac{1}{12u_n} + 0(1). \quad (32)$$

For finite values of n, we know[9] that $u \simeq 2n+1$; thus the sum (32) diverges when $n \to \infty$. Obviously a cut-off must be introduced and this cut-off is proportional to C; therefore,

$$\mathfrak{D}_1 \simeq \frac{1}{24}\ln C + O(1) \quad (33)$$

The term $\mathfrak{D}_2$ is the largest one and the main contribution to this term is obtained by replacing the sum by an integral. Thus, we write:

$$\mathfrak{D}_2 = \mathfrak{D}_{21} + \mathfrak{D}_{22}, \quad (34)$$

$$\mathfrak{D}_{21} = -2C\int_0^1 [\delta(\epsilon) + (\pi/4)(1-\epsilon)]\frac{dn}{d\epsilon}\,d\epsilon, \quad (35)$$

$$\mathfrak{D}_{22} = -2C \sum_{0<n<2C/\pi-1/2} [\delta(\epsilon_n) + (\pi/4)(1-\epsilon_n)]$$
$$+ 2C\int_0^1 [\delta(\epsilon) + (\pi/4)(1-\epsilon)]\frac{dn}{d\epsilon}\,d\epsilon. \quad (36)$$

For $0 < \epsilon < 1$, we have

$$\delta(\epsilon) + \frac{\pi}{4}(1-\epsilon) = (1-\epsilon)\int_0^{\pi/2} \frac{\sin^2\alpha}{[\sin^2\alpha + \epsilon\cos^2\alpha]^{1/2}}\,d\alpha \quad (37)$$

and from Eqs. (10) and (11), we obtain

$$\frac{dn}{d\epsilon} = \frac{C}{\pi}\left(\int_0^{\sqrt{\epsilon}} (\epsilon-y^2)^{-1/2}(1-y^2)^{-1/2}dy - \eta'(b)\right)$$
$$= \frac{C}{\pi}\left(\int_0^{\pi/2} \frac{d\alpha}{(1-\epsilon\sin^2\alpha)^{1/2}} - \eta'(b)\right). \quad (38)$$

Introducing the elliptic integrals

$$E(k) = \int_0^{\pi/2} (1-k^2\sin^2\alpha)^{1/2}\,d\alpha, \quad (39)$$

$$K(k) = \int_0^{\pi/2} (1-k^2\sin^2\alpha)^{-1/2}\,d\alpha. \quad (40)$$

We may also write

$$\delta(\epsilon) + \frac{\pi}{4}(1-\epsilon) = E(\sqrt{1-\epsilon}) - \epsilon K(\sqrt{1-\epsilon}), \quad (41)$$

$$\frac{dn}{d\epsilon} = \frac{C}{\pi}[K(\sqrt{\epsilon}) - \eta'(b)]. \quad (42)$$

Thus D_{21} is the sum of two terms [remember that $b = (C/2)(1-\epsilon)$]:

$$D_{21} = D_{211} + D_{212}, \quad (43)$$

$$D_{211} = -\frac{2C^2}{\pi}\int_0^1 d\epsilon[E(\sqrt{1-\epsilon}) - \epsilon K(\sqrt{1-\epsilon})]K(\sqrt{\epsilon}), \quad (44)$$

$$D_{212} = \frac{8}{\pi}\int_0^{C/2} db\,\eta'(b)b\int_0^{\pi/2} \times \sin^2\alpha\left(1-\frac{2b}{C}\cos^2\alpha\right)^{-1/2}d\alpha. \quad (45)$$

Replacing ϵ by $(1-\epsilon)$ in Eq. (44), we have also

$$D_{211} = -\frac{2C^2}{\pi}\int_0^1 d\epsilon[E(\sqrt{\epsilon}) - (1-\epsilon)K(\sqrt{\epsilon})]K(\sqrt{1-\epsilon})\,d\epsilon. \quad (46)$$

Adding Eqs. (44) and (46) and using Legendre's relation,[10]

$$E(\sqrt{1-\epsilon})K(\sqrt{\epsilon}) + E(\sqrt{\epsilon})K(\sqrt{1-\epsilon}) - K(\sqrt{\epsilon})K(\sqrt{1-\epsilon}) = \pi/2. \tag{47}$$

We get

$$D_{211} = -C^2/2. \tag{48}$$

On the other hand, from Eqs. (12), (13) and from Stirling's formula, we deduce that for $|b| \gg 1$, $\eta(b) \simeq 1/12b$. For this reason, it is convenient to write D_{212} as the sum of two terms:

$$D_{212} = \frac{8}{\pi} \int_0^{C/2} db\, \eta'(b) b \int_0^{\pi/2} \times d\alpha\, \sin^2\alpha \left[\left(1 - \frac{2b}{C}\cos^2\alpha\right)^{-1/2} - 1\right] + 2\int_0^{C/2} db\, \eta'(b) b. \tag{49}$$

In the limit $C \to \infty$, the first term is not changed if we replace $\eta'(b)$ by $-1/12b^2$. Then, putting $b = Ct/2$, we get for this term the convergent integral

$$-\frac{2}{3\pi}\int_0^1 \frac{dt}{t}\int_0^{\pi/2} d\alpha\, \sin^2\alpha\, [(1 - t\cos^2\alpha)^{-1/2} - 1]$$

which is a constant.

The second integral diverges logarithmically when $C \to \infty$, since $b\eta'(b) \simeq -1/12b$. Thus we obtain

$$D_{212} = -\tfrac{1}{6}\ln C + O(1). \tag{50}$$

From Eqs. (43), (48), and (50), we get

$$D_{21} = -C^2/2 - \tfrac{1}{6}\ln C + O(1) \tag{51}$$

The main contribution to the correction $\mathfrak{D}_{22}$ comes from small values of n. For n fixed and C large, we have $\epsilon_n = (2n+1)/C$. Thus, we may write

$$\mathfrak{D}_{22} = -2C \sum_{0 \le n < 2C/\pi}\left[\delta\left(\frac{2n+1}{C}\right) + \frac{\pi}{4}\left(1 - \frac{2n+1}{C}\right) - \frac{C}{2}\int_{2n/C}^{2(n+1)/C} [\delta(\epsilon) + (\pi/4)(1-\epsilon)]d\epsilon\right]. \tag{52}$$

By expanding $\delta(\epsilon) + (\pi/4)(1-\epsilon)$ in terms of $[\epsilon - (2n+1)/C]$ and integrating with respect to ϵ, we obtain

$$\mathfrak{D}_{22} \simeq \frac{1}{3C}\sum_{0 \le n < 2C/\pi} \delta''\left(\frac{2n+1}{C}\right). \tag{53}$$

As was shown in Ref. 9, for small values of ϵ

$$\delta(\epsilon) + \frac{\pi}{4}(1-\epsilon) \simeq 1 + \frac{\epsilon}{4}\ln\epsilon - \epsilon(\ln 2 + \tfrac{1}{4}). \tag{54}$$

Therefore,

$$\delta''(\epsilon) \simeq 1/4\epsilon \tag{55}$$

Thus, we find

$$\mathfrak{D}_{22} = \frac{1}{12}\sum_{0 \le n < 2C/\pi} \frac{1}{2n+1}. \tag{56}$$

The cut-off which appears in this formula may not be exactly right, but in any case, it must be of the order of C and therefore we find

$$\mathfrak{D}_{22} \simeq \frac{1}{24}\ln C + O(1). \tag{57}$$

Thus, Eqs. (34), (51), and (57) give

$$\mathfrak{D}_2 = -(C^2/2) - \tfrac{1}{8}\ln C + O(1). \tag{58}$$

Now, let us calculate the last term $\mathfrak{D}_3$. The values of b_n which contribute to the sum (31) must be small with respect to C. In this case, as was shown in Ref. 9, Eq. (10) can be approximately written

$$(n + \tfrac{1}{2})\frac{\pi}{2} = C + \varphi(b) - (b/2)\ln 4C. \tag{59}$$

On the other hand, the sum $\mathfrak{D}_3$ can be written as an integral

$$\mathfrak{D}_3 \simeq -\int_{-\infty}^{+\infty} \ln(1 + e^{-\pi|b|}) \left|\frac{dn}{db}\right| db. \tag{60}$$

From Eq. (59), we get

$$\frac{dn}{db} = \frac{2}{\pi}\varphi'(b) - \frac{1}{\pi}\ln 4C. \tag{61}$$

Keeping the main term, we have

$$\mathfrak{D}_3 \simeq -\frac{1}{\pi}\ln C\int_{-\infty}^{+\infty}\ln(1 + e^{-\pi|b|})db + O(1) = \frac{2}{\pi^2}\ln C\int_0^1 \ln(1+t)\frac{dt}{t} = -\tfrac{1}{6}\ln C + O(1). \tag{62}$$

The value of $\ln H(C)$ is found by using Eq. (28) and collecting the results of Eqs. (33), (58), and (62):

$$\ln H(C) \simeq -(C^2/2) - \tfrac{1}{4}\ln C + O(1). \tag{63}$$

With the help of Eqs. (26), we obtain the final result:

$$F(C) \simeq h\, 2^{1/4}\exp\left(-\frac{C^2}{4} - \frac{C}{2} - \frac{1}{8}\log C\right), \tag{64}$$

$$G(C) \simeq h\, 2^{-1/4}\exp\left(-\frac{C^2}{4} + \frac{C}{2} - \frac{1}{8}\log C\right), \tag{65}$$

$$H(C) \simeq h^2\exp\left(-\frac{C^2}{2} - \frac{1}{4}\log C\right), \tag{66}$$

where h is an unknown constant.

Incidentally, we note that the functions $A(C) \equiv F'(C)/F(C)$ and $B(C) \equiv G'(C)/G(C)$ are related by an exact relation[11]

$$\frac{d}{dC}[A(C) + B(C)] = -[A(C) - B(C)]^2 \tag{67}$$

and that the expressions (64), (65), (66) are compatible with this relation, as can be easily verified.

According to Equations (1), (2), (3) and (20), (21), (22)

$$E_1(0,S) = F(\pi S/2),$$
$$E_2(0,S) = H(\pi S/2),$$
$$E_4(0,S) = \tfrac{1}{2}[F(\pi S) + G(\pi S)]. \tag{68}$$

Thus, using Equations (64), (65), (66), we deduce immediately Equations (17), (18), (19) [where $g = h(2\pi)^{-1/8}$], i.e., the announced asymptotic behavior of $E_\beta(0,S)$ for $\beta = 1, 2, 4$.

[1] See, for example, M. L. Mehta, *Random Matrices* (Academic, New York, 1967).
[2] M. L. Mehta and F. J. Dyson, J. Math. Phys. 4, 715 (1963), Eqs. (16), (17), and (18).
[3] J. Meixner and F. Schafke, *Mathieusche Funktionen und Spharoidfunktionen* (Springer-Verlag, Berlin, 1954); C. Flammer, *Spheroidal wave functions* (Stanford U.P., Stanford, 1957); L. Robin, *Fonctions sphèriques de Legendre et fonction sphèroidales* (Gauthier-Villars, Paris, 1959).
[4] J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and P. J. Corbato, *Spheroidal Wave Functions* (Wiley, New York, 1956).
[5] M. Gaudin, Nucl. Phys. 25, 447 (1961); P. B. Kahn, Nucl. Phys. 41, 159 (1963).
[6] F. J. Dyson, J. Math. Phys. 3, 157 (1962).
[7] M. L. Mehta, Note CEA-N-1458(E), 1971, Centre d'Etudes Nucléaires de Saclay, (France).
[8] D. Slepian, J. Math. and Phys. 44, 99 (1965).
[9] J. des Cloizeaux and M. L. Mehta, J. Math. Phys. 13, 1745 (1972).
[10] See, for example, *Handbook of Mathematical Functions,* edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[11] See Ref. 1, Eq. (A.31.23).

Quark structure and octonions*

Murat Günaydin and Feza Gürsey†

Physics Department, Yale University, New Haven, Connecticut 06520
(Received 14 May 1973)

The octonion (Cayley) algebra is studied in a split basis by means of a formalism that brings out its quark structure. The groups $SO(8)$, $SO(7)$, and G_2 are represented by octonions as well as by 8×8 matrices and the principle of triality is studied in this formalism. Reduction is made through the physically important subgroups $SU(3)$ and $SU(2) \otimes SU(2)$ of G_2, the automorphism group of octonions.

1. INTRODUCTION

Octonions made their appearance in physics as a byproduct of an early attempt to generalize quantum mechanics through the associativity condition for physical observables.[1,2] In their algebraic approach to quantum mechanics, Jordan, von Neumann, and Wigner focused on the properties of Hermitian density matrices. Such matrices close under the commutative "Jordan" product which can be defined as their anticommutator. Thus, in switching from the matrix algebra of density matrices, we trade associativity for commutativity. The two formulations are equivalent except in the case of octonion hermitian 3×3 density matrices which form an exceptional Jordan Algebra.[2] In the latter case the nonassociativity is intrinsic and cannot be removed by going over to a corresponding operator algebra in a finite Hilbert space. In fact, it originates in the structure of the underlying octonion algebra which is a not commutative, not associative, but, alternative division algebra.

The Jordan approach has proved to be more fruitful in mathematics than physics. It has since been quietly dropped in favor of the associative Dirac algebra of operators in Hilbert space,[3] which generalizes the algebra of finite matrices.

The story took a new turn when the charge space made its appearance two decades ago in the Gell-Mann-Nishijima[4] scheme based on isospin and strangeness. A decade later, this led to the quark structure of elementary particles, revealing the underlying $SU(3)$ symmetry.[5] Meantime another group of rank two, namely G_2 was tried[6] and abandoned. Now, G_2 is the automorphism group of the octonion algebra and it admits $SU(3)$ as a subgroup. In fact, $SU(3)$ is the automorphism group of the multiplication rules among six of the octonion units. In terms of this subgroup the generators of G_2 split into an $SU(3)$ octet, a triplet and an antitriplet. Furthermore G_2 has a $SU(2) \times SU(2)$ subgroup under which the generators decompose as $(1, 0)$, $(0, 1)$, and $(1/2, 3/2)$. One of the $SU(2)$ is the isospin, while the other is a generalization of hypercharge to a rotation group. Hence the quark structure is manifest in G_2 and also in the other exceptional groups which are all related to octonions[7] and admit G_2 as a subgroup. An example is the exceptional Jordan algebra which has the exceptional group F_4 as its automorphism group.[8] The quark structure of this algebra was pointed out by Gamba.[9] Other possible connections of the Cayley algebra with internal symmetries were discussed by Pais[10] and others[11] while the admissibility of the elements of the exceptional Jordan algebra as observables was considered by Sherman[12] following the general algebraic framework of Segal.[13] Finally, we have shown recently[14] that the Poincaré group possesses an octonionic representation that leads to a quark structure arising from the breaking of G_2 with $SU(3)$ as the surviving subgroup.

An independent, and perhaps related line of research concerns the construction of Weinberg type renormalizable models[15] based on groups that do not give rise to triangular anomalies, including G_2, $SO(7)$, and $SO(8)$. The attempts enumerated above seem to provide sufficient motivation for a reformulation of the octonion algebra, and the groups $SO(8)$, $SO(7)$, and G_2 connected with it, in a manner which manifestly exhibits its quark structure and its $SU(3)$ content in charge space. Although a vast mathematical literature exists on these subjects,[7,16] it is not presented in a form directly usable by the particle physicist. The object of the present paper is to recast the mathematical theory in a quark language, in direct correspondence with Gell-Mann's treatment of $SU(3)$,[5] to develop an 8×8 matrix formalism, initiated by Seligman,[17] which allows us to treat $SO(8)$, $SO(7)$, G_2 in a unified way and prepare the ground for the investigation of the properties of an octonionic Hilbert space.[18]

The features which seem to be new consist of a matrix form for the octonion multiplication, the representation of G_2 through purely octonionic multiplication in a split basis and the reduction of $SO(8)$, $SO(7)$, and G_2 with respect to their physically important subgroups $SU(3)$ and $SU(2) \times SU(2)$. It is this reduction which exhibits the quark structure of the algebra.

The contents of the paper are as follows. The octonion algebra in the split basis is introduced in Sec 2 and its automorphism group G_2 derived in Sec. 3. The Lie algebra of G_2 and its imbedding in $SO(7)$ are given in Sec. 4. The following section 5 covers the $SU(3)$ and $SU(2) \times SU(2)$ subgroups of G_2. Section 6 is devoted to split octonions and split G_2 while the quark structure in split basis emerges in Sec. 7. A purely octonion representation of split G_2 appears in Sec. 8. The 8×8 matrix formulation of the Cayley algebra forms the object of Sec. 9. The imbedding of G_2 in $SO(7)$ and $SO(8)$ and its reduction with respect to its $SU(2) \times SU(2)$ subgroups are discussed respectively in Secs. 10 and 11. Finally the principle of triality is discussed within the formalism developed previously in Sec. 12. Additional details such as the structure constants of G_2, Zorn's vector-matrix method, theorems pertaining to triality and the realization of the Cayley algebra by means of Gell-Mann's 3×3 λ-matrices, and Dirac's 4×4 γ-matrices appear in the appendixes.

2. THE OCTONION ALGEBRA AND ITS SPLIT BASIS

A composition algebra is defined as an algebra A with identity and with a nondegenerate quadratic form Q defined over it such that Q permits composition, i.e. for $x, y \in A$.

$$Q(xy) = Q(x)Q(y). \tag{2.1}$$

According to the celebrated Hurwitz theorem, there

Copyright © 1973 by the American Institute of Physics

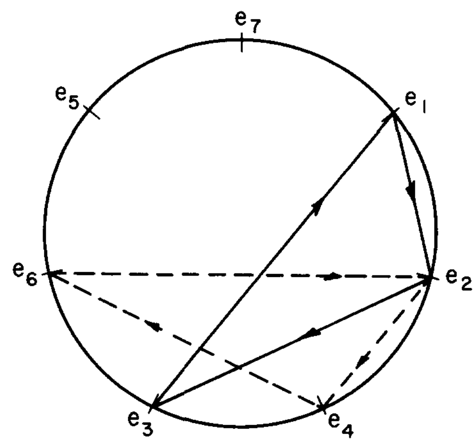

FIG. 1.

exist only four different composition algebras over the real or complex number fields. These are the real numbers **R** of dimension 1, complex numbers **C** of dimension 2, quaternions **H** of dimension 4, and octonions **O** of dimension 8. Of these algebras, the quaternions **H** are not commutative and the octonions **O** are neither commutative nor associative.[19] A composition algebra is said to be a division algebra if the quadratic form Q is anisotropic i.e.,

if $Q(x) = 0$ implies that $x = 0$.

Otherwise the algebra is called *split*.

Assuming that the reader is familiar with the algebras **R, C,** and **H,** we shall review briefly the properties of octonion algebra **O**[20] (sometimes called the Cayley algebra).

A basis for the real octonion **O** will contain eight elements including the identity

$$1, e_A, \quad A = 1, \ldots, 7, \quad \text{where } e_A^2 = -1.$$

For later application to the $SU(3)$ symmetry in particle physics, we label the elements e_A such that they satisfy the following multiplication table:

$$e_1e_2 = e_3, \quad e_5e_1 = e_6, \quad e_6e_2 = e_4, \quad e_4e_3 = e_5,$$
$$e_4e_7 = e_1, \quad e_6e_7 = e_3, \quad e_5e_7 = e_2$$

and

$$e_Ae_B + e_Be_A = -2\delta_{AB}$$

more concisely,

$$e_Ae_B = a_{ABC}e_C - \delta_{AB} \tag{2.2}$$

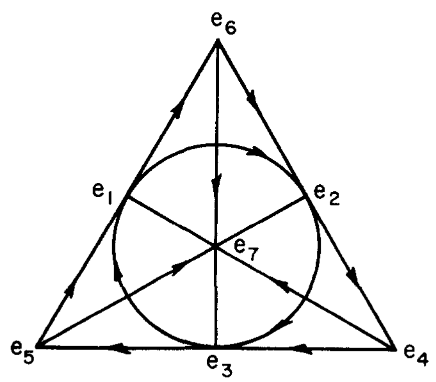

FIG. 2.

where a_{ABC} is totally antisymmetric and

$$a_{ABC} = +1 \text{ for } ABC = 123, 516, 624, 435, 471, 673, 572.$$

Note here the cyclic symmetry obtained by ordering seven points clockwise on a circle with the numbering (1243657) as given in Fig. 1. Then a triangle ABC is obtained from (123) by 6 successive rotations of angle $2\pi/7$. In Fig. 1 the elements corresponding to the corners of the triangle form a basis of a quaternion subalgebra (together with the identity element). Another convenient way of representing the multiplication table by singling out one of the elements is provided by the triangular diagram given in Fig. 2, where arrows show the directions along which the multiplication has a positive sign, e.g.

$$e_5e_1 = e_6, \quad e_1e_5 = -e_6, \quad e_6e_1 = -e_5, \quad e_6e_5 = e_1.$$

From the above multiplication table it is clear that the algebra **O** is not associative. Yet it satisfies a weaker condition than associativity, namely alternativity, i.e., the associator $[x,y,z]$ of the elements x,y,z defined as

$$[x,y,z] = (xy)z - x(yz) \tag{2.3}$$

is an alternating function of x,y,z:

$$[x,y,z] = [z,x,y] = [y,z,x] = -[y,x,z) = -[x,z,y].$$

This property if trivially satisfied by associative algebras **R, C,** and **H.**

The octonion algebra **O** with the above basis considered over the real numbers **R** is a division algebra with the quadratic form Q defined by

$$Q(x) = \bar{x}x = x\bar{x},$$

where $\bar{x}$ is the octonion conjugate of x obtained by replacing e_A in x by $-e_A$.

$$x = x_0 + x_Ae_A, \quad \bar{x} = x_0 - x_Ae_A.$$

This quadratic form is also called the norm form and denoted by $N(x)$. Then

$$N(x) = x\bar{x} = \bar{x}x = x_0^2 + \sum_{A=1}^{7} x_A^2. \tag{2.4}$$

For the split octonion algebra we choose the following basis:

$$u_1 = \tfrac{1}{2}(e_1 + ie_4), \quad u_2 = \tfrac{1}{2}(e_2 + ie_5),$$
$$u_3 = \tfrac{1}{2}(e_3 + ie_6), \quad u_0 = \tfrac{1}{2}(1 + ie_7)$$
$$u_1^* = \tfrac{1}{2}(e_1 - ie_4), \quad u_2^* = \tfrac{1}{2}(e_2 - ie_5),$$
$$u_3^* = \tfrac{1}{2}(e_3 - ie_6), \quad u_0^* = \tfrac{1}{2}(1 - ie_7),$$

where $i = \sqrt{-1}$ and is assumed to commute with all e_A. These basis elements satisfy the multiplication table:

$$u_iu_j = \epsilon_{ijk}u_k^*, \quad i,j,k = 1,2,3,$$
$$u_i^*u_j^* = \epsilon_{ijk}u_k,$$
$$u_iu_j^* = -\delta_{ij}u_0, \quad u_i^*u_j = -\delta_{ij}u_0^*,$$
$$u_iu_0 = 0, \quad u_iu_0^* = u_i, \quad u_i^*u_0 = u_i^*, \quad u_i^*u_0^* = 0,$$
$$u_0u_i = u_i, \quad u_0^*u_i = 0, \quad u_0u_i^* = 0, \quad u_0^*u_i^* = u_i^*,$$
$$u_0^2 = u_0, \quad u_0^{*2} = u_0^*, \quad u_0u_0^* = u_0^*u_0 = 0.$$

Clearly, the split octonion algebra contains divisors of zero and hence is not a division algebra. In Appendix B, we give a realization of split octonion algebra in terms of Zorn's vector matrices.

3. G_2 AS THE AUTOMORPHISM GROUP OF OCTONIONS

An automorphism of an algebra A is defined as an isomorphism of A onto itself. Under the automorphism, multiplication table of A is left invariant, i.e.,

$$x, y \in A, \qquad T \in \mathrm{Aut}A,$$

then

$$T(xy) = T(x)T(y)$$

and the automorphisms map the identity 1 into itself.

The set of all automorphisms of composition algebras form a group. For the real numbers **R** and complex numbers **C**, the groups of automorphisms are the trivial identity mapping and the cyclic group C_2, respectively. The automorphism group of quaternions is the $SU(2)$ group.[21] Below we shall investigate the automorphism group of the octonions, which is the exceptional Lie group G_2.

We use the following results of M. Zorn[22] as our starting point: Each automorphism of the Cayley algebra **O** is completely defined by the images of 3 "independent" basis elements.[23] Consider one such set, say $\{e_1, e_2, e_4\}$. Then there exists an automorphism σ such that

$$\begin{aligned}&\sigma(e_1) = e_1,\\ &\sigma(e_2) = \cos\phi_1 e_2 + \sin\phi_1 e_3,\\ &\sigma(e_4) = \cos\phi_2 e_4 + \sin\phi_2 e_7,\end{aligned} \tag{3.1}$$

The images of the other basis elements are determined by the conditions:

$$\begin{aligned}&\sigma(e_2)\sigma(e_4) = \sigma(e_6), \qquad \sigma(e_1)\sigma(e_2) = \sigma(e_3),\\ &\sigma(e_1)\sigma(e_4) = \sigma(e_7), \qquad \sigma(e_4)\sigma(e_3) = \sigma(e_5).\end{aligned}$$

It can easily be checked that $\sigma(e_A)$ satisfy the same multiplication table as e_A and hence σ is an automorphism. Conversely, one has the very important result that *each automorphism of* **O** *belongs in this manner to at least one Cayley basis*.

Now let us write down all the images of all basis elements under σ explicitly and observe some general patterns:

$$\begin{aligned}&\sigma(e_1) = e_1,\\ &\begin{pmatrix}\sigma(e_2)\\ \sigma(e_3)\end{pmatrix} = \begin{pmatrix}\cos\phi_1 & \sin\phi_1\\ -\sin\phi_1 & \cos\phi_1\end{pmatrix}\begin{pmatrix}e_2\\ e_3\end{pmatrix},\\ &\begin{pmatrix}\sigma(e_4)\\ \sigma(e_7)\end{pmatrix} = \begin{pmatrix}\cos\phi_2 & \sin\phi_2\\ -\sin\phi_2 & \cos\phi_2\end{pmatrix}\begin{pmatrix}e_4\\ e_7\end{pmatrix},\\ &\begin{pmatrix}\sigma(e_6)\\ \sigma(e_5)\end{pmatrix} = \begin{pmatrix}\cos(\phi_1+\phi_2) & -\sin(\phi_1+\phi_2)\\ \sin(\phi_1+\phi_2) & \cos(\phi_1+\phi_2)\end{pmatrix}\begin{pmatrix}e_6\\ e_5\end{pmatrix}.\end{aligned} \tag{3.2}$$

We see that under the automorphism σ we have three invariant planes (e_2, e_3), (e_4, e_7), (e_6, e_5) that undergo rotations through angles ϕ_1, ϕ_2, ϕ_3, respectively, such that

$$\phi_1 + \phi_2 + \phi_3 = 0 \bmod 2\pi.$$

We shall call the automorphisms of the form above *canonical automorphisms*. Each canonical automorphism has a fixed point and 3 invariant planes. If we denote the fixed point by e_k then the invariant planes (e_i, e_j) are determined by the conditions $e_i e_j = e_k$.

For each Cayley basis, there are seven independent canonical automorphisms. The canonical automorphism given above can be written more concisely as:

$$\begin{pmatrix}\sigma(e_2) + i\sigma(e_3)\\ \sigma(e_4) + i\sigma(e_7)\\ \sigma(e_6) + i\sigma(e_5)\end{pmatrix} = e^{(\alpha^1\lambda_3 + \beta^1\lambda_8)e_1}\begin{pmatrix}e_2 + ie_3\\ e_4 + ie_7\\ e_6 + ie_5\end{pmatrix}, \tag{3.3}$$

where λ_3 and λ_8 are the Gell-Mann matrices

$$\lambda_3 = \begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 0\end{pmatrix}, \qquad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix}$$

and α^1 and β^1 are related to ϕ_1 and ϕ_2 as:

$$\phi_1 = \alpha^1 + 3^{-1/2}\beta^1, \qquad \phi_2 = -\alpha^1 + 3^{-1/2}\beta^1.$$

This can be generalized to all the canonical automorphisms. First define seven octonionic 3-spinors $\psi(e_A)$:

$$\begin{aligned}
\psi(e_1) &= \frac{1}{2}\begin{pmatrix}e_6 + ie_5\\ e_2 + ie_3\\ e_4 + ie_7\end{pmatrix} = \frac{(1 + ie_1)}{2}\begin{pmatrix}e_6\\ e_2\\ e_4\end{pmatrix},\\
\psi(e_2) &= \frac{1}{2}\begin{pmatrix}e_4 + ie_6\\ e_3 + ie_1\\ e_5 + ie_7\end{pmatrix} = \frac{(1 + ie_2)}{2}\begin{pmatrix}e_4\\ e_3\\ e_5\end{pmatrix},\\
\psi(e_3) &= \frac{1}{2}\begin{pmatrix}e_5 + ie_4\\ e_1 + ie_2\\ e_6 + ie_7\end{pmatrix} = \frac{(1 + ie_3)}{2}\begin{pmatrix}e_5\\ e_1\\ e_6\end{pmatrix},\\
\psi(e_4) &= \frac{1}{2}\begin{pmatrix}e_3 + ie_5\\ e_6 + ie_2\\ e_7 + ie_1\end{pmatrix} = \frac{(1 + ie_4)}{2}\begin{pmatrix}e_3\\ e_6\\ e_7\end{pmatrix},\\
\psi(e_5) &= \frac{1}{2}\begin{pmatrix}e_1 + ie_6\\ e_4 + ie_3\\ e_7 + ie_2\end{pmatrix} = \frac{(1 + ie_5)}{2}\begin{pmatrix}e_1\\ e_4\\ e_7\end{pmatrix},\\
\psi(e_6) &= \frac{1}{2}\begin{pmatrix}e_2 + ie_4\\ e_5 + ie_1\\ e_7 + ie_3\end{pmatrix} = \frac{(1 + ie_6)}{2}\begin{pmatrix}e_2\\ e_5\\ e_7\end{pmatrix},\\
\psi(e_7) &= \frac{1}{2}\begin{pmatrix}e_1 + ie_4\\ e_2 + ie_5\\ e_3 + ie_6\end{pmatrix} = \frac{(1 + ie_7)}{2}\begin{pmatrix}e_1\\ e_2\\ e_3\end{pmatrix}.
\end{aligned} \tag{3.4}$$

A canonical automorphism leaving the element e_A fixed is defined by its action on $\psi(e_A)$:

$$\begin{aligned}\sigma^A : \psi(e_A) \to \sigma^A\psi(e_A) = \psi'(e_A) &= e^{(\alpha^A\lambda_3 + \beta^A\lambda_8)e_A}\psi(e_A)\\ &= e^{-i(\alpha^A\lambda_3 + \beta^A\lambda_8)}\psi(e_A),\end{aligned} \tag{3.5}$$

no sum over A.

The rows $(e_B + ie_C)$ of $\psi(e_A)$ are determined by the invariant planes (e_B, e_C) of the automorphism σ^A and the ordering of the rows is such that the first elements along the column define imaginary units of a quaternion subalgebra in a positive sense. Hence the cyclic permutation of the rows of $\psi(e_A)$ is immaterial for the subsequent discussion. Canonical automorphisms involve two independent parameters each and hence generate a 14-parameter Lie group. That this is the complete automorphism group of octonions follows from the well-known result that the automorphism group of octonions is a 14-parameter Lie group of type G_2.

Consequently, every automorphism of Cayley numbers can be written as a product of canonical automorphisms and; as stated above, every automorphism can be reduced to the canonical form in a suitably chosen basis.

4. THE LIE ALGEBRA OF G_2 AND ITS IMBEDDING IN $SO(7)$

Using the result that canonical automorphisms generate the Lie group G_2, let us now find its Lie algebra. As parameters corresponding to the generators of G_2, we shall take α^A and β^A rather than the angles ϕ_1^A and ϕ_2^A. Now consider a canonical automorphism σ_A with $\beta^A = 0$, then

$$\sigma_A\psi(e_A) = \psi'(e_A) = e^{\alpha^A\lambda_3 e_A}\begin{pmatrix} e_B + ie_C \\ e_D + ie_E \\ e_F + ie_G\end{pmatrix} = \begin{pmatrix} e'_B + ie'_C \\ e'_D + ie'_E \\ e'_F + ie'_G\end{pmatrix}$$

which gives

$$\begin{pmatrix} e'_B \\ e'_C\end{pmatrix} = \begin{pmatrix} \cos\alpha^A & \sin\alpha^A \\ -\sin\alpha^A & \cos\alpha^A\end{pmatrix}\begin{pmatrix} e_B \\ e_C\end{pmatrix},$$

$$\begin{pmatrix} e'_D \\ e'_E\end{pmatrix} \quad \begin{pmatrix} \cos\alpha^A & -\sin\alpha^A \\ \sin\alpha^A & \cos\alpha^A\end{pmatrix}\begin{pmatrix} e_D \\ e_E\end{pmatrix}.$$

Therefore the group action with parameter α^A induces rotations in the invariant planes (e_B, e_C) and (e_D, e_E) through angles α^A and $-\alpha^A$, respectively. Hence the generator of this group action is

$$(J_{BC} - J_{DE}),$$

where J_{BC} and J_{DE} are the anti-Hermitian rotation generators.

$$J_{BC} = -J_{CB} = -J_{BC}^{\dagger}.$$

Similarly, the generator corresponding to the group action with parameter β^A is

$$(1/\sqrt{3})(J_{BC} + J_{DE} - 2J_{FG}).$$

Since the indices go from 1 to 7, the 14 generators thus constructed will form a subalgebra of $SO(7)$ if they close under commutation. As will be shown below, they indeed close under commutation and hence establish the known result that G_2 is a subgroup of $SO(7)$. The remaining generators of $SO(7)$ can be taken as

$$(J_{BC} + J_{DE} + J_{FG})$$

which are generated by the mappings:

$$\psi(e_A) \to e^{\gamma^A I_3 e_A}\psi(e_A) = e^{-i\gamma^A I_3}\psi(e_A),$$

where I_3 is the 3×3 identity matrix.

For reasons that will be clear later, we shall modify the above basis for G_2 and $SO(7)$ and consider the following Hermitian basis:

$$\begin{aligned}
F_1 &= -i(J_{24} - J_{51}), & M_1 &= (i/\sqrt{3})(J_{24} + J_{51} - 2J_{73}), & N_1 &= i(J_{24} + J_{51} + J_{73}),\\
F_2 &= i(J_{54} - J_{12}), & M_2 &= (-i/\sqrt{3})(J_{54} + J_{12} - 2J_{67}), & N_2 &= i(J_{54} + J_{12} + J_{67}),\\
F_3 &= -i(J_{14} - J_{25}), & M_3 &= (i/\sqrt{3})(J_{14} + J_{25} - 2J_{36}), & N_3 &= i(J_{14} + J_{25} + J_{36}),\\
F_4 &= -i(J_{16} - J_{43}), & M_4 &= (i/\sqrt{3})(J_{16} + J_{43} - 2J_{72}), & N_4 &= i(J_{16} + J_{43} + J_{72}),\\
F_5 &= -i(J_{46} - J_{31}), & M_5 &= (i/\sqrt{3})(J_{46} + J_{31} - 2J_{57}), & N_5 &= i(J_{46} + J_{31} + J_{57}),\\
F_6 &= -i(J_{35} - J_{62}), & M_6 &= (i/\sqrt{3})(J_{35} + J_{62} - 2J_{71}), & N_6 &= i(J_{35} + J_{62} + J_{71}),\\
F_7 &= i(J_{65} - J_{23}), & M_7 &= (-i/\sqrt{3})(J_{65} + J_{23} - 2J_{47}), & N_7 &= i(J_{65} + J_{23} + J_{47}).
\end{aligned}$$

Note that the subscript A in M_A and N_A does not refer to the basis element left invariant by the corresponding canonical automorphism. We have used the above numbering to be consistent with Gell-Mann's notation for $SU(3)$, which is a subgroup of G_2 as is shown in the next section. The generators F_A and M_A, $A = 1, \ldots, 7$ close under commutation and form the Lie algebra of G_2. Denoting the Lie algebra of a group G by $\mathcal{L}G$ we have

$$\mathcal{L}G_2 = F_A \oplus M_A,$$
$$\mathcal{L}SO(7) = F_A \oplus M_A \oplus N_A.$$

The structure constants of $\mathcal{L}G_2$ are given in Appendix A. In the following sections, we shall denote the generators of $SO(7)$ by capital Latin letters F_A, M_A, and N_A and the corresponding $n\times n$ matrix representation of these generators by $\Lambda_A^{(n)}$, $\mu_A^{(n)}$, and $\nu_A^{(n)}$, respectively. The parameters corresponding to the generators M_A, N_A, and F_A will be denoted by m_A, n_A, and f_A, respectively.

5. THE $SU(3)$ AND $SU(2)\times SU(2)$ SUBGROUPS OF G_2

From the above table of the generators of G_2 one can easily observe that there are eight generators annihilating[24] a given basis element e_A. The generators annihilating, say, e_7 are F_A, $A = 1, \ldots, 7$, and $F_8 = -M_3$. They close under commutation and form the Lie algebra of $SU(3)$:

$$[F_a, F_b] = 2if_{abc}F_c, \qquad a, b, c = 1, \ldots, 8,$$

where f_{abc} are the usual structure constants of Gell-Mann. Hence the automorphisms of the Cayley algebra leaving a basis element e_A invariant form a subgroup $SU(3)$ of G_2. Since G_2 has only real representations,[25] only the real representations of $SU(3)$ can occur in the representations of G_2. For example, in the seven-dimensional representations of G_2, the only nontrivial real representation of $SU(3)$ that can occur is the 6-dimensional representation $3\oplus\bar{3}$. This six-dimensional representation can be constructed from Gell-Mann's matrices as follows:

$$\begin{aligned}
\Lambda_1^{(6)} &= \sigma_2 \otimes \lambda_1 = -i(\Sigma_{24} - \Sigma_{51}),\\
\Lambda_2^{(6)} &= 1_2 \otimes \lambda_2 = i(\Sigma_{54} - \Sigma_{12}),\\
\Lambda_3^{(6)} &= \sigma_2 \otimes \lambda_3 = -i(\Sigma_{14} - \Sigma_{25}),\\
\Lambda_4^{(6)} &= \sigma_2 \otimes \lambda_4 = -i(\Sigma_{16} - \Sigma_{43}),\\
\Lambda_5^{(6)} &= 1_2 \otimes \lambda_5 = -i(\Sigma_{46} - \Sigma_{31}),\\
\Lambda_6^{(6)} &= \sigma_2 \otimes \lambda_6 = -i(\Sigma_{35} - \Sigma_{62}),\\
\Lambda_7^{(6)} &= 1_2 \otimes \lambda_7 = i(\Sigma_{65} - \Sigma_{23}),\\
\Lambda_8^{(6)} &= \sigma_2 \otimes \lambda_8 = \frac{-i}{\sqrt{3}}(\Sigma_{14} + \Sigma_{25} - 2\Sigma_{36}),
\end{aligned} \tag{5.1}$$

where $\otimes$ denotes the direct product of matrices and λ_a are the Gell-Mann's λ matrices and σ_2 is the Pauli matrix

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad 1_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and Σ_{mn} are the 6×6 matrix representation of the generators J_{mn} of $SO(6)$. This construction shows clearly that $\Lambda_a^{(6)}$ can be imbedded into the seven-dimensional representations of G_2 as the representations of the generators F_a.

G_2 also has an $SU(2) \times SU(2)$ subgroup. The $SU(2) \times SU(2)$ subgroup involving the isospin subgroup of $SU(3)$ is generated by F_1, F_2, F_3 and M_1, M_2, M_3

$$\begin{aligned}
&[F_i, F_j] = 2i\epsilon_{ijk}F_k, \qquad i,j,k = 1,2,3,\\
&[F_i, M_j] = 0, \qquad [M_i, M_j] = (2i/\sqrt{3})\epsilon_{ijk}M_k.
\end{aligned} \tag{5.2}$$

$SU(2) \times SU(2)$ subgroup of G_2 arises from the fact that octonions can be constructed from two quaternions.[26]

The $SU(3)$ subgroup can be imbedded in G_2 in seven different ways and for each imbedding of $SU(3)$ there are three different imbeddings of $SU(2) \times SU(2)$ involving I, U, V spin subgroup of $SU(3)$.

Now consider the seven-dimensional action of G_2' on the octonion units e_A as the automorphism action. Then under the $SU(3)$ subgroup, six of the basis elements e_A transform like the six-dimensional real representation of $SU(3)$ $(3 \oplus \bar{3})$ and the seventh element is an $SU(3)$ scalar. Under $SU(2) \times SU(2)$, four of the elements e_A transform like the $(1/2, 1/2)$ representations and the remaining three transform as $(0, 1)$ representations.

6. SPLIT OCTONIONS AND SPLIT G_2

Above we have considered the automorphism group of real octonions with basis $1, e_A$. We saw that if we denote the parameters corresponding to the generators F_A and M_A by f_A and $\sqrt{3}m_A$ and the seven-dimensional representation of these generators by $\Lambda_A^{(7)}$ and $\mu_A^{(7)}$ then the most general automorphism of real octonions are given by the transformation

$$[e] \to [e'] = \exp[-if_A\Lambda_A^{(7)} - i\sqrt{3}m_A\mu_A^{(7)}][e] = e^X[e], \tag{6.1}$$

where

$$[e] = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{pmatrix} \quad \text{and} \quad X = -i(f_A\Lambda_A^{(7)} + \sqrt{3}m_A\mu_A^{(7)}).$$

X is given explicitly by

$$X = \begin{pmatrix}
0 & -(f_2+m_2) & -(f_5+m_5) & (m_3-f_3) & -(f_1+m_1) & (m_4-f_4) & 2m_6\\
(f_2+m_2) & 0 & -(f_7+m_7) & (m_1-f_1) & (f_3+m_3) & -(f_6+m_6) & 2m_4\\
(f_5+m_5) & (f_7+m_7) & 0 & -(f_4+m_4) & (m_6-f_6) & -2m_3 & 2m_1\\
(f_3-m_3) & (f_1-m_1) & (f_4+m_4) & 0 & (m_2-f_2) & (m_5-f_5) & 2m_7\\
(f_1+m_1) & -(f_3+m_3) & (f_6-m_6) & (f_2-m_2) & 0 & (m_7-f_7) & -2m_5\\
(f_4-m_4) & (f_6+m_6) & 2m_3 & (f_5-m_5) & (f_7-m_7) & 0 & 2m_2\\
-2m_6 & -2m_4 & -2m_1 & -2m_7 & 2m_5 & -2m_2 & 0
\end{pmatrix}$$

$$X^\dagger = -X$$

If we transform the real basis $[e]$ into what we shall call the split basis $[d]$ where

$$[d] = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_1^* \\ u_2^* \\ u_3^* \\ (i/\sqrt{2})e_7 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(e_1 + ie_4) \\ \frac{1}{2}(e_2 + ie_5) \\ \frac{1}{2}(e_3 + ie_6) \\ \frac{1}{2}(e_1 - ie_4) \\ \frac{1}{2}(e_2 - ie_5) \\ \frac{1}{2}(e_3 - ie_6) \\ (i/\sqrt{2})e_7 \end{bmatrix} = \begin{pmatrix} \psi(e_7) \\ \psi^*(e_7) \\ ie_7/\sqrt{2} \end{pmatrix}$$

then the automorphisms are generated by the mapping

$$[d] \to [d'] = e^{iZ}[d] \tag{6.2}$$

where Z is

$$Z = \begin{bmatrix} (f_3-m_3) & (f_1+if_2) & (f_4+if_5) & 0 & -(m_1-im_2) & (m_4+im_5) & -\sqrt{2}(m_6+im_7) \\ (f_1-if_2) & -(f_3+m_3) & (f_6+if_7) & (m_1-im_2) & 0 & -(m_6-im_7) & -\sqrt{2}(m_4-im_5) \\ (f_4-if_5) & (f_6-if_7) & 2m_3 & -(m_4+im_5) & (m_6-im_7) & 0 & -\sqrt{2}(m_1+im_2) \\ 0 & (m_1+im_2) & -(m_4-im_5) & -(f_3-m_3) & -(f_1-if_2) & -(f_4-if_5) & -\sqrt{2}(m_6-im_7) \\ -(m_1+im_2) & 0 & (m_6+im_7) & -(f_1+if_2) & (f_3+m_3) & -(f_6-if_7) & -\sqrt{2}(m_4+im_5) \\ (m_4-im_5) & -(m_6+im_7) & 0 & -(f_4+if_5) & -(f_6+if_7) & -2m_3 & -\sqrt{2}(m_1-im_2) \\ -\sqrt{2}(m_6-im_7) & -\sqrt{2}(m_4+im_5) & -\sqrt{2}(m_1-im_2) & -\sqrt{2}(m_6+im_7) & -\sqrt{2}(m_4-im_5) & -\sqrt{2}(m_1+im_2) & 0 \end{bmatrix}$$

$$Z^\dagger = Z,$$

and Z can be written in the form

$$Z = \begin{pmatrix} U_3 & O^\dagger & \mathbf{x}^T \\ O & -U_3^* & \mathbf{x}^\dagger \\ \mathbf{x}^* & \mathbf{x} & 0 \end{pmatrix},$$

where

$$U_3 = \begin{pmatrix} (f_3 - m_3) & (f_1 + if_2) & (f_4 + if_5) \\ (f_1 - if_2) & -(f_3 + m_3) & (f_6 + if_7) \\ (f_4 - if_5) & (f_6 - if_7) & 2m_3 \end{pmatrix},$$

$$O = \begin{pmatrix} 0 & (m_1 + im_2) & -(m_4 - im_5) \\ -(m_1 + im_2) & 0 & (m_6 + im_7) \\ (m_4 - im_5) & -(m_6 + im_7) & 0 \end{pmatrix},$$

$$\mathbf{x} = -\sqrt{2}[(m_6 + im_7)(m_4 - im_5)(m_1 + im_2)]$$

and $O_{ij} = -(1/\sqrt{2})\epsilon_{ijk}x_k$

$i, j, k = 1, 2, 3.$

Note that U_3 and O are the three-dimensional representations of the Lie algebras of $SU(3)$ and complex $SO(3)$.

If we further split the identity and consider the split octonions with basis $u_1u_2u_3u_0u_1^*u_2^*u_3^*u_0^*$, defined above, then the automorphism group G_2 will act on this basis by an 8-dimensional reducible representation.

$$G_2: [s] \to [s'] = e^{iY}[s], [s] = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_0 \\ u_1^* \\ u_2^* \\ u_3^* \\ u_0^* \end{bmatrix}, \tag{6.3}$$

where Y is given explicitly as

$$Y = \begin{bmatrix} (f_3-m_3) & (f_1+if_2) & (f_4+if_5) & -(m_6+im_7) & 0 & -(m_1-im_2) & (m_4+im_5) & (m_6+im_7) \\ (f_1-if_2) & -(f_3+m_3) & (f_6+if_7) & -(m_4-im_5) & (m_1-im_2) & 0 & -(m_6-im_7) & (m_4-im_5) \\ (f_4-if_5) & (f_6-if_7) & 2m_3 & -(m_1+im_2) & -(m_4+im_5) & (m_6-im_7) & 0 & (m_1+im_2) \\ -(m_6-im_7) & -(m_4+im_5) & -(m_1-im_2) & 0 & -(m_6+im_7) & -(m_4-im_5) & -(m_1+im_2) & 0 \\ 0 & -(m_1+im_2) & -(m_4-im_5) & -(m_6-im_7) & -(f_3-m_3) & -(f_1-if_2) & -(f_4-if_5) & (m_6-im_7) \\ -(m_1+im_2) & 0 & (m_6+im_7) & -(m_4+im_5) & -(f_1+if_2) & (f_3+m_3) & -(f_6-if_7) & (m_4+im_5) \\ (m_4-im_5) & -(m_6+im_7) & 0 & -(m_1-im_2) & -(f_4+if_5) & -(f_6+if_7) & -2m_3 & (m_1-im_2) \\ (m_6-im_7) & (m_4+im_5) & (m_1-im_2) & 0 & (m_6+im_7) & (m_4-im_5) & (m_1+im_2) & 0 \end{bmatrix}$$

$$Y^\dagger = Y,$$

and Y can be written in the form

$$Y = \begin{bmatrix} U_3 & \mathbf{x}^T/\sqrt{2} & O^\dagger & -\mathbf{x}^T/\sqrt{2} \\ \mathbf{x}^*/\sqrt{2} & 0 & \mathbf{x}/\sqrt{2} & 0 \\ O & \mathbf{x}^\dagger/\sqrt{2} & -U_3^* & -\mathbf{x}^\dagger/\sqrt{2} \\ -\mathbf{x}^*/\sqrt{2} & 0 & -\mathbf{x}/\sqrt{2} & 0 \end{bmatrix}$$

or alternatively as

$$Y = \begin{pmatrix} D & -E^* \\ E & -D^* \end{pmatrix},$$

where

$$D = \begin{pmatrix} U_3 & \mathbf{x}^T/\sqrt{2} \\ \mathbf{x}^*/\sqrt{2} & 0 \end{pmatrix}, \quad E = \begin{pmatrix} O & \mathbf{x}^\dagger/\sqrt{2} \\ -\mathbf{x}^*/\sqrt{2} & 0 \end{pmatrix}.$$

Note that the matrices D and E are not independent. D involves all the parameters of E. Keeping this point in mind, we see that $E \in$ complex $\mathcal{L}SO(4)$, $D \in \mathcal{L}SU(4)$. Matrices E close under the Lie product. Matrices D need one more generator to close under Lie product to form the four-dimensional representation of the Lie algebra of $SU(4)$.

The above form of G_2 as the automorphism group of split octonions is called the split G_2. Under the $SU(3)$ subgroup of split G_2 leaving u_0 and u_0^* invariant, the three split octonions (u_1, u_2, u_3) transform like a unitary triplet (quarks) and the complex conjugate octonions (u_1^*, u_2^*, u_3^*) transform like a unitary antitriplet (antiquarks). This property of split octonions is physically very important and plays a crucial role in obtaining a quark structure from the octonionic representations of Poincaré group.[14]

7. QUARK STRUCTURE IN THE SPLIT BASIS

To see another physically interesting property of split G_2 let us define the following basis for its Lie algebra[27]:

$$\begin{aligned}
&E_{12} = \tfrac{1}{2}(F_1 + iF_2), \quad E_{21} = \tfrac{1}{2}(F_1 - iF_2),\\
&E_{13} = \tfrac{1}{2}(F_4 + iF_5), \quad E_{31} = \tfrac{1}{2}(F_4 - iF_5),\\
&E_{23} = \tfrac{1}{2}(F_6 + iF_7), \quad E_{32} = \tfrac{1}{2}(F_6 - iF_7),\\
&F_3 = (E_{11} - E_{22}), \quad F_8 = (1/\sqrt{3})(E_{11} + E_{22} - 2E_{33}), \quad (7.1)\\
&Q_1 = \tfrac{1}{2}(M_6 + iM_7), \quad Q_1^\dagger = \tfrac{1}{2}(M_6 - iM_7),\\
&Q_2 = \tfrac{1}{2}(M_4 - iM_5), \quad Q_2^\dagger = \tfrac{1}{2}(M_4 + iM_5),\\
&Q_3 = \tfrac{1}{2}(M_1 + iM_2), \quad Q_3^\dagger = \tfrac{1}{2}(M_1 - iM_2),
\end{aligned}$$

where the expressions for Λ_3 and Λ_8 are purely formal at this point and will be explained shortly. In this basis commutation relations of split G_2 have the form:

$$\begin{aligned}
&[Q_i, Q_j] = -(2/\sqrt{3})\epsilon_{ijk}Q_k^\dagger,\\
&[Q_i, Q_j^\dagger] = T_{ij}, \qquad i,j = 1,2,3,\\
&[E_{ij}, Q_k] = \delta_{jk}Q_i,\\
&[T_{ii}, T_{jj}] = 0, \qquad [E_{ij}, E_{ji}] = (T_{ii} - T_{jj}), \qquad (7.2)\\
&[T_{ii}, E_{ij}] = E_{ij}, \qquad [T_{jj}, E_{ij}] = -E_{ij},\\
&[E_{ij}, E_{jk}] = E_{ik}, \qquad [E_{ji}, E_{kj}] = -E_{ki}
\end{aligned}$$

where T_{ij} is defined as

$$T_{ij} = E_{ij}, \qquad i \neq j,$$

and

$$\begin{aligned}
&T_{11} = \tfrac{1}{2}F_3 + (1/2\sqrt{3})F_8 = \tfrac{1}{3}(2E_{11} - E_{22} - E_{33}),\\
&T_{22} = -\tfrac{1}{2}F_3 + (1/2\sqrt{3})F_8 = \tfrac{1}{3}(-E_{11} + 2E_{22} - E_{33}),\\
&T_{33} = -(1/\sqrt{3})F_8 = \tfrac{1}{3}(-E_{11} - E_{22} + 2E_{33}). \qquad (7.3)
\end{aligned}$$

The generators T_{ij} form the subalgebra $SU(3)$. The generators F_3 and F_8 form a Cartan subalgebra of both $SU(3)$ and G_2. If we assign quantum numbers to the generators of split G_2, i.e., to its adjoint representation, using as the generators of third component of isospin and hypercharge the generators $I_3 = \frac{1}{2}F$ and $Y = (1/\sqrt{3})F_8$, we find that three quarks, three antiquarks, and eight mesons can be imbedded in the adjoint representation of split G_2, i.e., we can have the correspondence

$$\begin{aligned}
&Q_1 \leftrightarrow p \text{ quark}, \qquad Q_1^\dagger \leftrightarrow \bar{p},\\
&Q_2 \leftrightarrow n \text{ quark}, \qquad Q_2^\dagger \leftrightarrow \bar{n}\\
&Q_3 \leftrightarrow \lambda \text{ quark}, \qquad Q_3^\dagger \leftrightarrow \bar{\lambda}\\
&E_{12} \leftrightarrow \pi^+ (\text{or } \rho^+), \qquad E_{21} \leftrightarrow \pi^- (\text{or } \rho^-),\\
&E_{13} \leftrightarrow K^+ (K^{*+}), \qquad E_{31} \leftrightarrow K^- (K^{*-}),\\
&E_{23} \leftrightarrow K^0 (K^{*0}), \qquad E_{32} \leftrightarrow \overline{K^0}\ (\overline{K^{*0}}),\\
&\Lambda_3 \leftrightarrow \pi^0 (\rho^0), \qquad \Lambda_8 \leftrightarrow \eta\ (\omega_8),
\end{aligned}$$

This identification agrees with Gell-Mann's quark model in the assignment of the quantum numbers I_3 and Y and differs from it in the assignment of baryon number. If one uses the generator N_3 of $SO(7)$ as the baryon number generator, one gets the result that mesons are assigned zero baryon number as they must be but that the generators Q_i ($\leftrightarrow$ quarks) and the generators $Q_i^\dagger$ ($\leftrightarrow$ antiquarks) do not have well-defined baryon numbers.[28] These (pseudo-quark) generators Q_i have the interesting property that they generate the (anti-pseudo-quarks) $Q_i^\dagger$ under commutation, i.e.,

$$[Q_i, Q_j] = -(2/\sqrt{3})\epsilon_{ijk}Q_k^\dagger$$

and the $SU(3)$ subalgebra (mesons) under Lie triple product, i.e.

$$[Q_i, [Q_j, Q_k]] = -(2/\sqrt{3})\epsilon_{jkl}T_{il}. \qquad (7.4)$$

8. AN OCTONIONIC REPRESENTATION OF SPLIT G_2

The split octonions

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

transform as the three-dimensional irreducible representation of the $SU(3)$ subgroup of split G_2. But the elements

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_0 \end{pmatrix}$$

do not form a four-dimensional irreducible representation of split G_2. In fact the lowest nontrivial representation of G_2 is seven dimensional. Yet the action of G_2 on the basis

$$[s] = \begin{pmatrix} u \\ u^* \end{pmatrix}$$

is completely defined by its action on u, i.e., if

$$G_2: \begin{array}{l} u \to u' \\ u^* \to (u^*)' \end{array},$$

then

$$(u^*)' = (u')^*.$$

The action of G_2 generators on u can be represented by multiplication with octonion units in the following compact form:

$$E_{ij}u = u_i(u_j^* u), \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_0 \end{pmatrix}, \quad i,j = 1,2,3,$$

$$\begin{aligned}
Q_i u &= -(1/\sqrt{3})uu_i + (1/\sqrt{3})(u_i u_0)u = -(1/\sqrt{3})uu_i,\\
Q_i^\dagger u &= -(1/\sqrt{3})uu_i^* + (1/\sqrt{3})(u_i^* u_0)u = -(1/\sqrt{3})[u, u_i^*].
\end{aligned}$$

with

$$\begin{aligned}
F_3 u &= (E_{11} - E_{22})u = u_1(u_1^* u) - u_2(u_2^* u),\\
F_8 u &= (1/\sqrt{3})(E_{11} + E_{22} - 2E_{33})u\\
&= (1/\sqrt{3})\{u_1(u_1^* u) + u_2(u_2^* u) - 2u_3(u_3^* u)\},
\end{aligned}$$

which justifies the formal expressions for F_3 and F_8 given above. Thus the above form of the Lie algebra action of G_2 generates an octonionic representation of split $\mathcal{L}G_2$. The automorphism group of real octonions can also be shown in this form because a real octonion

$$\Phi = \Phi_0 + \Phi_A e_A$$

can be written as

$$\begin{aligned}\Phi = 2\ \mathrm{Re}\{&(\phi_0 - i\phi_7)\tfrac{1}{2}(1 + ie_7)\\ &+ (\phi_1 - i\phi_4)\tfrac{1}{2}(e_1 + ie_4)\\ &+ (\phi_2 - i\phi_5)\tfrac{1}{2}(e_2 + ie_5)\\ &+ (\phi_3 - i\phi_6)\tfrac{1}{2}(e_3 + ie_6)\},\end{aligned}$$

where Re refers to the real part with respect to the complex unit i. Then

$$\Phi = \phi^\dagger u + \text{c.c.} = \phi^\dagger u + \phi^T u^*,$$

where

$$\phi = \begin{pmatrix}\phi_1 + i\phi_4\\ \phi_2 + i\phi_5\\ \phi_3 + i\phi_6\\ \phi_0 + i\phi_7\end{pmatrix} \quad u = \begin{pmatrix}u_1\\ u_2\\ u_3\\ u_0\end{pmatrix}$$

Therefore, the action of Aut**O** on e_A is completely defined by its action on u.

9. 8 x 8 MATRIX FORMULATION OF THE CAYLEY ALGEBRA

In Appendix D, we give two constructions of Cayley algebra in terms of 3×3 λ-matrices and 4×4 γ-matrices. In this section, we shall study the 8×8 matrix construction of octonions. Consider the column matrix

$$[s] = \begin{bmatrix}u\\ u^*\end{bmatrix}$$

of split octonions. Define the conjugate matrix $[\bar{s}]^\dagger$ as $[\bar{s}]^\dagger = [\bar{s}]^{*T}$

$$= (-u_1^*, -u_2^*, -u_3^*, u_0, -u_1, -u_2, -u_3, u_0^*),$$

where the overbar denotes octonion conjugation, * denotes complex conjugation, and T is the usual transposition. Then the product $[s][\bar{s}]^\dagger$ can be written in the form

$$[s][\bar{s}]^\dagger = \tfrac{1}{2}(1 - i\Gamma_A e_A), \quad A = 1, \ldots, 7, \tag{9.1}$$

where Γ_A are 8×8 matrices given by:

$$\begin{aligned}\Gamma_1 &= -\sigma_1 \otimes \sigma_1 \otimes \sigma_2 = -\tau_1\rho_1\sigma_2,\\ \Gamma_2 &= -\sigma_1 \otimes \sigma_2 \otimes I = -\tau_1\rho_2,\\ \Gamma_3 &= \sigma_1 \otimes \sigma_3 \otimes \sigma_2 = \tau_1\rho_3\sigma_2,\\ \Gamma_4 &= \sigma_2 \otimes \sigma_2 \otimes \sigma_1 = \tau_2\rho_2\sigma_1,\\ \Gamma_5 &= -\sigma_2 \otimes \sigma_2 \otimes \sigma_3 = -\tau_2\rho_2\sigma_3,\\ \Gamma_6 &= \sigma_2 \otimes I \otimes \sigma_2 = \tau_2\sigma_2,\\ \Gamma_7 &= -\sigma_3 \otimes I \otimes I = -\tau_3\end{aligned} \tag{9.2}$$

($\otimes$ denotes direct product of the Pauli matrices $\sigma_1, \sigma_2, \sigma_3, I$),

where we have defined

$$\begin{aligned}I \otimes I \otimes \sigma_i &= \sigma_i,\\ I \otimes \sigma_i \otimes I &= \rho_i,\\ \sigma_i \otimes I \otimes I &= \tau_i,\end{aligned}$$

and chosen a representation in which $\Gamma_1, \Gamma_2, \Gamma_3$ are imaginary and $\Gamma_4, \Gamma_5, \Gamma_6, \Gamma_7$ are real. These seven matrices Γ_A are Hermitian and satisfy the anticommutation relations

$$\{\Gamma_A, \Gamma_B\} = 2\delta_{AB}, \quad \Gamma_A^\dagger = \Gamma_A. \tag{9.3}$$

Now number the rows of the column vector $[s]$ from 1 to 8 and define a mapping L_{u_i} on $[s]$ as the mapping induced by multiplication from the left by the element u_i.[29] Then a simple calculation gives the result that

$$\begin{aligned}L^{(8)}_{u_1} + L^{(8)}_{u_1^*}\ \ L^{(8)}_{u_1+u_1^*} &= L^{(8)}_{e_1} = \bar{\Sigma}_{18} - \bar{\Sigma}_{27} + \bar{\Sigma}_{36} - \bar{\Sigma}_{45},\\ L^{(8)}_{u_1-u_1^*} &= L^{(8)}_{ie_4} = \bar{E}_{81} + \bar{E}_{18}\\ &\qquad - \bar{E}_{54} - \bar{E}_{45} - \bar{E}_{36} - \bar{E}_{63} + \bar{E}_{27} + \bar{E}_{72},\\ L^{(8)}_{u_2+u_2^*} &= L^{(8)}_{e_2} = \bar{\Sigma}_{28} - \bar{\Sigma}_{46} + \bar{\Sigma}_{53} + \bar{\Sigma}_{17},\\ L^{(8)}_{u_2-u_2^*} &= L^{(8)}_{ie_5} = -\bar{E}_{71} - \bar{E}_{17}\\ &\qquad + \bar{E}_{53} + \bar{E}_{35} - \bar{E}_{46} - \bar{E}_{64} + \bar{E}_{28} + \bar{E}_{82},\\ L^{(8)}_{u_3+u_3^*} &= L^{(8)}_{e_3} = \bar{\Sigma}_{61} - \bar{\Sigma}_{52} - \bar{\Sigma}_{47} + \bar{\Sigma}_{38},\\ L^{(8)}_{u_3-u_3^*} &= L^{(8)}_{ie_6} = \bar{E}_{61} + \bar{E}_{16}\\ &\qquad - \bar{E}_{52} - \bar{E}_{25} - \bar{E}_{47} - \bar{E}_{74} + \bar{E}_{38} + \bar{E}_{83},\\ L^{(8)}_{u_4-u_4^*} &= L^{(8)}_{ie_7} = \bar{E}_{11} + \bar{E}_{22}\\ &\qquad + \bar{E}_{33} + \bar{E}_{44} - \bar{E}_{55} - \bar{E}_{66} - \bar{E}_{77} - \bar{E}_{88},\end{aligned} \tag{9.4}$$

where $\bar{E}_{ij}$ are the 8×8 matrix units and $\bar{\Sigma}_{ab} = \bar{E}_{ab} - E_{ba}$

Comparing the matrices Γ_A with the mappings L_{e_A} considered as matrices acting on the basis $[s]$ we have

$$\begin{aligned}L^{(8)}_{e_1} &= -i\Gamma_1, & L^{(8)}_{e_4} &= i\Gamma_4, & &\\ L^{(8)}_{e_2} &= -i\Gamma_2, & L^{(8)}_{e_5} &= i\Gamma_5, & L^{(8)}_{e_7} &= i\Gamma_7 \quad \text{on} \begin{bmatrix}u\\ u^*\end{bmatrix}.\\ L^{(8)}_{e_3} &= -i\Gamma_3, & L^{(8)}_{e_6} &= i\Gamma_6, & &\end{aligned} \tag{9.5}$$

From these equalities, the anticommutation relations of Γ_A follow automatically, since

$$L_{e_A}L_{e_B} + L_{e_B}L_{e_A} = L_{e_A e_B + e_B e_A} = L_{-2\delta_{AB}} \tag{9.6}$$

which in turn follows from the identity

$$O_1(O_2O_3) + O_2(O_1O_3) = (O_1O_2 + O_2O_1)O_3, \qquad O_1, O_2, O_3 \in O \tag{9.7}$$

for octonions.

Now define the matrices Γ_{AB} as

$$\begin{aligned}\Gamma_{AB} &= (1/2i)[\Gamma_A, \Gamma_B],\\ \Gamma_{AB} &= -\Gamma_{BA}, \quad \Gamma_{AB}^\dagger = \Gamma_{AB}.\end{aligned} \tag{9.8}$$

Twenty-one matrices Γ_{AB} form the Lie algebra of Spin (7) and $\Gamma_{AB} \oplus \Gamma_A$ form the Lie algebra of $SO(8)$.

Having constructed the matrices Γ_A and Γ_{AB} from the spinor $[s]$, we can forget about the octonionic character of $[s]$ and consider an eight-component spinor Ψ. Then

$$\{V_8 = \Psi^\dagger\Psi,\ V_A = \Psi^\dagger\Gamma_A\Psi\}$$

transform like a vector under $SO(8)$. Under the subgroup Spin (7), $\Psi^\dagger\Psi$ is a scalar and $\Psi^\dagger\Gamma_A\Psi = V_A$ is a vector. To characterize G_2, we need one more condition in addition to the requirement that $\Psi^\dagger\Psi$ be a scalar. Now G_2 is the automorphism group of octonions and it leaves the identity invariant. Therefore we would expect the G_2 subgroup of $SO(7)$ to leave $(\Psi_4^* + \Psi_8^*)(\Psi_4 + \Psi_8)$ invariant, since $(\Psi_4 + \Psi_8)$ corresponds to the identity of the octonions in the nonoctonionic formulation considered here. Thus, under G_2 both $\Psi^\dagger\Psi$ and $\Psi^\dagger K\Psi$ are scalars, where K is the matrix

$$K = \begin{pmatrix} 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&1\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&1 \end{pmatrix}.$$

In fact, the assertion that $\Psi^\dagger K\Psi$ is a scalar under G_2 can be rigorously proved by showing that only those linear combinations of the generators Γ_{AB} of Spin (7) that belong to G_2 commute with the matrix K. To do this it is convenient to use the following expression for K

$$K = \tfrac{1}{4}[\mathbf{1} - i(1/3!)a_{ABC}\Gamma_A\Gamma_B\Gamma_C], \qquad A,B,C, = 1,\ldots,7, \tag{9.9}$$

where a_{ABC} is a totally antisymmetric tensor and satisfies

$$a_{ABC} = 1 \text{ for } ABC = 123, 246, 435, 516, 572, 471, 673.$$

The matrix K is related to the octonion conjugation matrix O^c defined by $O^c[s] = [\bar{s}]$ as

$$O^c = K - \mathbf{1} \quad \text{or} \quad K = \mathbf{1} + O^c.$$

Therefore the conditions that $\Psi^\dagger K\Psi$ be a scalar is equivalent to the condition that $\Psi^\dagger O^c\Psi$ be a scalar.

The conditions that characterize G_2, i.e., that $\Psi^\dagger\Psi$ and $\Psi^\dagger K\Psi$ be invariant, are equivalent to saying that G_2 is the common subgroup of $SO(7)$ and Spin (7),[30] i.e.,

$$G_2 = \text{Spin}\ (7) \cap SO(7).$$

Above we showed that the matrices Γ_A correspond to the left multiplication by octonion units e_A acting on $[s]$. This does not mean that e_A can be represented by matrices Γ_A to form a Cayley algebra under the usual matrix multiplication. To get a Cayley algebra from Γ_A, we define the product of two Γ matrices as

$$\Gamma_A \circ \Gamma_B = \tfrac{1}{2}\{\Gamma_A, \Gamma_B\} + \tfrac{1}{2}\{[\Gamma_A, \Gamma_B], M\} + \Gamma_A M \Gamma_B - \Gamma_B M \Gamma_A, \tag{9.10}$$

where

$$M = K - \tfrac{1}{4}\mathbf{1} = -\tfrac{1}{4}i(1/3!)a_{ABC}\Gamma_A\Gamma_B\Gamma_C,$$

which gives

$$\Gamma_A \circ \Gamma_B = \delta_{AB} + i a_{ABC}\Gamma_C, \qquad A,B,C = 1,\ldots,7.$$

Hence defining the multiplication by a multiple c of the identity $\mathbf{1}$ as multiplication by the scalar c, we get a Cayley algebra with basis

$$\begin{aligned} &e_A \equiv -i\Gamma_A, \qquad 1 \equiv \mathbf{1},\\ &(-i\Gamma_A)\circ(-i\Gamma_B) = -\delta_{AB} + a_{ABC}(-i\Gamma_C). \end{aligned} \tag{9.11}$$

10. IMBEDDING IN $SO(7)$ AND $SO(8)$

In the above, we have decomposed the Lie algebra of $SO(7)$ as

$$SO(7) = F_A \oplus M_A \oplus N_A,$$

where $F_A \oplus M_A$ generate the subgroup G_2.

The 8-dimensional representation of G_2 as the automorphism group of octonions acting on the basis

$$[s] = \begin{bmatrix} u \\ u^* \end{bmatrix}$$

will induce an 8-dimensional spinor representation of $SO(7)$: In fact, after some algebra, one finds that the action of N_A on $[s]$ can be represented as

$$\begin{aligned} &\nu_1^{(8)} = \tfrac{1}{2}i(L_{e_6}^{(8)} - R_{e_6}^{(8)}), &&\nu_2^{(8)} = -\tfrac{1}{2}i(L_{e_3}^{(8)} - R_{e_3}^{(8)}),\\ &\nu_3^{(8)} = \tfrac{1}{2}i(L_{e_7}^{(8)} - R_{e_7}^{(8)}), &&\nu_5^{(8)} = -\tfrac{1}{2}i(L_{e_2}^{(8)} - R_{e_2}^{(8)}),\\ &\nu_4^{(8)} = \tfrac{1}{2}i(L_{e_5}^{(8)} - R_{e_5}^{(8)}), &&\nu_7^{(8)} = -\tfrac{1}{2}i(L_{e_1}^{(8)} - R_{e_1}^{(8)}),\\ &\nu_6^{(8)} = \tfrac{1}{2}i(L_{e_4}^{(8)} - R_{e_4}^{(8)}), \end{aligned} \tag{10.1}$$

where L_{e_A} and R_{e_A} are left and right multiplications by the element e_A, respectively. Explicit matrix form of L_{e_A} was given above. For the R_{e_A} we have

$$\begin{aligned} R_{e_1}^{(8)} &= \bar{\Sigma}_{27} + \bar{\Sigma}_{58} + \bar{\Sigma}_{14} + \bar{\Sigma}_{63},\\ R_{e_2}^{(8)} &= \bar{\Sigma}_{71} + \bar{\Sigma}_{35} + \bar{\Sigma}_{24} + \bar{\Sigma}_{68},\\ R_{e_3}^{(8)} &= \bar{\Sigma}_{78} + \bar{\Sigma}_{52} + \bar{\Sigma}_{34} + \bar{\Sigma}_{16},\\ R_{e_4}^{(8)} &= -i\{(\bar{E}_{14} + \bar{E}_{41})\\ &\quad + (\bar{E}_{36} + \bar{E}_{63}) - (\bar{E}_{58} + \bar{E}_{85}) - (\bar{E}_{27} + \bar{E}_{72})\},\\ R_{e_5}^{(8)} &= -i\{(\bar{E}_{17} + \bar{E}_{71})\\ &\quad + (\bar{E}_{24} + \bar{E}_{42}) - (\bar{E}_{35} + \bar{E}_{53}) - (\bar{E}_{68} + \bar{E}_{86})\},\\ R_{e_6}^{(8)} &= -i\{(\bar{E}_{25} + \bar{E}_{52})\\ &\quad + (\bar{E}_{34} + \bar{E}_{43}) - (\bar{E}_{16} + \bar{E}_{61}) - (\bar{E}_{78} + \bar{E}_{87})\},\\ R_{e_7}^{(8)} &= -i\{(\bar{E}_{44} + \bar{E}_{55}\\ &\quad + \bar{E}_{66} + \bar{E}_{77}) - (\bar{E}_{11} + \bar{E}_{22} + \bar{E}_{33} + \bar{E}_{88})\}. \end{aligned} \tag{10.2}$$

We have shown that the 8×8 matrices Γ_A and Γ_{AB} form an eight-dimensional representation of the Lie algebra of $SO(8)$. Since Γ_A correspond to the left multiplication by e_A acting on $[s]$, we have the result that the eight-dimensional representation of $\mathcal{L}SO(8)$ can also be decomposed as:

$$\mathcal{L}SO(8) = \Lambda_A^{(8)} \oplus \mu_A^{(8)} \oplus \nu_A^{(8)} \oplus \xi_A^{(8)},$$

where $\xi_A^{(8)} = \frac{1}{2}i(L_{e_A}^{(8)} + R_{e_A}^{(8)})$ corresponding to the generator

$$Z_A = \tfrac{1}{2}i(L_{e_A} + R_{e_A}).$$

Since the group $SO(8)$ has rank four, its Cartan subalgebra will be four dimensional. One can redefine the generators of $SO(8)$ such that F_3, M_3, N_3, and Z_3 form a Cartan subalgebra.

If we take $I_3 = \frac{1}{2}F_3$, $Y_3 = -(1/\sqrt{3})M_3$, $B = -\frac{1}{3}N_3$, $\frac{1}{2}Z_3$ as the Cartan subalgebra generators, we can assign the following quantum numbers to the basis elements:

	I_3	Y_3	B	$\frac{1}{2}Z_3$
u_1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{3}$	0
u_1^*	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0
u_2	$-\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{3}$	0
u_2^*	$\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0
u_3	0	$-\frac{2}{3}$	$\frac{1}{3}$	0
u_3^*	0	$+\frac{2}{3}$	$-\frac{1}{3}$	0
u_0	0	0	0	-1
u_0^*	0	0	0	1

Therefore, under the correspondence

$(u_1, u_2, u_3) \leftrightarrow (p, n, \lambda)$ quarks

$(u_1^*, u_2^*, u_3^*) \leftrightarrow (\bar{p}, \bar{n}, \bar{\lambda})$ antiquarks

$(u_0, u_0^*) \leftrightarrow$ (core, anticore)

we have the result that I_3, Y_3, and B act like the generators of third component of isospin, hypercharge, and baryon number. Subscript 3 in Y_3 refers to the fact that within G_2, Y is the generator of the third component of an $SU(2)$ subgroup just as I_3 is.

11. REDUCTION WITH RESPECT TO THE $SU(2) \times SU(2)$ [I-SPIN-G-SPIN] SUBGROUP OF G_2

The generators $I_i = F_i$, $G_i = \sqrt{3}M_i$, $i = 1, 2, 3$ form an $SU(2) \times SU(2)$ subalgebra of $\mathcal{L}G_2$:

$$[I_i, I_j] = 2i\epsilon_{ijk}I_k, \quad [G_i, G_j] = 2i\epsilon_{ijk}G_k,$$
$$[I_i, G_j] = 0, \quad i,j,k = 1,2,3. \tag{11.1}$$

I_i is the isospin subalgebra of the $SU(3)$ subalgebra of $\mathcal{L}G_2$ annihilating the basis element e_7. Now the spinors

$$\psi = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \leftrightarrow \begin{pmatrix} p \\ n \end{pmatrix}$$

and

$$\psi^* = \begin{pmatrix} u_1^* \\ u_2^* \end{pmatrix} \leftrightarrow \begin{pmatrix} \bar{p} \\ \bar{n} \end{pmatrix}$$

correspond to isospin doublets and the elements u_3, u_3^*, $(1/\sqrt{2})ie_7$ are isospin scalars.

Consider the infinitesimal group action generated by G_i

$$G: \psi \to \psi' = (1 - im^3)\psi - (m^2 + im^1)\psi^G,$$
$$\psi^{G\prime} = (m^2 - im^1)\psi + (1 + im^3)\psi^G,$$

where ψ^G is the G parity conjugate spinor defined by

$$\psi^G = i\tau_2\psi^* \quad \text{and} \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

Therefore under the G-spin subgroup (generated by G_1) of G_2 the spinor ψ and ψ^G form a G-spin doublet and transform as

$$G: \begin{pmatrix} \psi \\ \psi^G \end{pmatrix} \to \begin{pmatrix} \psi' \\ \psi^{G\prime} \end{pmatrix} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}\begin{pmatrix} \psi \\ \psi^G \end{pmatrix}, \tag{11.2}$$

where

$$|a|^2 + |b|^2 = 1.$$

Similarly we find that

$$\phi = \begin{pmatrix} u_3 \\ ie_7/\sqrt{2} \\ u_3^* \end{pmatrix}$$

forms a G-spin triplet which transforms infinitesimally as

$$G: \phi \to \phi' = \begin{pmatrix} 1 + 2im^3 & -i\sqrt{2}(m^1 + im^2) & 0 \\ -i\sqrt{2}(m^1 - im^2) & 1 & -i\sqrt{2}(m^1 + im^2) \\ 0 & -i\sqrt{2}(m^1 - im^2) & 1 - 2im^3 \end{pmatrix}\begin{pmatrix} u_3 \\ ie_7/\sqrt{2} \\ u_3^* \end{pmatrix},$$

the global form of which is

$$G: \phi \to \phi' = \begin{pmatrix} a^2 & \sqrt{2}ab & b^2 \\ -\sqrt{2}ab^* & |a|^2 - |b|^2 & \sqrt{2}a^*b \\ b^{*2} & -\sqrt{2}a^*b^* & a^{*2} \end{pmatrix}\phi. \tag{11.3}$$

An important property of G-spin is that its third component is proportional to hypercharge Y, i.e.,

$$Y = -\tfrac{1}{3}G_3$$

and hence it should properly be called hypercharge spin. This hypercharge-spin subgroup of G_2 commutes with the isospin subgroup generated by F_1, F_2, F_3. The isospin and hypercharge spin groups together generate a four-dimensional rotation group $SU(2) \times SU(2)$ which has been considered before[31] as applied to an isotopic doublet such as the nucleon or the p and n quarks. The multiplets $(u_1, u_2, u_2^*, -u_1^*)$ and $(u_3, (ie_7/\sqrt{2}), u_3^*)$ form the (1/2, 1/2) and (0, 1) representations of the subgroup $SU(2)_I \otimes SU(2)_Y$, respectively. The $SU(3)$ singlet $(ie_7/\sqrt{2})$ is not an hypercharge spin singlet. It transforms like the third component of an hypercharge triplet. We shall call it the vacuan v. Therefore the lowest-dimensional representation of G_2 has the root system shown is Fig. 3. Above we defined the G-parity conjugate spinor ψ^G of an isospin doublet ψ as

$$\psi^G = Ce^{i\pi I_2}\psi,$$

where I_2 is the second component of isospin and C is charge conjugation which in our case is taken as complex conjugation. We will generalize this G-parity concept to other charge space $SU(2)$ groups as follows: Write the above equation as

$$\Psi_I^{G^I} = Ce^{i\pi I_2}\psi_I; \tag{11.3a}$$

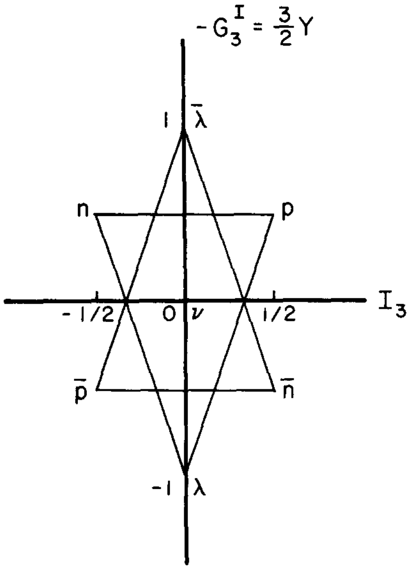

FIG. 3.

then for U and V spins we can define

$$\psi_U^{G^U} = Ce^{i\pi U_2}\psi_U \tag{11.3b}$$

$$\psi_V^{G^V} = Ce^{i\pi V_2}\psi_V \tag{11.3c}$$

Then under the $SU(2)_U \otimes SU(2)_{G^U}$ subgroup of G_2 generated by

$$\begin{aligned} &U_1 = \tfrac{1}{2}F_6, \quad U_2 = \tfrac{1}{2}F_7, \\ &U_3 = \tfrac{1}{4}(-\sqrt{3}M_3 - F_3), \\ &G_1^U = (\sqrt{3}/2)M_6, \quad G_2^U = (\sqrt{3}/2)M_7, \\ &G_3^U = (\sqrt{3}/4)\left(\frac{1}{\sqrt{3}}F_3 - M_3\right). \end{aligned} \tag{11.4a}$$

The spinor

$$\begin{pmatrix} \psi_U \\ \psi_U^{G^U} \end{pmatrix}$$

transforms like the (1/2, 1/2) representation, where

$$\psi_U = \begin{pmatrix} u_2 \\ u_3 \end{pmatrix} \leftrightarrow \begin{pmatrix} n \\ \lambda \end{pmatrix},$$

and the spinor

$$\phi_U = \begin{pmatrix} u_1 \\ (i/\sqrt{2})e_7 \\ u_1^* \end{pmatrix}$$

will transform as the (0, 1) representation. Same thing applies for the $SU(2)_V \otimes SU(2)_{G^V}$ subgroup generated by

$$\begin{aligned} &V_1 = \tfrac{1}{2}F_4, \quad V_2 = \tfrac{1}{2}F_5, \\ &V_3 = \tfrac{1}{4}(F_3 - \sqrt{3}M_3), \\ &G_1^V = (\sqrt{3}/2)M_4, \quad G_2^V = (\sqrt{3}/2)M_5, \end{aligned} \tag{11.4b}$$

$$G_3^V = (\sqrt{3}/4)(M_3 + \sqrt{3}F_3), \tag{11.4c}$$

except we have to replace ψ_U by

$$\psi_V = \begin{pmatrix} u_3 \\ u_1 \end{pmatrix} \leftrightarrow \begin{pmatrix} \lambda \\ p \end{pmatrix}$$

and ϕ_U by

$$\phi_V = \begin{pmatrix} u_2 \\ (i/\sqrt{2})e_7 \\ u_2^* \end{pmatrix}.$$

Hence, we have the result that just as the group $SU(3)$ contains three overlapping $SU(2)$ groups corresponding to I, U, V spins, the group G_2 contains three overlapping $SU(2) \times SU(2)_G$ groups corresponding to I, U, V spins together with their generalized G-parity extensions. We shall call the $SU(2)_G$ groups corresponding to I, U, V spins, the hypercharge spin, charge spin, and hypocharge spin groups, respectively. Under the $SU(2) \times SU(2)_G$ subgroup, the adjoint representation of G_2 decomposes as

$$14 = (1, 0) \oplus (0, 1) \oplus (1/2, 3/2).$$

So far we have considered the decomposition of $\mathcal{L}SO(8)$ in terms of seven anticommutating matrices Γ_A, which correspond to left multiplication by the basis elements e_A of octonions acting on the basis

$$[s] = \begin{vmatrix} u \\ u^* \end{vmatrix}.$$

A more consistent approach is to consider matrices corresponding to multiplication by the split octonions: Since

$$L_{u_i} = L_{1/2(e_i + ie_{i+3})} = \tfrac{1}{2}(L_{e_i} + iL_{e_{i+3}}), \qquad i = 1, 2, 3,$$

we have

$$\begin{aligned} &L_{u_i}^{(8)} = U_i = -\tfrac{1}{2}(\Gamma_{i+3} + \Gamma_i), \\ &L_{u_i^*}^{(8)} = \tilde{U}_i = \tfrac{1}{2}(\Gamma_{i+3} - i\Gamma_i), \\ &L_{u_0}^{(8)} = U_0 = \tfrac{1}{2}(1 - \Gamma_7), \\ &L_{u_0^*}^{(8)} = \tilde{U}_0 = \tfrac{1}{2}(1 + \Gamma_7). \end{aligned} \tag{11.5}$$

One can also change the basis on which the octonion units act and consider the real octonion basis on which real or split octonions may act. In any case, octonion multiplication operators L_a or R_a, $a \in \mathbf{O}$ is uniquely defined and choosing different bases on which they can act changes their matrix representations.

12. LIE MULTIPLICATION ALGEBRA OF OCTONIONS AND THE PRINCIPLE OF TRIALITY

A derivation D of an algebra A is defined as a linear transformation satisfying the property:

$$D(xy) = (Dx)y + x(Dy) \quad \text{for all } x, y \in A. \tag{12.1}$$

Derivations of an algebra A form a Lie algebra under Lie product, i.e.,

$$[D_i, D_j]_- = -[D_j, D_i]$$

and for all $D_i, D_j, D_k \in \mathrm{Der}A$,

$$[D_i, [D_j, D_k]] + [D_k, [D_i, D_j]] + [D_j, [D_k, D_i]] = 0,$$

Jacobi identity.

Derivation algebra of an algebra A is isomorphic to the Lie algebra of the automorphism group of A,[32] i.e., if $D \in \mathrm{Der}\ A$

$$D(xy) = (Dx)y + x(Dy) \Rightarrow e^D(xy) = (e^Dx)(e^Dy).$$

Therefore the derivation (Lie) algebra of octonions is isomorphic to the Lie algebra of G_2. Lie multiplication algebra of the octonions is defined as the Lie algebra with elements:

$$\mathcal{L}M\mathbf{O} = \mathrm{Der}\,\mathbf{O} \oplus L_{\mathbf{O}_0} \oplus R_{\mathbf{O}_0},$$

where $L_{\mathbf{O}_0}$ and $R_{\mathbf{O}_0}$ correspond to multiplication from the left and the right by traceless (or imaginary) octonion units. Since the octonions are not associative left and right multiplications do not commute. As was shown above, the Lie multiplication algebra of octonions is isomorphic to the Lie algebra of the group $SO(8)$.

$$\mathcal{L}SO(8) = F_A \oplus M_A \oplus N_A \oplus Z_A, \tag{12.2}$$

where

$$\mathrm{Der}\,\mathbf{O} \cong F_A \oplus M_A.$$

The usual real octonionic norm is invariant under the group $SO(8)$. Denoting the parameters corresponding to the generators $\Lambda_A^{(8)}, \mu_A^{(8)}, \nu_A^{(8)}, \xi_A^{(8)}$ by f_A, m_A, n_A, z_A, we can represent the action of $SO(8)$ on the split octonion basis $[s]$ by

$$SO(8): [s] \to e^{iL}[s], \qquad [s] = \begin{bmatrix} u \\ u^* \end{bmatrix}, \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_0 \end{pmatrix},$$

$$L = \begin{pmatrix} U & V \\ V^\dagger & -U_4^* \end{pmatrix},$$

where

$$U_4 = \begin{pmatrix} (f_3 - m_3 - n_3) & (f_1 + if_2) & (f_4 + if_5) & (\zeta_1 + i\zeta_2) \\ (f_1 - if_2) & -(f_3 + m_3 + n_3) & (f_6 + if_7) & (\zeta_3 + i\zeta_4) \\ (f_4 - if_5) & (f_6 - if_7) & (2m_3 - n_3) & (\zeta_5 + i\zeta_6) \\ (\zeta_1 - i\zeta_2) & (\zeta_3 - i\zeta_4) & (\zeta_5 - i\zeta_6) & -2z_3 \end{pmatrix}, \tag{12.3}$$

where

$$\zeta_1 = -m_6 + \tfrac{1}{2}(n_6 - z_6), \qquad \zeta_4 = m_5 - \tfrac{1}{2}(n_5 + z_5),$$

$$\zeta_2 = -m_7 - \tfrac{1}{2}(n_7 + z_7), \qquad \zeta_5 = -m_1 + \tfrac{1}{2}(n_1 - z_1),$$

$$\zeta_3 = -m_4 + \tfrac{1}{2}(n_4 - z_4), \qquad \zeta_6 = -m_2 - \tfrac{1}{2}(n_2 + z_2).$$

Matrices U_4 close under commutation and form the four-dimensional representation of $\mathcal{L}U(4)$.

The matrices V are antisymmetric.

$V_{\mu\nu} = -V_{\nu\mu}$ and form the Lie algebra of complex $SO(4)$:

$$V_{12} = -(m_1 + n_1) + i(m_2 - n_2),$$

$$V_{13} = (m_4 + n_4) + i(m_5 + n_5),$$

$$V_{14} = [m_6 - \tfrac{1}{2}(n_6 + z_6)] + i[m_7 + \tfrac{1}{2}(n_7 - z_7)],$$

$$V_{23} = -(m_6 + n_6) + i(m_7 - n_7),$$

$$V_{24} = [m_4 - \tfrac{1}{2}(n_4 + z_4)] - i[m_5 - \tfrac{1}{2}(n_5 - z_5)],$$

$$V_{34} = [m_1 - \tfrac{1}{2}(n_1 + z_1)] + i[m_2 + \tfrac{1}{2}(n_2 - z_2)].$$

Denoting U_4 as U, we have that U and V can be decomposed as

$$U = U_{G_2} \oplus U_{SO(8)/G_2}$$

$$L = L_{G_2} \oplus L_{SO(8)/G_2}, \tag{12.4}$$

$$V = V_{G_2} \oplus V_{SO(8)/G_2}$$

where $\oplus$ refers to vector space direct sum and V_{G_2}, U_{G_2} involve only the parameters f_A and m_A and $U_{SO(8)/G_2}$ and $V_{SO(8)/G_2}$ involve only n_A and z_A. Below we will construct the $\mathcal{L}SO(8)$ matrices that are in local triality with each other (see Appendix C for the principle of triality.) The principle of local triality states that for a given matrix $T^L \in \mathcal{L}SO(8)$ acting on the 8-dimensional space of octonions and which is skew with respect to the natural bilinear form (x, y) defined over the octonions, there exist uniquely determined matrices T^R and T^P belonging to the Lie algebra of $SO(8)$ (i.e., which are skew with respect to the norm form) such that

$$T^P(xy) = (T^L x)y + x(T^R y), \qquad x, y \in \mathbf{O},$$
$$T^P, T^L, T^R \in \mathcal{L}SO(8). \tag{12.5}$$

Decomposing T^P and T^R and T^L as above,

$$T^L = T^L_{G_2} \oplus T^L_{SO(8)/G_2},$$

$$T^R = T^R_{G_2} \oplus T^R_{SO(8)/G_2},$$

$$T^P = T^P_{G_2} \oplus T^P_{SO(8)/G_2},$$

we have

$$T^P_{G_2}(xy) + T^P_{SO(8)/G_2}(xy) = (T^L_{G_2}x)y + (T^L_{SO(8)/G_2}x)y + x(T^R_{G_2}y) + x(T^R_{SO(8)/G_2}y).$$

Now since G_2 is the automorphism group of octonions, its Lie algebra will be the derivation algebra of octonions satisfying

$$D(xy) = (Dx)y + x(Dy), \qquad D \in \text{Lie algebra of } G_2 = \mathcal{L}G_2.$$

Hence it follows that

$$D = T^P_{G_2} = T^L_{G_2} = T^R_{G_2} \in \mathcal{L}G_2.$$

In other words under the triality mappings

$$\begin{tikzcd} L \arrow[rr] & & R \arrow[dl] \\ & P \arrow[ul] & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} L \arrow[dr] & & R \arrow[ll] \\ & P \arrow[ur] & \end{tikzcd}$$

$\mathcal{L}G_2$ subalgebra of $\mathcal{L}SO(8)$ remains fixed. If we let

$$T^L_{SO(8)/G_2} = \begin{pmatrix} U_{SO(8)/G_2} & V_{SO(8)/G_2} \\ V^\dagger_{SO(8)/G_2} & -U^*_{SO(8)/G_2} \end{pmatrix}, \tag{12.6a}$$

$$T^R_{SO(8)/G_2} = \begin{pmatrix} A_{SO(8)/G_2} & B_{SO(8)/G_2} \\ B^\dagger_{SO(8)/G_2} & -A^*_{SO(8)/G_2} \end{pmatrix}, \tag{12.6b}$$

$$T^P_{SO(8)/G_2} = \begin{pmatrix} C_{SO(8)/G_2} & D_{SO(8)/G_2} \\ D^\dagger_{SO(8)/G_2} & -C^*_{SO(8)/G_2} \end{pmatrix}, \tag{12.6c}$$

where

$$U_{SO(8)/G_2}=\begin{pmatrix} -n_3 & 0 & 0 & [\frac{1}{2}(n_6-z_6)-\frac{1}{2}i(n_7+z_7)] \\ 0 & -n_3 & 0 & [\frac{1}{2}(n_4-z_4)-\frac{1}{2}i(n_5+z_5)] \\ 0 & 0 & -n_3 & [\frac{1}{2}(n_1-z_1)-\frac{1}{2}i(n_2+z_2)] \\ [\frac{1}{2}(n_6-z_6)+\frac{1}{2}i(n_7-z_7)] & [\frac{1}{2}(n_4-z_4)+\frac{1}{2}i(n_5-z_5)] & [\frac{1}{2}(n_1-z_1)+\frac{1}{2}i(n_2-z_2)] & -z_3 \end{pmatrix}, \tag{12.7a}$$

$$V_{SO(8)/G_2}=$$

$$\begin{pmatrix} 0 & -(n_1+in_2) & (n_4+in_5) & -[\frac{1}{2}(n_6+z_6)-\frac{1}{2}i(n_7-z_7)] \\ (n_1+in_2) & 0 & -(n_6+in_7) & -[\frac{1}{2}(n_4+z_4)-\frac{1}{2}i(n_5-z_5)] \\ -(n_4+in_5) & (n_6+in_7) & 0 & -[\frac{1}{2}(n_1+z_1)-\frac{1}{2}i(n_2-z_2)] \\ [\frac{1}{2}(n_6+z_6)-\frac{1}{2}i(n_7-z_7)] & [\frac{1}{2}(n_4+z_4)-\frac{1}{2}i(n_5-z_5)] & [\frac{1}{2}(n_1+z_1)-\frac{1}{2}i(n_2-z_2)] & 0 \end{pmatrix}. \tag{12.7b}$$

Then we find, after some calculation,

$$A_{SO(8)/G_2}=\begin{pmatrix} \frac{1}{2}(n_3+z_3) & 0 & 0 & -(n_6-in_7) \\ 0 & \frac{1}{2}(n_3+z_3) & 0 & -(n_4-in_5) \\ 0 & 0 & \frac{1}{2}(n_3+z_3) & -(n_1-in_2) \\ -(n_6+in_7) & -(n_4+in_5) & -(n_1+in_2) & -\frac{1}{2}(3n_3-z_3) \end{pmatrix}, \tag{12.7c}$$

$$B_{SO(8)/G_2}=$$

$$\frac{1}{2}\begin{pmatrix} 0 & [(n_1+z_1)+i(n_2-z_2)] & -[(n_4+z_4)+i(n_5-z_5)] & -[(n_6-z_6)-i(n_7+z_7)] \\ -[(n_1+z_1)+i(n_2-z_2)] & 0 & [(n_6+z_6)+i(n_7-z_7)] & -[(n_4-z_4)-i(n_5+z_5)] \\ [(n_4+z_4)+i(n_5-z_5)] & -[(n_6+z_6)+i(n_7-z_7)] & 0 & -[(n_1-z_1)-i(n_2+z_2)] \\ [(n_6-z_6)-i(n_7+z_7)] & [(n_4-z_4)-i(n_5+z_5)] & [(n_1-z_1)-i(n_2+z_2)] & 0 \end{pmatrix}, \tag{12.7d}$$

and

$$C_{SO(8)/G_2}=\begin{pmatrix} +\frac{1}{2}(n_3-z_3) & 0 & 0 & -(n_6-in_7) \\ 0 & \frac{1}{2}(n_3-z_3) & 0 & -(n_4-in_5) \\ 0 & 0 & \frac{1}{2}(n_3-z_3) & -(n_1-in_2) \\ -(n_6+in_7) & -(n_4+in_5) & -(n_1+in_2) & -\frac{1}{2}(3n_3+z_3) \end{pmatrix}, \tag{12.7e}$$

$$D_{SO(8)/G_2}=$$

$$\frac{1}{2}\begin{pmatrix} 0 & [(n_1-z_1)+i(n_2+z_2)] & -[(n_4-z_4)+i(n_5+z_5)] & -[(n_6+z_6)-i(n_7-z_7)] \\ -[(n_1-z_1)+i(n_2+z_2)] & 0 & [(n_6-z_6)+i(n_7+z_7)] & -[(n_4+z_4)-i(n_5-z_5)] \\ [(n_4-z_4)+i(n_5+z_5)] & -[(n_6-z_6)+i(n_7+z_7)] & 0 & -[(n_1+z_1)-i(n_2-z_2)] \\ [(n_6+z_6)-i(n_7-z_7)] & [(n_4+z_4)-i(n_5-z_5)] & [(n_1+z_1)-i(n_2-z_2)] & 0 \end{pmatrix}. \tag{12.7f}$$

We had shown earlier that the action of $\mathcal{L}G_2 \cong \mathrm{Der}\mathbf{O}$ on the octonion units can be represented by octonion multiplication and the action on the split octonion basis

$$[s]=\begin{pmatrix} u \\ u^* \end{pmatrix}$$

is uniquely determined by the action on u. Similarly, the action of $SO(8)$ on split octonions can be represented by octonion multiplication and the action on u uniquely determines the action on $[s]$. Below we give the expressions for the action of $\mathcal{L}SO(8)/G_2$ matrices that are in triality with each other in terms of octonion multiplication acting on u:

$$u=\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_0 \end{pmatrix},$$

$$\begin{aligned} T^L_{SO(8)/G_2}u = {} & \tfrac{1}{2}n_1([u_3^*,u]+(u+uu_0^*)u_3) \\ & -\tfrac{1}{2}in_2([u_3^*,u]-(u+uu_0^*)u_3) \\ & -n_3(uu_0^*) \\ & +\tfrac{1}{2}n_4([u_2^*,u]+(u+uu_0^*)u_2) \\ & -\tfrac{1}{2}in_5([u_2^*,u]-(u+uu_0^*)u_2) \\ & +\tfrac{1}{2}n_6([u_1^*,u]+(u+uu_0^*)u_1) \\ & -\tfrac{1}{2}in_7([u_1^*,u]-(u+uu_0^*)u_1) \\ & +z_1\tfrac{1}{2}(\{u_3^*,u\}-(uu_0)u_3) \\ & -iz_2\tfrac{1}{2}(-\{u_3^*,u\}-(uu_0)u_3) \\ & -z_3(uu_0) \\ & +z_4\tfrac{1}{2}(\{u_2^*,u\}-(uu_0)u_2) \\ & -iz_5\tfrac{1}{2}(-\{u_2^*,u\}-(uu_0)u_2) \\ & +z_6\tfrac{1}{2}(\{u_1^*,u\}-(uu_0)u_1) \\ & -iz_7\tfrac{1}{2}(-\{u_1^*,u\}-(uu_0)u_1), \end{aligned} \tag{12.8}$$

$$\begin{aligned} T^R_{SO(8)/G_2}u = {} & n_1(uu_3^* + \tfrac{1}{2}u_3^*u - (u - \tfrac{1}{2}uu_0^*)u_3) \\ & - in_2(uu_3^* + \tfrac{1}{2}u_3^*u + (u - \tfrac{1}{2}uu_0^*)u_3) \\ & + \tfrac{1}{2}n_3(uu_0^*) - \tfrac{3}{2}n_3(uu_0) \\ & + n_4(uu_2^* + \tfrac{1}{2}u_2^*u - (u - \tfrac{1}{2}uu_0^*)u_2) \\ & - in_5(uu_2^* + \tfrac{1}{2}u_2^*u + (u - \tfrac{1}{2}uu_0^*)u_2) \\ & + n_6(uu_1^* + \tfrac{1}{2}u_1^*u - (u - \tfrac{1}{2}uu_0^*)u_1) \\ & - in_7(uu_1^* + \tfrac{1}{2}u_1^*u + (u - \tfrac{1}{2}uu_0^*)u_1) \\ & + z_1\tfrac{1}{2}(-(uu_0^*)u_3 - u_3^*u) \\ & - iz_2\tfrac{1}{2}(-(uu_0^*)u_3 + u_3^*u) \\ & + \tfrac{1}{2}z_3u \\ & + z_4\tfrac{1}{2}(-(uu_0^*)u_2 - u_2^*u) \\ & - iz_5\tfrac{1}{2}(-(uu_0^*)u_2 + u_2^*u) \\ & + z_6\tfrac{1}{2}(-(uu_0^*)u_1 - u_1^*u) \\ & - iz_7\tfrac{1}{2}(-(uu_0^*)u_1 + u_1^*u), \end{aligned} \tag{12.9}$$

$$\begin{aligned} T^P_{SO(8)/G_2}u = {} & n_1(uu_3^* + \tfrac{1}{2}u_3^*u - (u - \tfrac{1}{2}uu_0^*)u_3) \\ & - in_2(uu_3^* + \tfrac{1}{2}u_3^*u + (u - \tfrac{1}{2}uu_0^*)u_3) \\ & + n_3(\tfrac{1}{2}u - 2uu_0) \\ & + n_4(uu_2^* + \tfrac{1}{2}u_2^*u - (u - \tfrac{1}{2}uu_0^*)u_2) \\ & - in_5(uu_2^* + \tfrac{1}{2}u_2^*u + (u - \tfrac{1}{2}uu_0^*)u_2) \\ & + n_6(uu_1^* + \tfrac{1}{2}u_1^*u - (u - \tfrac{1}{2}uu_0^*)u_1) \\ & - in_7(uu_1^* + \tfrac{1}{2}u_1^*u + (u - \tfrac{1}{2}uu_0^*)u_1) \\ & + z_1\tfrac{1}{2}((uu_0^*)u_3 + u_3^*u) \\ & - iz_2\tfrac{1}{2}((uu_0^*)u_3 - u_3^*u) \\ & - \tfrac{1}{2}z_3u \\ & + z_4\tfrac{1}{2}((uu_0^*)u_2 + u_2^*u) \\ & - iz_5\tfrac{1}{2}((uu_0^*)u_2 - u_2^*u) \\ & + z_6\tfrac{1}{2}((uu_0^*)u_1 + u_1^*u) \\ & - iz_7\tfrac{1}{2}((uu_0^*)u_1 - u_1^*u) \end{aligned} \tag{12.10}$$

ACKNOWLEDGMENTS

The authors gratefully acknowledge useful and stimulating conversations with Professor N. Jacobson, Professor G. Seligman, and Dr. S. Orfanidis.

APPENDIX A: STRUCTURE CONSTANTS OF G_2

Consider the basis of $\mathcal{L}G_2$ given in Sec. 1

$$\mathcal{L}G_2 = F_A \oplus M_A, \quad A = 1, \ldots, 7.$$

As was pointed out in Sec. 2, the generators F_A and $F_8 = -M_3$ form the $SU(3)$ subalgebra of $\mathcal{L}G_2$, i.e.,

$$[F_a, F_b] = 2if_{abc}F_c, \quad a, b, c = 1, 2, \ldots, 8, \tag{A1}$$

where f_{abc} are the totally antisymmetric structure constants of Gell-Mann, the nonzero elements of which are given in Table A1.

Now

$$\mathcal{L}G_2 = F_a \oplus M_s, \quad a = 1, 2, \ldots, 8,$$
$$m_s = s = 1, 2, 4, 5, 6, 7,$$

$$F_a \cong \mathcal{L}SU(3).$$

TABLE A1.

abc	f_{abc}
123	1
147	$\frac{1}{2}$
156	$-\frac{1}{2}$
246	$\frac{1}{2}$
257	$\frac{1}{2}$
345	$\frac{1}{2}$
367	$-\frac{1}{2}$
458	$\sqrt{3}/2$
678	$\sqrt{3}/2$

TABLE A2.

a	$m_s m_t$	$C_{am_sm_t}$
1	m_4m_7	1/2
1	m_5m_6	1/2
2	m_4m_6	$-1/2$
2	m_5m_7	1/2
3	m_4m_5	1/2
3	m_6m_7	1/2
4	m_1m_7	1/2
4	m_2m_6	$-1/2$
5	m_1m_6	$-1/2$
5	m_2m_7	$-1/2$
6	m_1m_5	$-1/2$
6	m_2m_4	$-1/2$
7	m_1m_4	$-1/2$
7	m_2m_5	1/2
8	m_4m_5	$-1/2\sqrt{3}$
8	m_6m_7	$1/2\sqrt{3}$
8	m_1m_2	-1

TABLE A3.

$m_sm_tm_u$	$C_{m_sm_tm_u}$
$m_1m_4m_7$	$-1/\sqrt{3}$
$m_1m_5m_6$	$1/\sqrt{3}$
$m_2m_4m_6$	$-1/\sqrt{3}$
$m_2m_5m_7$	$-1/\sqrt{3}$

The structure constants of the form C_{abm_s} vanish because F_a form a subalgebra. Hence the remaining nonvanishing structure constants of G_2 are of the form

$$C_{am_sm_t}, \quad a = 1, \ldots, 8, \quad s, t, u = 1, 2, 4, 5, 6, 7$$

or of the form

$$C_{m_sm_tm_u},$$

$$\begin{aligned} [F_a, F_b] &= 2if_{abc}F_c, \\ [F_a, M_s] &= 2iC_{am_sm_t}M_t, \\ [M_s, M_t] &= 2i(C_{m_sm_ta}F_a + C_{m_sm_tm_u}M_u, \end{aligned} \tag{A2}$$

where all the structure constants are totally antisymmetric. Below we list all the nonvanishing elements of $C_{am_sm_t}$ and $C_{m_sm_tm_u}$ (Tables A2 and A3).

APPENDIX B: ZORN'S VECTOR MATRICES

A realization of the split octonion algebra is via the Zorn's vector matrices

$$\begin{pmatrix} a & \mathbf{x} \\ \mathbf{y} & b \end{pmatrix},$$

where a and b are scalars and $\mathbf{x}$ and $\mathbf{y}$ are 3-vectors, with the product defined as

$$\begin{pmatrix} a & \mathbf{x} \\ \mathbf{y} & b \end{pmatrix}\begin{pmatrix} c & \mathbf{u} \\ \mathbf{v} & d \end{pmatrix} = \begin{pmatrix} ac + \mathbf{x}\cdot\mathbf{v} & a\mathbf{u} + d\mathbf{x} - \mathbf{y}\times\mathbf{v} \\ c\mathbf{y} + b\mathbf{v} + \mathbf{x}\times\mathbf{u} & \mathbf{y}\cdot\mathbf{u} + bd \end{pmatrix}. \tag{B1}$$

$\times$ denotes the usual vector product.

If the basis vectors of the three-dimensional space are $\mathbf{e}_i$, $i = 1, 2, 3$ with $\mathbf{e}_i \times \mathbf{e}_j = \epsilon_{ijk}\mathbf{e}_k$ and $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, then we can relate the split octonions to the vector matrices; namely

$$\begin{pmatrix}1&0\\0&0\end{pmatrix}=u_0^*,\quad \begin{pmatrix}0&-\mathbf{e}_i\\0&0\end{pmatrix}=u_i^*,$$

$$\begin{pmatrix}0&0\\0&1\end{pmatrix}=u_0,\quad \begin{pmatrix}0&0\\\mathbf{e}_i&0\end{pmatrix}=u_i.$$

Octonion conjugation defined above induces a natural involution for the vector matrices, i.e., if

$$A=\begin{pmatrix}a&-\mathbf{x}\\\mathbf{y}&b\end{pmatrix},\quad \bar{A}=\begin{pmatrix}b&+\mathbf{x}\\-\mathbf{y}&a\end{pmatrix},$$

$$A=au_0^*+x_iu_i^*+bu_0+y_iu_i,$$

$$\bar{A}=au_0-x_iu_i^*+bu_0^*-y_iu_i,$$

$$N(A)=A\bar{A}=\bar{A}A=(ab+\mathbf{x}\cdot\mathbf{y}). \tag{B2}$$

APPENDIX C: PRINCIPLE OF TRIALITY

The usual octonionic norm is invariant under the group $SO(8)$ or equivalently the bilinear form induced by the octonionic norm is skew with respect to the Lie algebra of $SO(8)$, i.e.,

$$(x,y)\equiv\tfrac{1}{2}(\bar{x}y+y\bar{x}) \tag{C1}$$

then for $T\in\mathcal{L}SO(8)$

$$(Tx,y)+(x,Ty)=0\quad \text{for all } x,y\in\mathbf{O} \tag{C2}$$

For the elements D of the derivation algebra of octonions we have

$$D\in \operatorname{Der}\mathbf{O}\cong\mathcal{L}G_2,\qquad D(xy)=(Dx)y+x(Dy). \tag{C3}$$

Integrated form of this (local) identity gives us the automorphisms of $\mathbf{O}$, i.e.,

$$e^D(xy)=(e^Dx)(e^Dy) \tag{C4}$$

or

$$d=e^D,$$

$$d(xy)=(dx)(dy)\Rightarrow d\in G_2.$$

The principle of triality is nothing but a generalization of the identities (C3)–(4) and is unique to octonions.[8] According to the principle of local triality (PLT) it is possible to generalize identity (C3) to all the elements of the Lie multiplication algebra $\mathcal{L}SO(8)$. Namely, given an element $T^L\in\mathcal{L}SO(8)$ acting on the octonions there exist *unique* T^R and $T^P\in\mathcal{L}SO(8)$ such that

$$(\mathrm{PLT}):(T^Lx)y+x(T^Ry)=T^P(xy)\quad \text{for all } x,y\in\mathbf{O}. \tag{C5}$$

Just as it is possible to integrate the derivations of octonion algebra to get its automorphisms, one can also integrate the PLT to get the principle of global triality (PGT), which is a generalization of the concept of automorphism. According to the PGT, given $t^l\in SO(8)$ acting on the octonions there exist t^r and $t^p\in SO(8)$, unique up to a sign, such that[33]

$$\mathrm{PGT}:(t^lx)(t^ry)=t^p(xy)\quad \text{for all } x,y\in\mathbf{O}. \tag{C6}$$

Since the group $SO(8)$ is the "Lie multiplication group" of octonions (i.e., that every action of $SO(8)$ on O can be represented by octonion multiplication), one can reformulate the PGT as follows[34]:

Given $d^1\in SO(8)$ $d^2,d^3\in SO(8)$

$$(d^1x)(d^2y)=\overline{d^3\overline{(xy)}}\qquad \text{for all } x,y\in\mathbf{O}, \tag{C7a}$$

where the overbar denotes octonion conjugation.

In this form of the PGT we have cyclic symmetry between d^1, d^2 and d^3, i.e.,

$$(d^1x)(d^2y)=\overline{d^3\overline{(xy)}}$$

implies

$$(d^2x)(d^3y)=\overline{d^1\overline{(xy)}}, \tag{C7b}$$

$$(d^3x)(d^1y)=\overline{d^2\overline{(xy)}}. \tag{C7c}$$

Since given d^1, d^2, and d^3 are determined uniquely up to a sign, the subgroup of $SO(8)\times SO(8)\times SO(8)$ consisting of elements which are in triality will form a twofold covering group of $SO(8)$, i.e., it will be isomorphic to Spin (8). The group $SO(8)$ has the subgroup $SO(7)$ and given $t\in SO(7)$ there exist $\bar{t}\in SO(8)$

$$(tx)(\bar{t}y)=\bar{t}(xy)\quad \text{for all } x,y\in\mathbf{O} \tag{C8}$$

then the elements t form the covering group Spin (7) of $SO(7)$.

APPENDIX D: REALIZATIONS OF THE CAYLEY ALGEBRA IN TERMS OF GELL-MANN λ MATRICES AND DIRAC'S γ-MATRICES

1. The λ-matrices

We want to define a product between the λ matrices of Gell-Mann such that they will form the nonassociative Cayley algebra. Since there are eight λ matrices and seven imaginary octonion units e_A, the product will be defined between seven of the λ matrices and will involve the eighth λ matrix. In view of the broken $SU(3)$, this eighth λ matrix will be taken to be λ_8. The general form of the product consistent with octonion multiplication can be parameterized as follows:

$$\begin{aligned}\lambda_A\circ\lambda_B=&\tfrac{1}{2}\beta\operatorname{Tr}(\lambda_A\lambda_B)1+\tfrac{1}{2}\delta\operatorname{Tr}(\lambda_8\{\lambda_A,\lambda_B\})1\\&-(2/\sqrt{3})(\alpha+\tfrac{1}{6}\gamma)\operatorname{Tr}(\lambda_8[\lambda_A,\lambda_B])1\\&+\{\alpha 1+\sqrt{3}(\alpha+\tfrac{1}{6}\gamma)\lambda_8,[\lambda_A,\lambda_B]\}\\&+\gamma[\{\lambda_8,\lambda_A\},\{\lambda_8,\lambda_B\}],\end{aligned} \tag{D1}$$

where { , } and [,] denote anticommutation and commutation, respectively. Then, for $A=1,2,3$ we have

$$\lambda_A\circ\lambda_A=\beta+(2/\sqrt{3})\delta\equiv 1/s^2,\qquad \text{no sum over } A, \tag{D2}$$

and for $A=4,5,6,7$

$$\lambda_A\circ\lambda_A=\beta-(1/\sqrt{3})\delta\equiv 1/t^2. \tag{D3}$$

In addition, the octonion multiplication imposes the conditions:

$$\alpha=-\tfrac{5}{4}(2\gamma/9),\qquad \beta=15(2\gamma/9)^2,$$

$$\delta=5\sqrt{3}(2\gamma/9)^2. \tag{D4}$$

Hence, we get the result that the 3×3 matrices

$$e_i = is\lambda_i, \quad i = 1, 2, 3,$$
$$e_4 = it\lambda_4, \quad e_6 = -it\lambda_6,$$
$$e_5 = it\lambda_5, \quad e_7 = -it\lambda_7 \tag{D5}$$

satisfy the octonion multiplication table of the imaginary units e_A under the product defined above and generate a Cayley algebra with identity being the scalar identity:

$$e_A \circ e_B = -\delta_{AB} + a_{ABC}e_C \tag{D6}$$

An interesting property of this product is that the coefficient multiplying the λ matrices is different for different isospin multiplets.

2. The γ-matrices

Let us define a product between 4×4 Hermitian matrices of the form:

$$A = \begin{pmatrix} \alpha 1_2 & -i\sigma\cdot\mathbf{a} \\ i\sigma\cdot\mathbf{b} & \beta 1_2 \end{pmatrix}, \quad C = \begin{pmatrix} \gamma 1_2 & -i\sigma\cdot\mathbf{c} \\ i\sigma\cdot\mathbf{d} & \delta 1_2 \end{pmatrix}.$$

Such that they form a Cayley algebra. First, note that the matrix A can be written in terms of γ matrices as:

$$A = \tfrac{1}{2}(\alpha+\beta) + \tfrac{1}{2}(\alpha-\beta)\gamma_5 + \tfrac{1}{2}\gamma_5\boldsymbol{\gamma}\cdot(\mathbf{a}-\mathbf{b}) + \tfrac{1}{2}\boldsymbol{\gamma}\cdot(\mathbf{a}+\mathbf{b}), \tag{D7}$$

where γ matrices are taken in the Weyl basis and the parameters $\alpha, \beta, \mathbf{a}, \mathbf{b}$ are all real.

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3) = \rho_2 \otimes \sigma,$$
$$\gamma_4 = \rho_1 \otimes I, \quad \gamma_5 = \rho_3 \otimes I. \tag{D8}$$

To get a product which is not associative, we are led to defining a new operation $\tilde{}$ over the 4×4 matrices:

$$\tilde{M} = \widetilde{\begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix}} \equiv \begin{pmatrix} \mathbf{a}^\dagger & \mathbf{b} \\ \mathbf{c} & \mathbf{d}^\dagger \end{pmatrix}, \tag{D9}$$
$$= \tfrac{1}{2}(1+\gamma_5)A^\dagger\tfrac{1}{2}(1+\gamma_5) + \tfrac{1}{2}(1-\gamma_5)A^\dagger\tfrac{1}{2}(1-\gamma_5)$$
$$+ \tfrac{1}{2}(1+\gamma_5)A\tfrac{1}{2}(1-\gamma_5) + \tfrac{1}{2}(1-\gamma_5)A\tfrac{1}{2}(1+\gamma_5),$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ are 2×2 matrices.

Then under the product

$$A * C = \tfrac{1}{2}(AC + \widetilde{AC}) + \tfrac{1}{2}\gamma_4(AC^\dagger - \widetilde{AC^\dagger}) \tag{D10}$$

the matrices of the form shown above form a split Cayley algebra equivalent to the Zorn's vector matrices

$$A*C = \begin{pmatrix} (\alpha\gamma + \mathbf{a}\cdot\mathbf{d}) & (-i\alpha\sigma\cdot\mathbf{c} - i\delta\sigma\cdot\mathbf{a} + i\sigma\cdot\mathbf{b}\times\mathbf{d}) \\ (i\gamma\sigma\cdot\mathbf{b} + i\beta\sigma\cdot\mathbf{d} + i\sigma\cdot(\mathbf{a}\times\mathbf{c}) & (\beta\delta + \mathbf{b}\cdot\mathbf{c}) \end{pmatrix}. \tag{D11}$$

Writing A in the form

$$A = \tfrac{1}{2}(1+\gamma_5)(\alpha + \boldsymbol{\gamma}\cdot\mathbf{a}) + \tfrac{1}{2}(1-\gamma_5)(\beta + \boldsymbol{\gamma}\cdot\mathbf{b}),$$
$$= \tfrac{1}{2}(1 - ie_7)(\alpha + e_i a_i) + \tfrac{1}{2}(1 + ie_7)(\beta + e_i b_i),$$
$$= \alpha u_0^* + u_i^* a_i + \beta u_0 + u_i b_i, \tag{D12}$$

it is easily seen that the split octonion basis u_i, u_0, u_i^*, u_0^* is realized in this case by

$$u_0^* = \tfrac{1}{2}(1+\gamma_5), \quad u_0 = \tfrac{1}{2}(1-\gamma_5),$$
$$u_i^* = \tfrac{1}{2}(1+\gamma_5)\gamma_i, \quad u_i = \tfrac{1}{2}(1-\gamma_5)\gamma_i, \; i = 1, 2, 3. \tag{D13}$$

Therefore, the role played by ie in extending the quaternion algebra $(1, e_1, e_2, e_3)$ into the split octonion algebra is played in the above realization by γ_5, i.e.,

$$ie_7(1, e_1, e_2, e_3) = (ie_7, ie_4, ie_5, ie_6),$$
$$\gamma_5 * (1, \gamma_1, \gamma_2, \gamma_3) = \gamma_5(1, \gamma_1, \gamma_2, \gamma_3)$$
$$= (\gamma_5, \gamma_5\gamma_1, \gamma_5\gamma_2, \gamma_5\gamma_3).$$

$*$ multiplication by γ_5 reduces to the ordinary matrix multiplication. Conversely, the crucial role played by γ_5 in constructing projection operators into lh and rh states is reflected in the octonion algebra by the important role played by u_0 and u_0^* as projection operators into quark and antiquark states in the octonionic representations of the Poincaré group.[14]

*Research (Yale Report 3075-27) supported by the U. S. Atomic Energy Commission under contract AT(11-1)3075.
†Now at the Middle East Technical University, Ankara, on leave from Yale University.
[1]P. Jordan, Z. Phys. **80**, 285 (1933).
[2]P. Jordan, J. Von Neumann, and E. P. Wigner, Ann. Math. **35**, 29 (1934); A. A. Albert, Ann. Math. 35, 65 (1934).
[3]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U. P., Oxford, 1958).
[4]M. Gell-Mann, Phys. Rev. **92**, 833 (1953). K. Nishijima, Prog. Theor. Phys. **13**, 285 (1955).
[5]M. Gell-Mann and Y. Neeman, *The Eightfold Way* (Benjamin, New York, 1965).
[6]R. E. Behrends and A. Sirlin, Phys. Rev. **121**, 324 (1961).
[7]R. D. Schafer, *Non-Associative Algebras* (Academic, New York, 1966).
[8]C. Chevalley and R. D. Schafer, Proc. Natl. Acad. Sci. USA **36**, 137 (1950).
[9]A. Gamba, in *High Energy Physics and Elementary Particles*, edited by A. Salam (IAEA, Vienna, 1965), p. 641. A. Gamba, J. Math. Phys. **8**, 775 (1967).
[10]A. Pais, Phys. Rev. Lett. **7**, 291 (1961); A. Pais, J. Math. Phys. **3**, 1135 (1962).
[11]R. Penney, Nuovo Cimento B 3, 95 (1971).
[12]S. Sherman, Ann. Math. **64**, 593 (1956).
[13]I. E. Segal, Ann. Math. **48**, 930 (1947).
[14]M. Günaydin and F. Gürsey, Nuovo Cimento Lett. **6**, 401 (1973).
[15]H. Georgi and S. Glashow, Phys. Rev. D **6**, 429 (1972).
[16]See, for example, A. A. Albert, *Studies in Modern Algebra* (Prentice-Hall, Princeton, NJ, 1963). N. Jacobson, "Structure and Representations of Jordan Algebras," Amer. Math. Soc. Coll. Publ. Vol. 39, and references contained therein.
[17]G. B. Seligman, Trans. Am. Math. Soc. **94**, 452 (1960) and Trans. Am. Math. Soc. **97**, 286 (1960).
[18]See Ref. 14 and H. H. Goldstine and L. P. Horwitz, Proc. Natl. Acad. Sci. USA **48**, 1134 (1962); Math. Ann. **154**, 1 (1964); Math. Ann. **164**, 291 (1966); L. P. Horwitz and L. C. Biedenharn, Helv. Phys. Acta **38**, 385 (1965).
[19]Note the distinction between the terms nonassociative and not associative. The former is generally used to denote all the composition algebras mentioned above which satisfy the property of alternativity defined below.
[20]H. Freudenthal, "Oktaven, Ausnahmegruppen and Oktavengeometrie," (mimeographed), Utretch (1951).

[21]See, for example: D. Finkelstein, J. M. Jauch, S. Schiminovich, and D. Speiser, J. Math. Phys. **3**, 207 (1962). D. Finkelstein, J. M. Jauch, and D. Speiser, J. Math. Phys. **4**, 136 (1963).

[22]M. Zorn, Proc. Natl. Acad. Sci. USA **21**, 355 (1935).

[23]By three "independent" elements we mean any three elements e_i, e_j, e_k such that none of them is proportional to a product of the other two, i.e., $e_k \neq a e_i e_j$.

[24]An element left invariant by the Lie group is said to be annihilated by the Lie algebra.

[25]See, e. g., M. L. Mehta, J. Math. Phys. **7**, 1824 (1966) and M. L. Mehta and P. K. Srivastava, J. Math. Phys. **7**, 1833 (1966).

[26]For example, take $\{1, e_1, e_2, e_3\}$ as a basis generating a quaternion algebra H. Then each octonion can be written as $z = q_1 + q_2 e_7$, $w = r_1 + r_2 e_7 \in O$, where $q_1, q_2, r_1, r_2 \in H$, with the product defined by $zw = (q_1 r_1 - \bar{r}_2 q_2) + (r_2 q_1 + q_2 \bar{r}_1) e_7$. (The bar denotes quaternion conjugation).

[27]An equivalent form of this basis was first studied by G. Seligman as the derivation algebra of Zorn's vector matrices given in Appendix B. See Ref. 17.

[28]In fact, the generator N_3 extends $SU(3)$ subgroup into $U(3)$ and the group G_2 into $SO(7)$.

[29]Note that we put a bar over the indexed matrices when they act on the split octonions, i.e., under the numbering: $(u_1 u_2 u_3 u_0 u_1^* u_2^* u_3^* u_0^*)$ $(s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8)$ we have $\bar{E}_{ab} s_c = \delta_{bc} s_a$, $a, b, c = 1, \ldots, 8$. For the real octonions we have the numbering $(e_A, 1) \leftrightarrow (e_A, e_8)$, $A = 1, \ldots, 7$. Then $E_{ab} e_c = \delta_{bc} e_a$, $a, b, c, = 1, \ldots, 8$. We also defined Σ_{ab} as $\Sigma_{ab} = E_{ab} - E_{ba}$, $\bar{\Sigma}_{ab} = \bar{E}_{ab} - \bar{E}_{ba}$.

[30]I. Yokota, J. Fac. Sci. Shinshu Univ. 2, 125 (1967).

[31]O. Hara, Y. Fujii, and Y. Ohnuki, Prog. Theor. Phys. **19**, 129 (1958); B. Touschek, Nuovo Cimento **18**, 181 (1958); A. Gamba and E. C. G. Sudarshan, Nuovo Cimento **10**, 407 (1958); T. D. Lee and G. C. Wick, Phys. Rev. **148**, B1385 (1966); F. Gürsey and M. Koca, Nuovo Cimento Lett. **1**, 228 (1969).

[32]For a proof and more details see N. Jacobson *Lie Algebras* (Interscience, New York, 1962.)

[33]Here we use capital letters for the elements of the Lie algebra and small letters for the group elements.

[34]See Y. Matşushima, Nagoya Math. J. **4**, 83 (1952).

Projective manifolds and projective theory of relativity*

G. T. Evans† and D. K. Sen

Department of Mathematics, University of Toronto, Toronto M5S 1A1, Canada

(Received 27 March 1973; revised manuscript received 5 June 1973)

We present a *global* formulation of projective theories of relativity in the framework of projective manifolds, that is, manifolds based on the pseudogroup of homogeneous transformations in $\mathbb{R}^5$. Apart from formulating every previously considered geometric object and physical relation in an invariant manner, some new results, such as the theorem on the semidirect product structure of the invariance group of Einstein–Maxwell equations, and theorems on topological restrictions on the underlying five-dimensional projective manifold, etc. have been obtained. The relationship between space–time and the auxiliary 5-manifold is clarified and investigated in detail. A more general geometric definition of the electromagnetic field tensor and a geometric interpretation of the charge/mass ratio is given.

INTRODUCTION

Projective theory of relativity originated as one of the various attempts to formulate a unified and geometrized theory of gravitation and electromagnetism. Following the pioneering work of Weyl,[1] Kaluza[2] and Klein[3] introduced a five-dimensional manifold whose fifteen component metric tensor, under some rather artificial assumptions, could be interpreted as the combined field tensor satisfying the Einstein–Maxwell equations. The physical significance of the fifth dimension remained, however, unclear until Veblen and Hoffmann,[4,5] Schouten and van Dantzig[6] showed that the Kaluza–Klein theory could be regarded as a four-dimensional projective theory, in which a four-dimensional projective space was attached to every point of space–time. A somewhat different approach was taken by van Dantzig[7] who considered homogeneous coordinates in a five-dimensional space and introduced geometric objects whose components were homogeneous functions of coordinates. Jordan[8] demonstrated the homomorphism of the group of homogeneous coordinate transformations in $\mathbb{R}^5$ with the invariance group of Einstein–Maxwell equations and gave a generalized version of the theory[9,10,11].

All these works were however formulated in local coordinates. The object of this paper is to provide a *global* formulation of projective theory of relativity in the framework of so-called projective manifolds, that is, manifolds based on the pseudogroup of homogeneous transformations of degree one in $\mathbb{R}^5$. This provides a global formulation of van Dantzig and Jordan's version of projective theory of relativity.

In the first part, the invariance pseudogroup of Einstein–Maxwell equations is investigated in some detail and its relationship with the pseudogroup of homogeneous transformations of degree one in $\mathbb{R}^5$ is established. This leads naturally to projective manifolds. All geometric objects are defined globally and topological restrictions for the existence of such manifolds are pointed out.

In the second part, the projective theory of relativity is formulated and projections onto space-time of various geometrical objects are studied, leading finally to the basic field equations of the theory. A more general geometric definition of the electro-magnetic field tensor and a geometric interpretation of the charge/mass ratio is given.

1. C^∞-PROJECTIVE MANIFOLDS

The pseudogroup of C^∞-homogeneous transformations

Consider the Einstein–Maxwell equations for the combined electromagnetic and gravitational fields in vacuum

$$G_{ik} + \kappa E_{ik} = 0,$$
$$f^{ik}_{;k} = 0, \quad f_{ik} = \phi_{i,k} - \phi_{k,i}. \tag{1.1}$$

[Here commas and semicolons denote partial and covariant derivatives, respectively. The summation convention is used throughout and we use $+ - - -$ for the signature of g_{ik} $(i, k = 0, 1, 2, 3)$.]

where ϕ_i is the electromagnetic four-potential, $E_{ik} = f_{il}f^l_k - \frac{1}{4}g_{ik}f_{lm}f^{lm}$ the energy–momentum tensor of the electromagnetic field f_{ik}, and G_{ik} the Einstein tensor for space–time with metric g_{ik}. Equation (1.1) is, of course, invariant under the pseudogroup K of C^∞ transformations in $\mathbb{R}^4$

$$K \ni \Lambda : x^k \to x^{k'} = x^{k'}(x^k). \tag{1.2}$$

K is strictly speaking a pseudogroup because coordinate transformations in a C^∞-manifold, i.e. C^∞-diffeomorphisms between open sets in $\mathbb{R}^n$ do not quite satisfy all the properties of a group. For a precise definition of pseudogroup of transformations, see Ref. 12.

But (1.1) is also invariant under the (Abelian) group E of gauge transformations

$$E \ni [\phi] : \phi_i \to \phi_i + \phi_{,i}, \tag{1.3}$$

where ϕ is a scalar function in $\mathbb{R}^4$. The symmetry group of the Einstein–Maxwell equations is therefore the combined pseudogroup G, a typical element of which will be denoted by $g = (\Lambda, [\phi])$ to mean a gauge transformation $[\phi]$ followed by a coordinate transformation Λ, e.g.,

$$(\Lambda, [\phi]) : \phi_i(x^k) \xrightarrow{[\phi]} \phi_i(x^k) + \phi(x^k)_{,i} \xrightarrow{\Lambda} [\phi_i(x^k) + \phi(x^k)_{,i}]x^i_{,i'}. \tag{1.4}$$

We have the following structure theorem for G.

Theorem 1.1: G is a semidirect product of K and E.

Proof: This follows from the group product rule in G. Note that the unit element of E is $[c]$, where c is any constant. Denote by e the unit element of K. Now consider a gauge transformation $[\phi_1]$ followed by a coordinate transformation Λ_1 and then $[\phi_2]$ and Λ_2—all in that order. Then we have

$$\phi_i(x^k) \xrightarrow{[\phi_1]} \phi_i(x^k) + \phi_1(x^k)_{,i} \xrightarrow{\Lambda_1} (\phi_i(x^k) + \phi_1(x^k)_{,i})x^i_{,i'}$$
$$\xrightarrow{[\phi_2]} \phi_i(x^k)x^i_{,i'} + \phi_1(x^k)_{,i}x^i_{,i'} + \phi_2(x^{k'})_{,i'} \xrightarrow{\Lambda_2} \phi_i(x^k)x^i_{,i'}x^{i'}_{,i''}$$
$$+ \phi_1(x^k)_{,i}x^i_{,i'}x^{i'}_{,i''} + \phi_2(x^{k'})_{,i'}x^{i'}_{,i''}$$
$$= (\phi_i(x^k) + \phi_1(x^k)_{,i} + \phi_2(x^{k'}(x^k))_{,i})x^i_{,i''} \tag{1.5}$$

or

$$(\Lambda_2, [\phi_2]) \cdot (\Lambda_1, [\phi_1]) = (\Lambda_2\Lambda_1, [\phi_2 \circ \Lambda_1 + \phi_1]), \tag{1.6}$$

where $\phi_2 \circ \Lambda_1 : x \xrightarrow{\Lambda_1} x' \xrightarrow{\phi_2} \phi_2(x'(x))$.

Copyright © 1973 by the American Institute of Physics

The inverse rule is therefore

$$(\Lambda,[\phi])^{-1}=(\Lambda^{-1},[-\phi\circ\Lambda^{-1}]). \tag{1.7}$$

We see that $G = KE$, i.e., $(\Lambda, [\phi])$ can be written uniquely as $(\Lambda, [\phi]) = (\Lambda, [c])\cdot(e, [\phi])$.

K induces an automorphism of E as follows:

$$\begin{aligned}(\Lambda,[c]):(e,[\phi])&\to(\Lambda,[c])(e,[\phi])(\Lambda,[c])^{-1}\\&=(\Lambda,[\phi])(\Lambda^{-1},[-c\circ\Lambda^{-1}])\\&=(e,[\phi\circ\Lambda^{-1}])\in E.\end{aligned}$$

And E is a normal subpseudogroup of G, because

$$\begin{aligned}(\Lambda,[\phi])^{-1}(e,[\psi])(\Lambda,[\phi])&=(\Lambda^{-1},[-\phi\circ\Lambda^{-1}])(e,[\psi])(\Lambda,[\phi])\\&=(\Lambda^{-1},[-\phi\circ\Lambda^{-1}])(\Lambda,[\psi\circ\Lambda+\phi])=(e,[\psi\circ\Lambda])\in E.\end{aligned}$$

QED

Jordan[8] showed that the symmetry group of (1.1) is intimately connected with homogeneous transformations of degree one in $\mathbb{R}^5$. This is made more precise as follows.

Definition: A mapping $f: U\to\mathbb{R}$ is said to be a locally homogeneous function of degree μ on an open set $U\subset\mathbb{R}^n$, if for all $x\in U$ and all $t\in(1-\epsilon_x, 1+\epsilon_x)$, for some $1>\epsilon_x>0$ such that $tx\in U$, we have $f(tx)=t^\mu f(x)$.

We shall recall here some elementary facts about such functions. Denote by $C_\mu^\infty(U)$, the class of C^∞ locally homogeneous functions of degree μ on U (we consider only the C^∞ case, although the C^k case could be treated without much change):

$$\begin{aligned}&\text{(i)}\quad f\in C_\mu^\infty(U)\Leftrightarrow x^if_{,i}=\mu f\quad\text{(Euler's equation)},\\&\text{(ii)}\quad f\in C_\mu^\infty(U)\Rightarrow f_{,i}\in C_{\mu-1}^\infty(U),\\&\text{(iii)}\quad f\in C_\mu^\infty(U),\quad g\in C_\lambda^\infty(U)\Rightarrow f\cdot g\in C_{\mu+\lambda}^\infty(U).\end{aligned}\tag{1.8}$$

The notion of C^∞ locally homogeneous functions of degree μ can now be generalized to maps $f: U\to V$, where U and V are open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. That is, each coordinate function is to be locally homogeneous.

Proposition 1.1: Let $f: U\to V$ be a smooth diffeomorphism (i.e., both f and f^{-1} are C^∞) *in $\mathbb{R}^n$ of open sets U onto V. Then $f^{-1}\in C_{\mu^{-1}}^\infty(V)$ if $f\in C_\mu^\infty(U)$ and $\mu\neq 0$.*

Proof: Let $y=[y^1(x),\ldots,y^n(x)]=f(x)$ and $x=[x^1(y),\ldots,x^n(y)]=f^{-1}(y)$. Then $x^ly^j_{,l}=\mu y^j$, and $y^jx^i_{,j}=\mu^{-1}x^ly^j_{,l}x^i_{,j}=\mu^{-1}x^i$. QED

Proposition 1.2: Let $f: U\to V$ and $g: V\to W$, where U, V, W are open sets in $\mathbb{R}^p, \mathbb{R}^q$, and $\mathbb{R}^s$, respectively. If $f\in C_\mu^\infty(U)$ and $g\in C_\lambda^\infty(V)$, then $g\circ f\in C_{\lambda\mu}^\infty(U)$.

Proof: Let

$$y=[y^1(x),\ldots,y^q(x)]=f(x),$$

$$z=[z^1(y),\ldots,z^s(x)]=g(y).$$

Then

$$x^iy^\alpha_{,i}=\mu y^\alpha\quad(i=1,\ldots,p,\quad\alpha=1,\ldots,q),$$

$$y^\alpha z^A_{,\alpha}=\lambda z^A\quad(A=1,\ldots,s).$$

So

$$x^iz^A_{,i}=x^iz^A_{,\alpha}y^\alpha_{,i}=\mu y^\alpha z^A_{,\alpha}=\mu\lambda z^A.$$

QED

In particular if $\lambda=1$, $g\circ f\in C_\mu^\infty(U)$. Note that, *if $\mu=0$, $g\circ f$ is locally homogeneous of degree zero irrespective of whether g is homogeneous or not.*

As a Corollary of Proposition 1.2, we have,

Proposition 1.3: Smooth diffeomorphisms, which are locally homogeneous of degree one, form a proper subpseudogroup H^n of the pseudogroup of all smooth diffeomorphisms in $\mathbb{R}^n$.

We can now state the homomorphism theorem due to Jordan.

Theorem 1.2: The pseudogroup H^5 is homomorphic to the symmetry group G of the Einstein–Maxwell equations in vacuum.

Proof: An element $h\in H^5$ can be written as

$$h: X^\mu\to X^{\mu'}=X^{\mu'}(X^\mu),\qquad\mu=0,1,2,3,4, \tag{1.9}$$

where $X^{\mu'}(X^\mu)$ are C^∞, invertible, and locally homogeneous functions of degree one. It can also be written as

$$h: X^\mu\to X^{\mu'}=X^\mu f^{(\mu)}(X^\mu)\quad\text{(no summation)}, \tag{1.10}$$

where $f^{(\mu)}(X^\mu)$ are C^∞, invertible, and locally homogeneous functions of degree zero. H^5 has the following subpseudogroups:

$$J^5=\left\{h\in H^5\;\middle|\;h: X^\mu\to X^{\mu'}=X^\mu f(X^\mu)=X^\mu F\left(\frac{X^1}{X^0},\ldots,\frac{X^4}{X^0}\right)\right\}$$

$$N^5=\left\{h\in H^5\;\middle|\;h:\begin{array}{l}X^0\to X^{0'}=X^0\\ X^k\to X^{k'}=X^0f^k\left(\dfrac{X^1}{X^0},\ldots,\dfrac{X^4}{X^0}\right)\end{array}\right\}. \tag{1.11}$$

The theorem is proved by considering the homomorphisms

$$\begin{aligned}&J^5\to E,\qquad J^5\ni j\to[\phi]=[\log F],\\&N^5\simeq K,\qquad N^5\ni n\Leftrightarrow\Lambda: x^k\to x^{k'}=f^k(x^k),\end{aligned}\tag{1.12}$$

and by establishing that J^5 is a normal subpseudogroup of H^5. QED

For details of the proof we refer to Ref. 9. Actually G is *isomorphic* to a subpseudogroup of H^5. But it is H^5 that we shall be concerned with.

Projective manifolds

A C^k manifold is, by definition, a topological space with a maximal atlas compatible with the pseudogroup of C^k transformations in $\mathbb{R}^n$. We wish to consider manifolds based on the pseudogroup H^n.

Definition: An $(n+1)$-dimensional C^∞ projective manifold is a Hausdorff space with a maximal atlas compatible with the pseudogroup H^{n+1} of C^∞ homogeneous transformations of degree one in $\mathbb{R}^{n+1}$.

From now on M will denote an $(n+1)$-dimensional C^∞ projective manifold. We shall assume that M is also paracompact. Let $\{U_i,\phi_i\}$ be an atlas of M. Then for all pairs (i,j) such that $U_i\cap U_j\neq\Phi$ the coordinate maps $\phi_j\circ\phi_i^{-1}: X^\mu\to X^{\mu'}$ are locally homogeneous maps of degree one. Note that a maximal atlas compatible with H^{n+1} need not be a maximal atlas compatible with the pseudogroup of all C^∞ transformations in $\mathbb{R}^{n+1}$. So strictly speaking a C^∞ projective manifold is not a C^∞ manifold, although an atlas compatible with H^{n+1} is compatible with the pseudogroup of all C^∞ transformations in $\mathbb{R}^{n+1}$; in other words, a C^∞ projective structure defines uniquely a C^∞ structure. We shall consider the converse problem later.

Definition: A mapping $f: M \to N$, where M and N are C^∞-projective manifolds, is called a C^∞ homogeneous map of degree μ if for every point $x \in M$, there exists a local chart (U, ϕ) at x and a local chart (V, ψ) in $y = f(x)$ such that the map $\psi \circ f \circ \phi^{-1}$ is a locally C^∞-homogeneous map of degree μ from $\phi(f^{-1}(V) \cap U)$ into $\psi(V)$.

It follows from Proposition 1.2 that the homogeneity condition is independent of the choice of local charts since coordinate maps are homogeneous maps of degree one.

Note that the definition makes sense even if N is an ordinary C^∞-manifold, if $\mu = 0$.

If in the above we take $N = \mathbb{R}$, we obtain homogeneous functions $f: M \to \mathbb{R}$ of degree μ. In other words, $f: M \to \mathbb{R}$ is a C^∞ homogeneous function of degree μ if f is a C^∞-function on M and for every $x \in M$ there exists a local chart (U, ϕ) at x such that $f \circ \phi^{-1}$ is a locally homogeneous function of degree μ from $\phi(U)$ into $\mathbb{R}$. Again the homogeneity condition is independent of the choice of charts. For, if (U_i, ϕ_i) and (U_j, ϕ_j) are two charts with $U_i \cap U_j \neq \Phi$, we have $f \circ \phi_j^{-1} = (f \circ \phi_i^{-1}) \circ (\phi_i \circ \phi_j^{-1})$. Since $\phi_i \circ \phi_j^{-1}$ is locally homogeneous of degree one, it follows from Proposition 1.2 that $f \circ \phi_i^{-1}$ and $f \circ \phi_j^{-1}$ have both the same degree of homogeneity.

Proposition 1.4: Let $f: M \to N, g: N \to P$ be C^∞ homogeneous maps of degree μ and λ, respectively, where M, N and P are C^∞ projective manifolds. The composite map $g \circ f$ is a C^∞ homogeneous map of degree $\mu\lambda$.

Proof: Follows from Proposition 1.2. QED

Definition: A C^∞ homogeneous diffeomorphism $f: M \to N$ is a C^∞ diffeomorphism such that f and f^{-1} are homogeneous maps of degree one.

A C^∞-homogeneous diffeomorphism of course implies C^∞ diffeomorphism. The converse problem will be considered in Theorem 1.3.

Definition: A point $p \in M$ is an origin of M if there exists a chart (U, ϕ) at p such that $\phi(p) = (0, \ldots, 0)$. Let (V, ψ) be another chart at p with $\psi(p) = (X_0^{0'}, \ldots, X_0^{n'})$. Then $\psi \circ \phi^{-1}((0, \ldots, 0)) = (X_0^{0'}, \ldots, X_0^{n'})$ and for some $t \neq 0$, $\psi \circ \phi^{-1}(t(X^0, \ldots, X^n)) = (tX^{0'}, \ldots, tX^{n'})$. Therefore, $X_0^{0'} =, \ldots, = X_0^{n'} = 0$.

The definition of origins is therefore independent of the choice of charts.

Theorem 1.3: A maximal atlas of an $n+1$ dimensional C^∞-manifold contains a subatlas without origins compatible with H^{n+1} if and only if the manifold admits a C^∞ vector field X with no singularities. (We are indebted to S. Halperin for this theorem.)

Proof: Let F_t denote the flow of X. We can cover the manifold by a collection of charts (U, ϕ) with the following properties. There exists a hypersurface Σ of U which intersects the flow lines of X nontangentially such that $\phi(\Sigma)$ intersects the radial lines from the origin in $\mathbb{R}^{n+1}$ also nontangentially, and that for every point $p \in U$ there exists a unique point $p_0 \in \Sigma$ and a real number $\epsilon > 0$ such that $p = F_t p_0$ for some $-\epsilon < t < \epsilon$. If (U', ϕ') is another such a chart with $U \cap U' \neq \Phi$ then we can always find such a hypersurface Σ in the overlap $U \cap U'$, so that the above holds for all its points.

Then define a map $\psi: U \to \mathbb{R}^{n+1}$ by $p \to \psi(p) = e^t \phi(p_0)$, which is a local C^∞ diffeomorphism and (U, ψ) is an admissible chart. Let (U', ψ') be another such chart derived from (U', ϕ') with $U \cap U' \neq \Phi$. Then $\psi' \circ \psi^{-1}(\phi(p_0)) = \psi'(p_0) = \phi'(p_0)$ and $\psi' \circ \psi^{-1}(e^t \phi(p_0)) = \psi'(F_t p_0) = e^t \phi'(p_0)$. Thus $\psi' \circ \psi^{-1}$ is locally homogeneous of degree one. The set of all such charts (U, ψ) is therefore compatible with H^{n+1}.

This proves the sufficiency part. Necessity is established in the next section where we construct a nonvanishing vector field X. QED

Theorem 1.4: A C^∞ manifold admits an atlas compatible with H^{n+1} if it is noncompact, or if it is compact and orientable with Euler characteristic zero.

Proof: Follows from Theorem 1.3. QED

Projectors

From now on we shall consider M to be *without origins*. We shall denote the set of all C^∞ homogeneous functions of degree μ on any open subset U of M by $C_\mu^\infty(U)$. The set $C_0^\infty(M)$, which is a subring of the ring of all C^∞- functions on M will play an important role and the homogeneous functions of degree zero will be called *projective invariants*.

A C^∞ vector field on a C^∞ manifold can be defined as a derivation of the algebra of differentiable functions. A projective vector field, or in van Dantzig's terminology, a *projector* of type $(1, 0)$ satisfies an additional property.

Definition: A projector of type $(1, 0)$ ξ is a C^∞-vector field such that $\xi(f) \in C_\mu^\infty(U)$ for every $f \in C_\mu^\infty(U)$.

In local coordinates (X^μ), $\mu = 0, 1, \ldots, n$, on a coordinate chart (U, ϕ), ξ can be expressed as $\xi = \xi^\mu(X)\partial/\partial X^\mu$ where $\xi^\mu(X) = \xi(X^\mu)$. Since $X^\mu \in C_1^\infty(\phi(U))$, $\xi^\mu(X) \in C_1^\infty(\phi(U))$. Thus a projector of type $(1, 0)$ is completely determined by its action on $C_\mu^\infty(U)$ and its components are locally homogeneous functions of degree one.

A particular and important example is the *coordinate projector* X, defined as follows.

Definition: $X(f) = \mu f$ for every $f \in C_\mu^\infty(U)$.

That X is a derivation of $C_\mu^\infty(U)$ is easily seen. Let $f \in C_\mu^\infty(U)$ and $g \in C_\lambda^\infty(U)$. Then $fg \in C_{\mu+\lambda}^\infty(U)$ from (1.7) and $fX(g) + gX(f) = f\lambda g + g\mu f = X(fg)$. Thus $X(f) = 0$ for all $f \in C_0^\infty(U)$. In local coordinates $X = X(X^\mu)\partial/\partial X^\mu = X^\mu \partial/\partial X^\mu$, that is, the components of X are the coordinate functions themselves. Since M is without origins X *is a nonvanishing vector field.* This establishes the necessity condition in Theorem 1.3.

The integral curves of X, in local coordinates, are given by $dX^\mu/dt = X^\mu$, or $X^\mu(t) = e^t X^\mu(0)$.

Definition: A projector ω of type $(0, 1)$ (or a projective 1-form) is a C^∞ 1-form such that $\omega(\xi) \in C_0^\infty(U)$ for every projector ξ of type $(1, 0)$.

In local coordinates let $\omega(\xi) = \omega_\mu \xi^\mu$. Then from (1.8) we have $0 = (\omega_\mu \xi^\mu)_{,\nu} X^\nu = (\omega_{\mu,\nu} X^\nu + \omega_\mu)\xi^\mu$. Or $\omega_{\mu,\nu} X^\nu = -\omega_\mu$, i.e., $\omega_\mu \in C_{-1}^\infty(\phi(U))$. The components of a projector of type $(0, 1)$ are locally homogeneous functions of degree minus one.

Definition: A projector θ of type (r, s) is a C^∞ tensor of type (r, s) such that $\theta(\omega_1, \ldots, \omega_r, \xi_1, \ldots, \xi_s) \in C_0^\infty(U)$ for all projectors $\omega_1, \ldots, \omega_r$ of type $(0, 1)$ and $\xi_1, \ldots, \xi_s$ of type $(1, 0)$.

In local coordinates the components of θ are locally homogeneous functions of degree $r - s$.

Proportion 1.5: The Lie derivative of any projector θ with respect to the coordinate projector X vanishes, i.e., $L_X\theta = 0$.

Proof: Let θ be a projector of type $(0, 0)$, i.e., $\theta = f \in C_P^\infty(M)$. Then $L_X f = X(f) = 0$. If ξ is a projector of type $(1, 0)$ then $L_X\xi = X\xi - \xi X$, so that $(L_X\xi)f = 0$ if $f \in C_P^\infty(M)$. Or $L_X\xi = 0$. And if ω is a projector of type $(0, 1)$, $(L_X\omega)\xi = X(\omega(\xi)) - \omega(L_X\xi) = 0$ since $\omega(\xi) \in C_P^\infty(M)$. For an arbitrary projector θ of type (r, s), $(L_X\theta)(\omega_1, \ldots, \omega_r, \xi_1, \ldots, \xi_s) = X\theta(\omega_1, \ldots, \omega_r, \xi_1, \ldots, \xi_s) - \theta(L_X\omega_1 \ldots, \omega_r, \xi_1, \ldots, \xi_s) \cdots - \theta(\omega_1, \ldots, \omega_r, L_X\xi_1, \ldots, \xi_s) \cdots = 0$. QED

Connections and metric

Definition: An affine connection ∇ on M will be called a projective connection[13] *if $\nabla_\xi\eta$ is a projective vector field whenever ξ and η are.*

Definition: A semi-Riemannian metric g will be called projective if it is a projector of type $(0, 2)$.

Then $g(\xi, \eta) = \langle\xi, \eta\rangle$ is a projective invariant. The following two propositions are easily demonstrated.

Proposition 1.6: Given a projective inner product $\langle\,,\rangle$, the unique Levi-Civita connection ∇, which satisfies

$$\begin{aligned} &\nabla_\xi\eta - \nabla_\eta\xi - [\xi, \eta] = 0 \\ &\zeta\langle\xi, \eta\rangle - \langle\nabla_\zeta\xi, \eta\rangle - \langle\xi, \nabla_\zeta\eta\rangle = 0 \end{aligned} \tag{1.13}$$

for all projective vector fields ξ, η, ζ, is a projective connection

Proposition 1.7:

$$\begin{aligned} &\nabla_X\xi - \nabla_\xi X = 0 \\ &\langle\nabla_X\xi, \eta\rangle + \langle\xi, \nabla_X\eta\rangle = 0 \end{aligned} \tag{1.14}$$

for all ξ, η. In particular, $\langle\nabla_X X, X\rangle = 0$.

From now on we shall assume that M is endowed with a projective metric and the unique projective connection defined in Proposition 1.6.

Proposition 1.8: X generates a local one-parameter group of local isometries of M.

Proof: From Proposition 1.5, $L_X g = 0$. QED

Let C_X be the contraction operator with respect to X. Since $L_X\omega = d \circ C_X\omega + C_X \circ d\omega$ where ω is any projector of degree $(0, 1)$ and d the exterior derivative, we have in view of Proposition 1.5 again

$$d \circ C_X\omega + C_X \circ d\omega = 0. \tag{1.15}$$

In particular, if $\omega = X_*$, the induced projective 1-form of X by $\langle\,,\rangle$, we get

$$dJ + C_X \circ dX_* = 0. \tag{1.16}$$

In contrast to the Riemannian case we have the following negative result which is of physical interest.

Proposition 1.9: There exists no local coordinate system on M in which the connection components $\Gamma^\mu_{\alpha\beta}$ vanish at any given point.

Proof: Suppose $\Gamma^\mu_{\alpha\beta} = 0$ at $p \in M$. Then $g_{\mu\lambda,\nu} = 0$ at p. But since $g_{\mu\lambda}$ is a locally homogeneous function of degree minus two, $g_{\mu\lambda,\nu}X^\nu = -2g_{\mu\lambda}$. Or $g_{\mu\lambda} = 0$ at p, contradicting the nondegeneracy of g. QED

Thus a physical field represented by $g_{\mu\lambda}$ would not satisfy an equivalence principle.

Orthogonal projection on X

The coordinate projector X will play an important role in the theory. We shall assume that $J = \langle X, X\rangle \neq 0$. We write $\xi \perp \eta$ if $\langle\xi, \eta\rangle = 0$. Denote by $\xi^\perp$ the orthogonal complement of the projection of ξ on X, i.e.,

$$\xi^\perp = \xi - J^{-1}\langle\xi, X\rangle X. \tag{1.17}$$

Then $\xi^\perp \perp X$, $X^\perp = 0$; and $\xi \perp X$ implies $\xi^\perp = \xi$. In local cordinates the components of $\xi^\perp$ are $\xi^{(\mu)} = \xi^\mu - J^{-1}\xi^\nu X_\nu X^\mu$. Recall that $\langle\,,\rangle$ induces an isomorphism between projective vector fields and projective 1-forms. Let ξ_* denote the induced 1-form of ξ. Then the orthogonal complement of the projection of a projective 1-form ω on X is given by

$$\omega^\perp = \omega - J^{-1}\omega(X)X_*. \tag{1.18}$$

Then $(X_*)^\perp = 0$, $\omega^\perp(\xi) = \omega(\xi^\perp)$. This can be easily extended to an arbitrary projector θ.

Definition: Two projectors θ_1 and θ_2 are said to be congruent, and we write $\theta_1 \equiv \theta_2$, if $\theta_1^\perp = \theta_2^\perp$.

If $\theta \equiv 0$, the covariant differential of θ, $\nabla\theta$ need not be congruent to zero. For example, let ξ be parallel to X. Then $\nabla\xi(X_*, \eta) = X_*(\nabla_\eta\xi) = \langle X, \nabla_\eta\xi\rangle \neq 0$. Following Jordan we wish to consider a new connection K such that if $\theta \equiv 0$, then the covariant differential $K\theta$ of θ relative to K is also congruent to zero.

Theorem 1.5: Given a projective inner product $\langle\,,\rangle$, there is a unique projective connection K such that

$$K_\eta\xi^\perp = (\nabla_\eta\xi^\perp)^\perp, \tag{1.19}$$

$$\zeta\langle\xi, \eta\rangle = \langle K_\zeta\xi, \eta\rangle + \langle\xi, K_\zeta\eta\rangle \tag{1.20}$$

for all projective vector fields ξ, η, ζ. In fact,

$$K_\eta\xi = \nabla_\eta\xi + J^{-1}\langle\xi, \nabla_\eta X\rangle X - J^{-1}\langle\xi, X\rangle\nabla_\eta X. \tag{1.21}$$

Proof: From (1.13) and (1.20)

$$\langle K_\zeta\xi, \eta\rangle + \langle\xi, K_\zeta\eta\rangle = \langle\nabla_\zeta\xi, \eta\rangle + \langle\xi, \nabla_\zeta\eta\rangle. \tag{1.22}$$

From (1.19)

$$\langle K_\eta\xi, \zeta^\perp\rangle = \langle\nabla_\eta\xi^\perp, \zeta^\perp\rangle. \tag{1.23}$$

for all $\zeta \in P(M)$, and therefore

$$K_\eta\xi = \nabla_\eta\xi - J^{-1}\langle\xi, X\rangle\nabla_\eta X + f(X, \xi, \eta)X, \tag{1.24}$$

where $f(X, \xi, \eta)$ is a projective invariant, which remains to be determined.

If we put $\zeta = \eta$ and $\eta = X$ in (1.22) and use (1.24), we obtain finally

$$\langle\xi, X\rangle[f(X, X, \eta) - J^{-1}\langle\nabla_\eta X, X\rangle] = \langle\xi, \nabla_\eta X\rangle - f(X, \xi, \eta)J. \tag{1.25}$$

Putting $\xi = X$ in (1.25) we get

$$f(X, \xi, \eta) = J^{-1}\langle\xi, \nabla_\eta X\rangle. \qquad \text{QED}$$

Corollary:

$$K_\eta \xi^\perp = (K_\eta \xi)^\perp. \tag{1.26}$$

Thus $\xi^\perp = \xi$ implies $(K_\eta \xi)^\perp = K_\eta \xi$ and $\xi^\perp = 0$ implies $(K_\eta \xi)^\perp = 0$. The property (1.26) applies to any arbitrary projector θ, i.e., $K_\eta \theta^\perp = (K_\eta \theta)^\perp$.

Proposition 1.10: $\theta \equiv 0$ implies $K\theta \equiv 0$.

Proof: It follows from $K_\eta \theta^\perp = (K_\eta \theta)^\perp$ that if θ is parallel to X, then so is $K_\eta \theta$. QED

The property (1.19) also applies to any arbitrary projector θ, i.e.,

$$(K_\eta \theta)^\perp = (\nabla_\eta \theta^\perp)^\perp. \tag{1.27}$$

2. PROJECTIVE THEORY OF RELATIVITY

The projective formalism

The projective formalism of a unified field theory of gravitation and electromagnetism is based on the triple $\{M, \phi, V\}$ where M is a five-dimensional projective manifold with a nondegenerate projective metric, V the (four-dimensional) space-time and $\phi : M \to V$ a surjective C^∞-map such that the induced map $\phi^* : C^\infty(V) \to C^\infty(M)$ given by $\phi^* h = h \circ \phi, h \in C^\infty(V)$ is a surjective map from the C^∞-functions in V onto the C^∞ homogeneous functions of degree *zero* in M.

ϕ is necessarily a homogeneous map of degree zero[14] and, as noted in Proposition 1.2, such a map makes sense even though V is not a projective manifold.

It turns out that a geometrized field theory—formulated exactly along the lines of general theory of relativity—on the five-dimensional manifold M gives, on projection onto the space-time V, a unified and geometrical interpretation of both the gravitational and electromagnetic fields.

Projection onto space-time

We will be considering projections of geometric objects in M onto V. Since ϕ is surjective, ϕ^* is injective. Thus we have

Proposition 2.1: ϕ^* is an isomorphism between the ring of C^∞-functions in V and the ring of C^∞-homogeneous functions of degree zero in M.

In general a smooth surjective map from one manifold onto another does not carry vector fields into vector fields. But in view of Proposition 2.1 and the fact that $\xi(C_0^\infty(M)) \subset C_0^\infty(M)$ for all projective vector fields ξ, we have

Proposition 2.2: If ξ is a C^∞-projective vector field on M, then $\phi_* \xi$ defined by

$$\phi^*((\phi_* \xi)h) = \xi(h \circ \phi) \quad \text{for all } h \in C^\infty(V) \tag{2.1}$$

is a C^∞ vector field on V.

In local coordinates[15] let ϕ be given by $X^\mu \to x^k(X^\mu)$. Then $\xi^k = x^k_{,\mu} \xi^\mu$, where $\xi^\mu = \xi(X^\mu)$ and $(\phi_* \xi)(x^k) = \xi^k$. Note that $\phi_* \xi = 0$ if ξ is parallel to X.

The set $P(M)$ of all C^∞-projective vector fields on M is a module over $C_0^\infty(M)$ [but not over $C^\infty(M)$], and the sets of all C^∞-projective vector fields on M which are orthogonal and parallel to X, to be denoted by $P^\perp(M)$ and $Q(M)$, respectively, are submodules over $C_0^\infty(M)$. In fact, $P(M)$ is the direct sum of $P^\perp(M)$ and $Q(M)$.

In view of Proposition 2.1 and 2.2, ϕ_* is a module homomorphism from $P(M)$, which is a module over $C_0^\infty(M)$, into the set $\chi(V)$ of all C^∞-vector fields on V, which is a module over $C^\infty(V)$. Let $\phi_*^\perp$ be the restriction of ϕ_* to $P^\perp(M)$.

Theorem 2.1: $\phi_*^\perp : P^\perp(M) \to \chi(V)$ is a module isomorphism.

Proof: ϕ_* is surjective, because, according to Proposition 2.1 every derivation of $C^\infty(V)$ induces a derivation of $C_0^\infty(M)$, which in turn induces a derivation of $C_1^\infty(M)$ and defines a projective vector field. Now

$$\chi(V) \simeq P(M)/\mathrm{Ker}\phi_* = \frac{P(M)}{Q(M)}. \tag{2.2}$$

But $P(M)/Q(M) \simeq P^\perp(M)$. Thus $P^\perp(M) \simeq \chi(V)$ and the isomorphism is given by $\phi_*^\perp$. QED

If θ is a projector of type $(0, s)$, we define $\phi_* \theta$ by

$$\phi^*((\phi_* \theta)(\eta_1, \ldots, \eta_s)) = \theta(\phi_*^\perp \eta_1, \ldots, \phi_*^\perp \eta_s) \tag{2.3}$$

for all $\eta_i \in \chi(V)$. The projection $\phi_* g$ of the metric projective tensor g on M is again a nondegenerate metric on V and signature of g = (signature of $\phi_* g$, sign of J). We have

$$\langle \phi_* \xi_1, \phi_* \xi_2 \rangle_V = \langle \xi_1^\perp, \xi_2^\perp \rangle_M = \langle \xi_1, \xi_2 \rangle_M - J^{-1} \langle \xi_1, X \rangle_M \langle \xi_2, X \rangle_M \tag{2.4}$$

for all $\xi_i \in P(M)$. If $\eta, \xi \perp X$, it follows from (1.21) that $\phi_*(K_\eta \xi) = \phi_*(\nabla_\eta \xi)$.

Theorem 2.1: The map $D : \chi(V) \times \chi(V) \to \chi(V)$ given by

$$(\eta, \omega) \to D_\eta \omega = \phi_*(\nabla_{\psi\eta} \psi\omega), \tag{2.5}$$

where $\psi = (\phi_*^\perp)^{-1}$, is the symmetric Riemannian connection on V relative to $\phi_* g$.

Proof: We have evidently $D_\eta(\omega_1 + \omega_2) = D_\eta \omega_1 + D_\eta \omega_2$ and $D_{\eta_1 + \eta_2}(\omega) = D_{\eta_1} \omega + D_{\eta_2} \omega$. And $\psi(h\eta) = (h \circ \phi)\psi\eta$, $h \in C^\infty(V)$. So that $D_{h\eta}\omega = hD_\eta \omega$. Also $D_\eta(h\omega) = \phi_*(\nabla_{\psi\eta}((h \circ \phi)\psi\omega)) = hD_\eta \omega + \eta(h)\omega$. Finally, $\langle D_\zeta \eta, \omega \rangle_V + \langle \eta, D_\zeta \omega \rangle_V = \langle \nabla_{\psi\zeta} \psi\eta, \psi\omega \rangle_M + \langle \psi\eta, \nabla_{\psi\zeta} \psi\omega \rangle_M = \psi\zeta \langle \psi\eta, \psi\omega \rangle_M = \psi\zeta(\phi^* \langle \eta, \omega \rangle_V) = \zeta \langle \eta, \omega \rangle_V$. Therefore D is the symmetric Riemannian connection relative to $\phi_* g$. Note that $\phi_*(\nabla_{\psi\eta} \psi\omega) = \phi_*(K_{\psi\eta} \psi\omega)$. QED

Consider now the curvature projective tensor relative to the connection K on M,

$$R_K(\xi_1, \xi_2)\xi_3 = K_{\xi_1} K_{\xi_2} \xi_3 - K_{\xi_2} K_{\xi_1} \xi_3 - K_{[\xi_1, \xi_2]} \xi_3. \tag{2.6}$$

The projection map ϕ_* is easily extended to arbitrary projectors.

Theorem 2.3: $\phi_* R_K = R_D$, the curvature tensor in V relative to the connection D.

Proof:

$$D_\eta D_\omega \zeta = \phi_*(K_{\psi\eta} \psi(D_\omega \zeta)) = \phi_*(K_{\psi\eta} \psi(\phi_*(K_{\psi\omega} \psi\zeta)))$$
$$= \phi_*(K_{\psi\eta}(K_{\psi\omega} \psi\zeta)^\perp) = \phi_*(K_{\psi\eta} K_{\psi\omega} \psi\zeta).$$

Also,

$$D_{[\eta, \omega]} \zeta = \phi_*(K_{\psi[\eta, \omega]} \psi\zeta) = \phi_*(K_{[\psi\eta, \psi\omega]} \psi\zeta).$$

Thus

$$R_D(\eta, \omega)\zeta = \phi_*(R_K(\psi\eta, \psi\omega)\psi\zeta) = (\phi_* R_K)(\eta, \omega)\zeta.$$

QED

Field equations

Suppose we have a projector equation on M.

For example, let

$$\theta(\xi,\eta)=0 \tag{2.7}$$

for all projective vector fields ξ,η on M, where θ is some projector of type (0, 2). Then

$$\theta(\xi,\eta)=\theta(\xi^\perp+fX,\eta^\perp+gX)=\theta(\xi^\perp,\eta^\perp)+f\theta(X,\eta^\perp)+g\theta(\xi^\perp,X)+fg\theta(X,X)=0, \tag{2.8}$$

where $f,g\in C_0^\infty(M)$. Define two projectors θ_1,θ_2 of type (0, 1) by $\theta_1(\eta)=\theta(X,\eta)$, $\theta_2(\xi)=\theta(\xi,X)$ and a projective invariant $\theta_{12}=\theta(X,X)$. Then (2.8) implies

$$\theta(\xi,\eta)=\theta^\perp(\xi,\eta)+f\theta_1^\perp(\eta)+g\theta_2^\perp(\xi)+fg\theta_{12}^\perp=0. \tag{2.9}$$

Thus $\theta=0$ implies $\theta^\perp=0$, $\theta_1^\perp=0$, $\theta_2^\perp=0$, $\theta_{12}^\perp=0$, each of which are orthogonal to X and, therefore, in one-one correspondence with tensors on V. A projector equation on M therefore corresponds to several tensor equations on V. In terms of components, $\theta_{\mu\nu}=0$ corresponds to $\theta_{ik}=0$, $\theta_{i\mu}X^\mu=0$, $\theta_{\mu k}X^\mu=0$, and $\theta_{\mu\nu}X^\mu X^\nu=0$, where $\theta_{i\mu}=\theta_{\lambda\mu}g_i^\lambda$, $\theta_{\mu k}=\theta_{\mu\lambda}g_k^\lambda$, $\theta_{ik}=\theta_{\lambda\mu}g_i^\lambda g_k^\mu$, $g_k^\lambda x^l_{,\lambda}=\delta_k^l$.

Let F be an antisymmetric projector of type (0, 2) defined as follows:

$$F(\xi,\eta)=2J^{-1}\langle\xi,K_X\eta\rangle, \tag{2.10}$$

$$[KF](\xi,\eta,\zeta)=(K_\zeta F)(\xi,\eta)+(K_\xi F)(\eta,\zeta)+(K_\eta F)(\zeta,\xi). \tag{2.11}$$

Theorem 2.4:

$$[KF]=0. \tag{2.12}$$

Proof: We first show that $[KF]=[KF]^\perp$ and then $[KF]^\perp=0$. $F(\xi,X)=2J^{-1}\langle\xi,K_XX\rangle=0$ because $K_XX=J^{-1}\langle X,\nabla_XX\rangle X=0$. Since F is antisymmetric, $F\perp X$, so that $KF\perp X$, $[KF]\perp X$. Or $[KF]=[KF]^\perp$.

Now $(K_\zeta F)(\xi,\eta)=\zeta(F(\xi,\eta))-F(K_\zeta\xi,\eta)-F(\xi,K_\zeta\eta)$. So that

$$\begin{aligned}\tfrac12(K_\zeta F)(\xi,\eta)&=\zeta(J^{-1})\langle\xi,K_X\eta\rangle+J^{-1}\{\zeta(\langle\xi,K_X\eta\rangle)-\langle K_\zeta\xi,K_X\eta\rangle-\langle\xi,K_XK_\zeta\eta\rangle\}\\&=-2J^{-2}\langle K_\zeta X,X\rangle\langle\xi,K_X\eta\rangle+J^{-1}\langle\xi,K_\zeta K_X\eta-K_XK_\zeta\eta\rangle.\end{aligned}$$

Suppose now that $\xi,\eta,\zeta\perp X$. Then $K_\zeta K_X\eta=K_\zeta(\nabla_X\eta)^\perp=(\nabla_\zeta(\nabla_X\eta)^\perp)^\perp$. Since $\xi\perp X$,

$$\begin{aligned}\langle\xi,K_\zeta K_X\eta\rangle&=\langle\xi,\nabla_\zeta(\nabla_X\eta)^\perp\rangle\\&=\langle\xi,\nabla_\zeta\nabla_X\eta\rangle-J^{-1}\langle\nabla_X\eta,X\rangle\langle\xi,\nabla_\zeta X\rangle.\end{aligned}$$

So that

$$\langle\xi,K_\zeta K_X\eta-K_XK_\zeta\eta\rangle=\langle\xi,\nabla_\zeta\nabla_X\eta-\nabla_X\nabla_\zeta\eta\rangle-J^{-1}\{\langle\eta,\nabla_XX\rangle\langle\zeta,\nabla_X\xi\rangle+\langle\xi,\nabla_XX\rangle\langle\eta,\nabla_X\zeta\rangle.$$

Also $\langle\xi,K_X\eta\rangle=\langle\xi,\nabla_X\eta\rangle$ and $\langle K_\zeta X,X\rangle=\langle\nabla_\zeta X,X\rangle=\langle\nabla_X\zeta,X\rangle=-\langle\nabla_XX,\zeta\rangle$ by (1.14).

Substituting above we find that

$$\begin{aligned}\tfrac12(K_\zeta F)(\xi,\eta)&=J^{-2}\{3\langle\nabla_XX,\zeta\rangle\langle\xi,\nabla_X\eta\rangle-[\langle\nabla_XX,\zeta\rangle\langle\xi,\nabla_X\eta\rangle\\&\quad+\langle\nabla_XX,\xi\rangle\langle\eta,\nabla_X\zeta\rangle+\langle\nabla_XX,\eta\rangle\langle\zeta,\nabla_X\xi\rangle]\}\\&\quad+J^{-1}\langle\xi,\nabla_\zeta\nabla_X\eta-\nabla_X\nabla_\zeta\eta\rangle.\end{aligned}$$

If we now cyclically permute ζ,ξ,η and add, both the terms vanish—the second because of Bianchi identities. Thus $[KF]^\perp=0$. QED

The projection of F onto space-time can be interpreted as the electromagnetic field tensor and that of (2.12), as one set of Maxwell's equations. In local coordinates the components of F are given by $F_{\rho\sigma}=J^{-1}(X_{\rho\sigma}+J^{-1}X_{\mu\rho}X^\mu X_\sigma-J^{-1}X_\rho X_{\mu\sigma}X^\mu)$ where $X_\mu=g_{\mu\lambda}X^\lambda$ and $X_{\rho\sigma}=X_{\sigma,\rho}-X_{\rho,\sigma}$.

In view of the relationship (1.21) between the two connections K and ∇ (and, consequently between the curvature tensors R_K and R_∇) and Theorem 2.3, it is now easy to express the projections of the curvature and Ricci tensors, Ricci scalar, etc., relative to the connection ∇ in M onto the corresponding quantities relative to the connection D in V.

For example, consider the Einstein equations in M in local coordinates

$$R_{\mu\lambda}=0, \tag{2.13}$$

where $R_{\mu\lambda}$ are the components of the Ricci tensor relative to ∇.

Then according to the procedure (2.9), one obtains[9]

$$R_{kl}=R_{kl}^D-\tfrac12J^{-1}X_{kj}X^j_{\ l}-\tfrac12J^{-1}J_{,k;l}+\tfrac14J_{,k}J_{,l}=0,$$

$$R_{k\nu}X^\nu=\tfrac12X^l_{k;l}+\tfrac14J^{-1}X^l_kJ_{,l}=0,$$

$$R_{\mu\nu}X^\mu X^\nu=\tfrac12J^{;k}_{;k}-\tfrac14J^{-1}J^{,k}J_{,k}-\tfrac14X_{kl}X^{kl}=0,$$

where R_{kl}^D are the components of the Ricci tensor relative to the connection D in V.

In order to obtain exact correspondence with the Einstein–Maxwell equations (1.1) it is necessary to make the additional assumption $J=-1$, in which case for any projective vector field ξ, $0=\xi(\langle X,X\rangle)=2\langle\nabla_\xi X,X\rangle=-2\langle\nabla_XX,\xi\rangle$. Since $\langle\,,\rangle$ is nondegenerate, $\nabla_XX=0$. Then $F(\xi,\eta)=\frac12\langle\xi,\nabla_X\eta\rangle$ and $F_{\rho\sigma}=X_{\rho\sigma}$.

We refer to Ref. 9 for detailed derivation of the field equations.

Geodesics

We have seen that any $\xi\in P(M)$ can be written as $\xi=\alpha+aX$, where $\alpha\in P^\perp(M)$ and $a\in C_0^\infty(M)$. Conversely, any object of the form $\alpha+aX$ is in $P(M)$; thus α and a are independent of each other.

Let $\gamma(t)$ be a ∇-geodesic in M with tangent vector $\gamma_*(t)$ of the form $\alpha+aX$, for some t, i.e.,

$$\nabla_{\gamma_*(t)}\gamma_*(t)=0 \tag{2.14}$$

or

$$\nabla_\alpha\alpha+a\nabla_X\alpha+\alpha(a)X+a\nabla_\alpha X+a^2\nabla_XX=0. \tag{2.15}$$

The curve $\phi\cdot\gamma$ has tangent $\phi_*(\alpha+aX)=\phi_*\alpha$:

$$\begin{aligned}D_{\phi_*\alpha}\phi_*\alpha&=\phi_*(\nabla_\alpha\alpha)\\&=-2(\phi\cdot a)\phi_*(\nabla_X\alpha)\quad\text{if } J=\text{const.}\end{aligned} \tag{2.16}$$

This is equivalent to

$$\langle D_{\phi_*\alpha}\phi_*\alpha,Y\rangle_V+2(\phi\cdot a)\langle\phi_*(\nabla_X\alpha),Y\rangle_V=0 \tag{2.17}$$

for all $Y\in\chi(V)$, or

$$\langle\psi(D_{\phi_*\alpha}\phi_*\alpha),\psi Y\rangle_M+2a\langle(\nabla_X\alpha)^\perp,\psi Y\rangle_M=0, \tag{2.18}$$

$$\langle\nabla_\alpha\alpha,\psi Y\rangle+2a\langle\nabla_X\alpha,\psi Y\rangle=0, \tag{2.19}$$

$$\langle \nabla_\alpha \alpha, \psi Y\rangle + aF(\psi Y, \alpha) = 0. \qquad (2.20)$$

Projecting this into V, in a chart with $\alpha = dx^i/ds$, this becomes

$$\frac{D^2x^i}{Ds^2} + aF^i_k \frac{dx^k}{ds} = 0 \qquad (2.21)$$

which corresponds to the equation of motion of a charged particle if $a = -e/m$. Now

$$\begin{aligned}\alpha(a) &= \xi(J^{-1}\langle \xi, X\rangle)\\ &= -2J^{-2}\langle \nabla_\xi X, X\rangle\langle \xi, X\rangle + J^{-1}\langle \nabla_\xi \xi, X\rangle + J^{-1}\langle \xi, \nabla_\xi X\rangle\\ &= 2J^{-2}\langle \nabla_X X, \xi\rangle\langle \xi, X\rangle + J^{-1}\langle \nabla_\xi \xi, X\rangle + 0 \qquad (2.22)\\ &= 0 \quad \text{if } J = \text{const}\end{aligned}$$

since $\nabla_\xi \xi = 0$ along a geodesic. Therefore, a is constant along γ, and the assumption that $J = \text{const}$ is equivalent to assuming a constant charge/mass ratio.

In particular, of course, if $a = 0$ then $\phi \cdot \sigma$ is a geodesic in V.

4. CONCLUSION

One of the advantages of the intrinsic formulation is that it can bring into the open the main ideas of the theory, which tend to hide behind coordinates. It becomes clear that Jordan's theory does not use homogeneity in any essential way; it is an artifice to ensure a relation between projectors and space–time tensors. For example, it can be shown[16] that the converse of Proposition 1.5 also holds, that is, if $L_X\theta = 0$ for any tensor field θ, then θ must be a projector. Thus one could start with a 5-dimensional manifold with a fixed nonvanishing vector field X and consider those tensors whose Lie derivatives with respect to X vanishes. The quotient space, defined by the equivalence relation $p \sim q$ if p and q lie on the same integral curve of X, can under certain conditions be made into a proper manifold which then could be identified with space–time.

ACKNOWLEDGMENTS

We would like to thank Professors W. Greub, S. Halperin, D. Norman, and J. R. Vanstone for their critical comments.

*Supported in part by NRC (Canada) under operating Grant A4054.
†NRC (Canada) 1967 Science Predoctoral Fellow.

[1] H. Weyl, Sitz. Ber. Preuss, Akad. Wiss. 465 (1918).
[2] Th. Kaluza, Sitz. Ber.Preuss, Akad. Wiss. 966 (1921).
[3] O. Klein, Z. Phys. **37**, 895 (1926).
[4] O. Veblen and B. Hoffmann, Phys. Rev. **36**, 810 (1930).
[5] O. Veblen, *Projektive Relativitätstheorie* (Springer, Berlin, 1933).
[6] J. A. Schouten and D. van Dantzig, Proc. Akad. Wetensch. Amsterdam **34**, 1398 (1931).
[7] D. van Dantzig, Math. Ann 106, 400 (1932).
[8] P. Jordan, Göttinger Nachr. 74 (1945).
[9] P. Jordan, *Schwerkraft und Weltall* (Vieweg, Braunschweig, 1955), 2nd. ed.
[10] G. Ludwig, *Fortschritte der projektiven Relativitästheorie* (Vieweg, Braunschweig, 1951).
[11] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'electromagnetisme* (Masson, Paris, 1955).
[12] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963), Vol I.
[13] Our definition differs from the standard one given by, for example, E. Cartan *Leçons sur la théorie des espaces à connexion projective* (Gauthier-Villars, Paris, 1937).
[14] Jordan, for example, assumed only that ϕ is a homogeneous map of degree zero. However, rigourously speaking one needs somewhat more in order to project projective vector fields from M onto V.
[15] From now on X^μ and x^k will denote local coordinates in M and V respectively, and all Greek indices and all Latin indices take values 0,1,...,4 and 0,1,...,3 respectively.
[16] G. T. Evans, Ph.D. thesis (Toronto, 1973).

Optimal analytic extrapolation for the scattering amplitude from cuts to interior points

S. Ciulli

NORDITA, Copenhagen, Denmark, and Institute for Atomic Physics, Bucharest, Romania

G. Nenciu

Institute for Atomic Physics, Bucharest, Romania

(Received 23 November 1971)

Given a data function together with an error corridor for the scattering amplitude along some finite part of the cuts, one can construct effectively the whole set of analytic functions ("admissible amplitudes"), compatible with these conditions and bounded by a certain number M on the remaining part of the cuts. Depending on the actual value of an important constant ϵ_0 computed from the data function and the bound M, this set may be void. If not, in every point of the cut plane the set of values of the admissible amplitudes fills densely a circle; explicit formulas are given for its radius $\hat{\eta}(z)$ and center $\hat{f}(z)$, the latter being the best possible estimate for the whole set. In contrast to the linear extrapolation obtained by Poisson weighted dispersion relations, here nonlinear functional methods were used. This paper contains an appendix written by Professor C. Foias, on some functional analytical methods used in connection with the computation of the numerical value of the constant ϵ_0.

I. INTRODUCTION

In realistic particle-physics problems, information is available only along some limited parts of the cuts of the energy (or momentum transfer) complex plane of the scattering amplitude, and the problem one is usually faced with is to extract from this limited, error-affected knowledge, information on the behavior of the amplitude of other reactions or at energies outside the initial range.[1–5]

This is in general an ill-posed mathematical problem, in the sense that small changes (errors) in the input data could provide incontrollable responses in the output. Nevertheless, following an idea first emphasized[6] at the 1969 Lund Conference, Carleman weight functions can be used to write down[7,8] those dispersion relations (sum rules) which exploit this limited, error affected information in the most economical way, in the sense that any other weighted dispersion relation would lead to greater error-bounds in the results. As already emphasized in Ref. 7, the dispersion relations do not exhaust the optimization problem of the extrapolation procedures: The aim of the present paper is precisely to find by nonlinear methods this absolute optimum, as well as to construct *all possible* analytic functions $f(z)$ compatible with some given, error-affected, histogram $h(z)$, on some limited part of the energy (or momentum) cut complex plane. [z is here the relevant (energy, momentum, cosine, and so on) variable and $f(z)$ is the amplitude itself, or one of its combinations with some given complex functions].

We have perhaps all experienced the trying situation of being asked by some experimentalist friend to find a close form, say for the transverse momentum dependence of some cross-section, in terms of "usual" functions—cosines, logarithms, and so on. If possible, even exhibiting a Regge behavior. To our question "what for?", his answer would probably be "in order to have some easy-to-remember formula instead of these long intricate tables"; but this answer usually hides also the secret hope that our formula could apply to a much wider range of momenta than that where the measurements were made, representing, so to say, an "objective" physical reality.

This standpoint might seem naive, but under a more attentive consideration one sees that it prevails over the whole of theoretical physics. Indeed, we ought to remember that each successful theory is, in a metaphorical sense, a curve which lies inside the error corridor along the whole range of the present physical informations. (This goes for succesful theories only! For, usually, we are content with theories which only partially pass through the present error corridor!) Such a theory makes definite predictions also outside the range of the actual physical information, but, of course, one can have no confidence in these "predictions" for, in general, there are many possible theories passing through the same error corridor whose predictions can differ considerably outside the range of the present information. This is a serious problem, encountered, of course, not only in physics but in every branch of science.

A sensible solution to this problem would be to work simultaneously with the *whole set* of theories passing through the error corridor. The drawback of such an

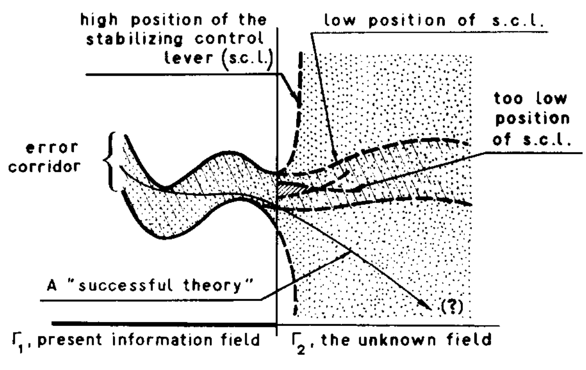

FIG. 1.

Copyright © 1973 American Institute of Physics

approach is, of course, evident, since these theories can in principle make arbitrary predictions outside the range of the present informations (see Fig. 1). In the good old days when the theorist felt the results before the actual calculations were done, there was an easier choice between the possible outcomes. Nowadays the situation has changed drastically, first of all because of the considerable broadening of the range of choice owing to the informational boom but also due to the ever increasing degree of abstraction. Research has become *more and more indirect*, with the consequence that the leading principles are now far outside the reach of the physical measurements. Even the most common concepts of theoretical physics, such as particles, resonances, exchanges and so on, are beyond the actual experimental range, and Landau, for instance, raised the question whether it would not be wiser to leave some of these principles out. (He included even the concept of "interaction" among the other presumable ill concepts.) Indeed, Calucci, Fonda and Ghiraldi[9] showed that a suitably chosen nonresonant background can simulate as well as one would like a Breit–Wigner resonance curve, so that those who are fitting cross-section bumps with two parameter resonance formulas get exactly what they had expected, from the beginning, to get!

In principle the new scientific approach would have to cope with these two, interwoven problems: (i) working with the whole set of theories passing through the error corridor of the present knowledge; (ii) finding new theoretical concepts to replace the dated ones and *controlling the behavior* of the set of possible theories *outside* the range of our present knowledge. Of great importance among these concepts are those which control "the opening" of the set of possible theories (see Fig. 1): We shall call them "regularizers" or "stabilizing" control levers. (Although the former term is already used in the theory of ill-posed problems of mathematical physics,[10] we shall give preference to the latter because of its more specific content.)

To make things more palpable, consider the problem of solving in some function space the equation $Af=h$, with f the input unknown function and h the experimental data, from which we try to deduce f. We assume that A is a continuous operator, with a unique inverse A^{-1}. The uniqueness of the inverse does not yet imply, in practice, the unique determination of f from h, since, if A^{-1} is discontinuous, arbitrarily small variation in h will cause uncontrollably large changes in f. This is a very frequently met situation, e.g., A is the operator of taking the restriction of an analytic function to an open curve or to a part of the boundary (note that A^{-1} is then the operator of analytic continuation of the function to all interior points, and hence, is unique); or consider the Fredholm operator of the first kind encountered in the theory of diffusion (solving backwards the heat equation), or in geophysics (solving in homogeneous Laplace equations with data given on an open boundary), etc.

The problem is: what complementary conditions should be imposed in order to stabilize the problem? Obviously a stabilizing lever of this problem would be every condition which restricts the set in which one searches for the solution (the admissible f's) to a compact set, since that is a sufficient condition to make A^{-1} continuous too.

The problem of devising a general stable approach for particle physics, as a whole, might seem a formidable task; nevertheless, special problems can be treated rather easily. An instructive example is provided by the problem of finding *all possible* amplitudes $f(z)$ passing for $z \in \Gamma_1$ (the actual range of measurements) through the error corridor

$$|f(z)-h(z)|_{z\in\Gamma_1}<\epsilon, \tag{1.1}$$

where the complex function $h(z)$—the experimentally measured histogram—is given along the part Γ_1 of the cuts of the analyticity domain (D) of $f(z)$. The analyticity of $f(z)$ in (D) represents itself a stabilizing concept, but alone it is unsufficient, as there are many analytic functions satisfying (1.1) which, however, differ arbitrarily much outside Γ_1. To turn the set of the admissible functions into a compact, we shall introduce also the stabilizing parameter M, adding to (1.1) the boundness condition

$$|f(z)|<M, \quad z\in\Gamma_2, \tag{1.2}$$

where $\Gamma=\Gamma_1+\Gamma_2$ represents the whole boundary (cuts) of the complex cut plane (D). As it was stated above, the aim of this paper is to construct effectively every possible analytic function $f(z)$ in D, satisfying the inequality (1.1) on Γ_1—$h(z)$ and ϵ being given. It will be shown that these functions $f(z)$ are labelled, not only by the value of the stabilizing parameter M, but also by some "running index" $\psi(\zeta)$, more precisely by a general unimodular function in the unit disk, to be defined later. The difference between the role of the "control level" M and the "running index" $\psi(\zeta)$ will appear clearly throughout this paper.

As a by-product of this theory, we shall find the value of the center $\hat{f}(z)$ of the set of the values of all possible $f(z)$ in every given point z, this center $\hat{f}(z)$ being the best estimate ever found for the extrapolation of the scattering amplitudes satisfying (1.1). A comparison with the optimal dispersion relation yield $\hat{h}(z)$ (see Refs. 7 and 8) is then performed.

II. DESCRIPTION OF THE PROBLEM

Our problem amounts to the construction of *analytic* functions to be used in the analytic continuation of the physical data in the holomorphy domain of the scattering amplitude. For what follows, it is convenient to transform this holomorphy domain[11] into the unit circle of a suitably chosen conformal variable $\zeta(z)$ (for technical details see Sec. 2 of Ref. 7), the physical region Γ_1, where the measurements were performed [where the function $h(\zeta)$ is given] being depicted on the right semicircle $\zeta=e^{i\theta}$, $-\pi/2<\theta<\pi/2$ (see Fig. 2). Following Ref. 6, in order to express the conditions (1.1) and (1.2) as a single one, we multiply both the amplitude $f(\zeta)$ ($\zeta\in D$) and the histogram $h(\zeta)$ ($\zeta\in\Gamma_1$) with a suitable chosen function of Carleman type

$$C_0(M/\epsilon;\zeta)=\exp\{-\ln(M/\epsilon)[\omega(\zeta)+i\tilde{\omega}(\zeta)]\}, \tag{2.1}$$

$$\tilde{f}(\zeta)=C_0(M/\epsilon;\zeta)f(\zeta), \tag{2.2}$$

$$\tilde{h}(\zeta)=C_0(M/\epsilon;\zeta)h(\zeta), \tag{2.3}$$

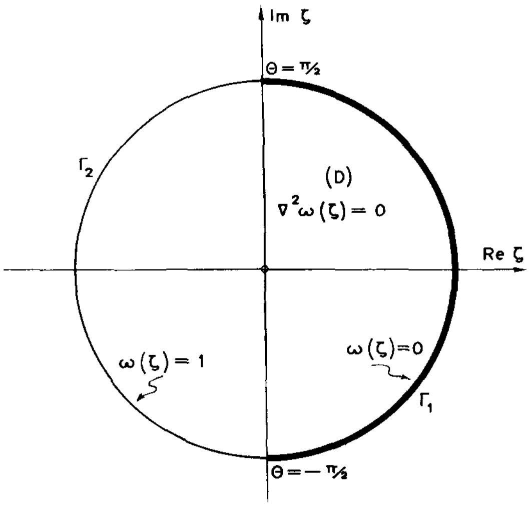

FIG. 2. The unit $\zeta(z)$ circle in which the z-cut plane was mapped. If the z cuts are $(-\infty, z_1)$ and (z_2, ∞) with data given along $\Gamma_1=(z_2, z_3)$, then $\zeta(z)=\{1-(1-u^2)^{1/2}\}/u$, where $u=[(z_1+z_2-2z_3)z+z_3(z_1+z_2)-2z_1z_2]/[(z-z_3)(z_2-z_1)]$.

where $\omega(\zeta)$ is a potential (a harmonic measure) defined to be *zero* on Γ_1 and *one* on the remainder part, Γ_2, of the cuts $\Gamma=\Gamma_1+\Gamma_2$,

$$\begin{aligned}\nabla^2\omega(\zeta)&=0 \quad \text{for } \zeta\in D,\\ \omega(\zeta)&=0 \quad \text{for } \zeta\in\Gamma_1,\\ \omega(\zeta)&=1 \quad \text{for } \zeta\in\Gamma_2,\end{aligned} \tag{2.4}$$

and where $\tilde{\omega}(\zeta)$ is its harmonic conjugate (the stream line function). For the case depicted in Fig. 2,

$$\begin{aligned}\omega(\zeta)+i\tilde{\omega}(\zeta)&=\tfrac{1}{2}-(2/\pi)\arctan\zeta\\ &\equiv\tfrac{1}{2}+(i/\pi)\ln[(1+i\zeta)/(1-i\zeta)].\end{aligned} \tag{2.5}$$

As, owing to (2.1) and (2.4), the modulus of $C_0(M/\epsilon;\zeta)$ is equal to 1 on Γ_1 and to ϵ/M on Γ_2, taking *by definition* $h(\zeta)$ equal to zero on Γ_2,

$$\tilde{h}(\zeta)=h(\zeta)\overset{\mathrm{DEF}}{=}0 \quad \text{for } \zeta\in\Gamma_2, \tag{2.6}$$

[initially the histogram was defined only on the "known cut", so that we are free to complement this definition with Eq. (2.6)] the conditions (1.1) and (1.2) are, equivalent with the unique one,

$$|\tilde{f}(\zeta)-\tilde{h}(\zeta)|_{\zeta\in\Gamma_1+\Gamma_2}<\epsilon \tag{2.7}$$

for the weighted amplitude [see Eqs. (2.2) and (2.3)] $\tilde{f}(\zeta)$.

We are thus left with the well-stated problem that, given a function $\tilde{h}(\zeta)$ on the unit circle $\Gamma=\Gamma_1+\Gamma_2$,

$$\begin{aligned}&\tilde{h}(e^{i\theta})=h(e^{i\theta})\exp[-(i/\pi)\ln(M/\epsilon)\ln\tan(\pi/4-\theta/2)]\\ &\qquad \text{for } -\pi/2<\theta<\pi/2,\\ &\tilde{h}(e^{i\theta})=0 \text{ for } \pi/2<\theta<3\pi/2,\end{aligned} \tag{2.8}$$

to find *all* functions $\tilde{f}(\zeta)$ *analytic* inside the unit circle D which approximate $\tilde{h}(\zeta)$ on Γ according to (2.7).

The stabilizing role of the parameter M is now transparent; indeed, we shall show in Sec. 4 that for every two "admissible" functions $\tilde{f}_1(\zeta)$ and $\tilde{f}_2(\zeta)$ we have

$$|\tilde{f}_1(\zeta)-\tilde{f}_2(\zeta)|\leq 2\eta(\zeta), \tag{2.9}$$

so that, owing to (2.2), the difference between two admissible amplitudes cannot exceed

$$|f_1(\zeta)-f_2(\zeta)|\leq 2\eta(\zeta)/|C_0(M/\epsilon;\zeta)|. \tag{2.10}$$

From (2.7) it is obvious that we have at least $\eta(\zeta)\leq\epsilon$ and in Sec. 4 an algorithm will be given for the actual form of $\eta(\zeta)$.

In all previous extrapolation procedures we have, so far, assumed the existence of at least one analytic function (the amplitude itself) satisfying conditions (1.1) and (1.2), or the equivalent condition (2.7). Nevertheless, for some histograms $\tilde{h}(\zeta)$ and for some ϵ there *may be no analytic function* $\tilde{f}(\zeta)$ at all satisfying the condition (2.7). For instance, let us suppose that the Carleman weighted histogram $\tilde{h}(\zeta)$ has the special form

$$\tilde{h}(e^{i\theta})=\tilde{h}_1(e^{i\theta})+e^{-i\theta}, \tag{2.11}$$

where

$$\tilde{h}_1(e^{i\theta})=\sum_{n=0}^{\infty}c_n e^{in\theta} \tag{2.12}$$

is the limit of a function $\tilde{h}_1(\zeta)$ holomorphic inside (D). It can readily be shown that there are no holomorphic functions inside (D) which can approximate $\tilde{h}$ given by (2.11) with an error ϵ smaller than one. Indeed, putting

$$\tilde{f}(\zeta)=\tilde{h}_1(\zeta)-\chi_1(\zeta), \tag{2.13}$$

condition (2.7) reads

$$|-\chi_1(\zeta)-1/\zeta|_{\zeta\in\Gamma}\leq\epsilon. \tag{2.14}$$

But on the unit circle we have

$$|\chi_1(\zeta)+1/\zeta|\equiv|1+\zeta\cdot\tilde{\chi}_1(\zeta)| \quad \text{for } |\zeta|=1,$$

and, as $1+\zeta\cdot\chi_1(\zeta)$ (in contradistinction to χ_1+1/ζ) is holomorphic inside (D) and equal to 1 at the origin, from the principle of the maximum of the modulus it follows that the value of ϵ of (2.14) cannot be smaller than 1!

Coming back to the general problem, we note that under very general conditions (Fourier expandibility of $\tilde{h}$ on the unit circle) the histogram can be cast into the form

$$\tilde{h}(\zeta)=\tilde{h}_1(\zeta)+\tilde{h}_2(\zeta), \tag{2.15}$$

where $\tilde{h}_1$ and $\tilde{h}_2$ are limits of functions holomorphic respectively inside and outside the unit circle:

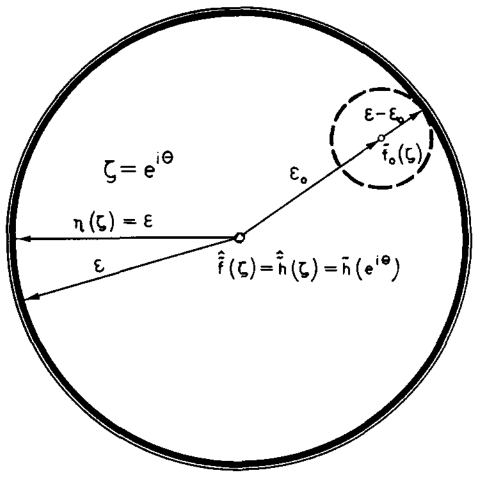

FIG. 3. In a *boundary point* $|\zeta|=1$ the circle filled by the admissible amplitudes coincides with the circle of radius ϵ, the "Nevanlinna bound" for the error encountered in Poisson weighted dispersion relations (Ref. 7). The point ζ being on the boundary, the harmonic function $\hat{h}(\zeta)$ coincides also with the w weighted histogram $\tilde{h}(\zeta)$. In every boundary point, the distance between $\hat{h}(\zeta)$ and the minimal function $\tilde{f}_0$ is ϵ_0. Admissible amplitudes which are nearer to the center at a particular point $\zeta \in \Gamma_1$, have to go away from it at other boundary points $|\zeta|=1$.

$$\tilde{h}_1(\zeta)=\sum_{n=0}^{\infty} c_n\zeta^n, \tag{2.16a}$$

$$\tilde{h}_2(\zeta)=\sum_{m=1}^{\infty} c_{-m}\zeta^{-m}, \tag{2.16b}$$

$$\tilde{h}(e^{i\theta})=\sum_{-\infty}^{\infty} c_n e^{in\theta}=\tilde{h}_1(e^{i\theta})+\tilde{h}_2(e^{i\theta}), \tag{2.16c}$$

where

$$c_{-n}=\frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} h(e^{i\theta})\exp\left\{\frac{-i}{\pi}\ln\left(\frac{M}{\epsilon}\right)\ln\left[\tan\left(\frac{\pi}{4}-\frac{\theta}{2}\right)\right]\right\}e^{in\theta}\,d\theta. \tag{2.16d}$$

It is clear that in our problem, the trouble comes from the nonanalytic component $\tilde{h}_2(\zeta)$. Indeed, again writing $\tilde{f}(\zeta)$ in the form (2.13), we are left to find those holomorphic functions $\chi_1(\zeta)$ which according to (2.7) satisfy

$$|\chi_1(\zeta)+\tilde{h}_2(\zeta)|_\Gamma \leqslant \epsilon. \tag{2.17}$$

From the previous example [in which $\tilde{h}_2$ was set equal to $1/\zeta$, see (2.11)] it is apparent that we cannot approximate arbitrarily well the nonanalytic component $\tilde{h}_2(\zeta)$ of the histogram by analytic functions $-\chi_1(\zeta)$; in other words, for a given $\tilde{h}_2(\zeta)$ there exists a number ϵ_0 so that for every holomorphic function $\chi_1(\zeta)$ in D we have

$$\max_{\zeta\in\Gamma}|\chi_1(\zeta)+\tilde{h}_2(\zeta)| \geqslant \epsilon_0[\tilde{h}_2]. \tag{2.18}$$

That holomorphic function which reaches in (2.18) the lower bound ϵ_0 will be called "the minimalizing analytic function" and will be denoted by $\chi_1^0(\zeta)$ [in the previous example we had $\epsilon_0=1$ and $\chi_1^0(\zeta)\equiv 0$]. We should also like to stress that unlike ϵ, the quantity ϵ_0 does not depend only on the accuracy of the experiment, but on the experiment itself, ϵ_0 being a functional of $\tilde{h}_2(\zeta)$. Of course, in order to have at least one analytic function satisfying (2.7), ϵ has to be greater than ϵ_0,

$$\epsilon \geqslant \epsilon_0[\tilde{h}_2]. \tag{2.19}$$

Unless we are in the exceptional case $\epsilon=\epsilon_0$, the "minimal amplitude", constructed with the minimalizing function χ_1^0,

$$\tilde{f}_0=\tilde{h}_1-\chi_1^0 \tag{2.20}$$

is not the best approximation ["$\hat{\tilde{f}}(\zeta)$"] of all analytic functions in (D) satisfying (2.7): Indeed, if ϵ is strictly greater than ϵ_0, the set of the admissible function $\tilde{f}(\zeta)$ for every $\zeta\in\bar{D}$ fills densely a disk [the boldface circle in Figs. 3 and 4, corresponding, respectively, to the cases when ζ is a boundary and an interior point of (D)]. Obviously the center $\hat{\tilde{f}}(\zeta)$ of the circle represents for each ζ the best approximation, but as it will be proved in Sec. 4, this center *does not* coincide with the $f_0(\zeta)$.

The reader, making the intuitive assumption that the histogram is the limit of an analytic function (the amplitude itself), might get the feeling that all this trouble and soul searching is due only to the inaccuracy of the experiment which would produce a (small) nonanalytic term $\tilde{h}_2(\zeta)$. Contrary to the common belief, $\tilde{h}_2(\zeta)$ is by no means small (regardless of the experimental accuracy), being the direct product of a well-known theorem of Fourier decomposition applied to functions

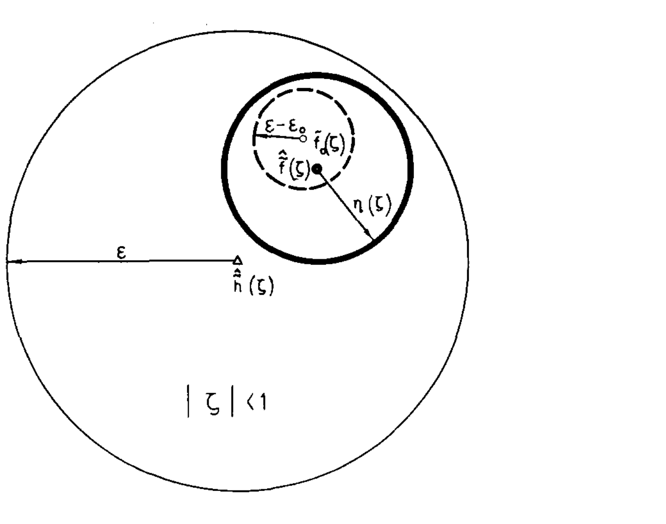

FIG. 4. Typical situation in an interior point $|\zeta|<1$. The values of all weighted holomorphic functions $\tilde{f}(\zeta)$ compatible with conditions (1.1) and (1.2) fill the boldface circle of radius $\eta(\zeta)$ around $\hat{\tilde{f}}(\zeta)$. There are no admissible amplitudes outside it, so that $\hat{\tilde{f}}(\zeta)$ is the best estimate for a random-taken amplitude. The dashed circle of radius $\epsilon-\epsilon_0$ around $\tilde{f}_0(\zeta)$ is still contained in the latter one, so that $\epsilon-\epsilon_0<\eta(\zeta)<\epsilon$. Especially for ϵ_0 close to ϵ, $\hat{\tilde{f}}(\zeta)$ may differ considerably from $\hat{h}(\zeta)$.

which are identically zero over a segment (Γ_2) of the boundary. (If a Fourier expandable function is identically zero over some segment of its period, the negative and positive frequencies are simultaneously present.) In other words $\tilde{h}_2(\zeta)$ is the consequence of our incomplete knowledge, reflecting our complete lack of knowledge on Γ_2 rather than the lack of accuracy on Γ_1!

Previous methods of extrapolation variously took into account this nonholomorphic component. For instance, the Carleman weighted dispersion relations,[6] written with a conventional Cauchy kernel,

$$\check{h}(\zeta)=\frac{1}{2\pi i}\int_\Gamma \frac{\tilde{h}(\zeta')}{\zeta'-\zeta}\,d\zeta', \tag{2.21a}$$

are completely insensitive to the presence of $\tilde{h}_2$, as

$$\check{h}(\zeta)\equiv\tilde{h}_1(\zeta),\quad \check{h}_2(\zeta)\equiv 0. \tag{2.21b}$$

On the other hand, the Poisson kernel used in Refs. 7 and 8 transforms $\tilde{h}_2(\zeta)$ into the complex valued harmonic (but not holomorphic!) function

$$\hat{\tilde{h}}_2(\zeta)=\sum_{p=1}^{\infty} c_{-p}\zeta^{*p}\quad (\zeta^*\text{ being the complex conjugate of }\zeta), \tag{2.22}$$

so that the extrapolated function[7,8]

$$\hat{\tilde{h}}(\zeta)=\frac{1}{2\pi i}\int \tilde{h}(\zeta')d(G(\zeta,\zeta')+iH(\zeta,\zeta'))\equiv\tilde{h}_1(\zeta)+\hat{\tilde{h}}_2(\zeta) \tag{2.23}$$

provides a "hundred per cent approximation" of the histogram on Γ. Indeed, owing to the limiting properties of harmonic functions on the boundary, we have on Γ

$$|\hat{\tilde{h}}(\zeta)-\tilde{h}(\zeta)|_{\zeta\in\Gamma}=0, \tag{2.24}$$

but inside the unit circle the harmonic function $\hat{\tilde{h}}(\zeta)$ could differ considerably (of course, less than ϵ!) from the values of the admissible holomorphic functions $f(\zeta)$. This is especially apparent from the previous example: If in Eq. (2.14) one sets ϵ equal to 1, the unique analytic function satisfying (2.14) is the minimalizing function $\chi_1^0(\zeta)\equiv 0$, i.e., the amplitude coincides both with the unique admissible function $\tilde{f}_0(\zeta)$ (the minimal amplitude) and with the Cauchy weighted integral $\tilde{h}_1(\zeta)$, while the corresponding Poisson extrapolation $\hat{\tilde{h}}=\tilde{h}_1(\zeta)+\zeta^*$ represents, with the exception of the origin $\zeta=\zeta^*=0$, a worse approximation. The coincidence between the minimal amplitude $\tilde{f}_0(\zeta)$ and the weighted Cauchy dispersion integral $\tilde{h}_1(\zeta)$ which occured in this example is purely incidental, as, in general, $\tilde{f}_0(\zeta)$, $\hat{\tilde{h}}(\zeta)$ and, especially, $\tilde{h}_1(\zeta)$ could differ considerably among them. This is especially apparent from the well-known example of the step function

$$\tilde{h}(e^{i\theta})=\begin{matrix}0 & \text{for } -\pi<\theta<0\\ 1 & \text{for } 0<\theta<\pi\end{matrix}$$
$$=1+\frac{2}{\pi}\sum_{p=1}^{\infty}\frac{\sin(2p-1)\theta}{2p-1}, \tag{2.25}$$

which also emphasizes the importance of the $\tilde{h}_2$ term, as the corresponding

$$\tilde{h}_1(\zeta)=1-i\sum_{p=1}^{\infty}\frac{\zeta^{2p-1}}{(2p-1)} \tag{2.26}$$

becomes extremely great in the neighborhoods of the points $\zeta=1$ and $\zeta=-1$, while the sum $\tilde{h}_1+\hat{\tilde{h}}_2$—and, hence, also the weighted Poisson integral $\hat{\tilde{h}}(\zeta)=\tilde{h}_1(\zeta)+\hat{\tilde{h}}_2(\zeta)$, as well as the minimal amplitude $f_0(\zeta)$ and all the other admissible amplitudes—remain finite.

If, as in the step-function example, the Carleman weighted Cauchy dispersion integral could differ in an incontrollable manner from the true amplitude, the Carleman weighted Poisson dispersion relation always secures the error bound ϵ prescribed by the maximum of the modulus principle—or, if one comes back to the unweighted (real) amplitude $f(\zeta)$, (2.2), the Poisson extrapolation

$$\hat{h}(\zeta)\overset{\text{DEF}}{=}\hat{\tilde{h}}(\zeta)/C_0(M/\epsilon;\zeta) \tag{2.27}$$

secures the "Nevanlina bound"

$$|\hat{h}(\zeta)-f(\zeta)|_{\zeta\in\bar{D}}<\epsilon/C_0(\zeta)\equiv\epsilon\exp[\ln(M/\epsilon)\omega(\zeta)]$$
$$=\epsilon^{1-\omega(\zeta)}M^{\omega(\zeta)}. \tag{2.28}$$

As we have already stated above and as it will be proved in Sec. 4, all admissible weighted amplitudes $\tilde{f}(\zeta)$ fill densely, for every ζ, a disk of radius $\eta(\zeta)$, contained, of course, in the circle of radius ϵ centered around $\hat{\tilde{h}}(\zeta)$ (see Figs. 3 and 4).

Without going into the details of Sec. 4, we can immediately show that the circle of the admissible weighted amplitudes contains a circle of (constant) radius $\epsilon-\epsilon_0$, centered around the minimal amplitude $\tilde{f}_0(\zeta)$; for each function

$$\tilde{f}_{\alpha,\varphi}(\zeta)=\tilde{f}_0(\zeta)+\alpha(\epsilon-\epsilon_0)e^{i\varphi}\quad (0\leqslant\alpha<1,\ 0\leqslant\varphi<2\pi), \tag{2.29}$$

where α, φ are constants, is holomorfic inside D and satisfies condition (2.7) (the dashed line circles of Figs. 3 and 4). This provides us with a simple criterion to decide which method to use in some precise practical application. For that we need the numerical value only of ϵ_0 which will be computed in Sec. 3 [see Eq. (3.13)]:

(1) If $\epsilon_0\ll\epsilon$ so that the dashed circle of radius $\epsilon-\epsilon_0$ [and thus, *a fortiori*, the bold face circle of radius $\eta(\zeta)$, $\epsilon-\epsilon_0<\eta<\epsilon$] fills most of the circle of radius ϵ, the function $\hat{\tilde{h}}(\zeta)$ represents fairly well the middle point $\bar{\tilde{f}}(\zeta)$ of all possible weighted amplitudes $\tilde{f}(\zeta)$ (the center of the bold face circle), and thus the Carleman weighted dispersion relation [see Eqs. (2.23), (2.27) and Ref. 7] would give very satisfactory results.

(2) If the accuracy of the data along the initial physical region Γ_1 is high enough so that ϵ is small and only slightly greater than ϵ_0, it could happen that all admissible amplitudes would pile up in a very small region containing the minimal weighted amplitude as in Eq. (2.14). Therefore, if the radius $\epsilon-\epsilon_0$ of the dashed circle is small in comparison to ϵ, one has to compute the radius $\eta(\zeta)$ [Eq. (4.20)] and, if it differs also considerably from ϵ, find the center $\bar{\tilde{f}}(\zeta)$ of the circle of all the weighted admissible amplitudes. Of course, in both

cases the optimal approximation for the entire set of amplitudes is given by the nonanalytic function:

$$\hat{f}(\zeta)=\hat{\hat{f}}(\zeta)/C_0(M/\epsilon;\zeta). \tag{2.30}$$

In Sec. 3 we shall deal with the problem of the minimal weighted amplitude $\tilde{f}_0(\zeta)$ and the computation of the constant ϵ_0. In Sec. 4 we shall then write down all possible amplitudes $f(\zeta)$ compatible with a given ϵ, as well as the explicit form of the (nonanalytic) function $\hat{f}(\zeta)$ giving the value of the center of the set of all the admissible amplitudes, as well as the value of the radius $\eta(\zeta)/C_0(M/\epsilon;\zeta)$ of this set. So far, the width of the error channel ϵ was kept constant: The physically important variable-error case is discussed in the concluding Sec. 5, where an out line of the L^2-norm problems is also given, together with a discussion about the two minimal amplitudes, $f_{\epsilon_0}(\zeta)$ (which approximates the best the histogram on Γ_1 *at given* M) and $f_{M_0}(\zeta)$ (the amplitude with least modulus M_0 on Γ_2, *at given* ϵ). Both extremal amplitudes, $f_{\epsilon_0}(\zeta)$ and $f_{M_0}(\zeta)$, as well as $f_0(\zeta)$, are contained in the circle of radius $\eta(\zeta)/C_0(M/\epsilon;\zeta)$ around the optimal function $\hat{f}(\zeta)$.

III. COMPUTATION OF $\epsilon_0[h;M/\epsilon]$

One of the problems described in Sec. 2 we shall now have to deal with is the construction of the holomorphic function $-\chi_1^0(\zeta)$, which approximates on the unit circle Γ, in the best way, the nonholomorphic part $\tilde{h}_2(\zeta)$ of the weighted histogram, i.e.,

$$\max_{\substack{\zeta\in\Gamma\\ \chi_1\to\chi_1^0}} |\chi_1(\zeta)+\tilde{h}_2(\zeta)|\to\min=\epsilon_0. \tag{3.1}$$

This is an important problem which has been under scrutiny for a long while by mathematicians, although it is virtually solved by the Schwartz lemma in the form of Pick and Schur,[12] which is nothing but a special case of the Lindelöf principle. Nevertheless, the last-time powerful function-analytical methods (Nehari,[13] and Krein[14]) have had a great impact on this problem, and in the present section we shall show how one can compute the constant ϵ_0 from Eq. (3.1) using the Foiaş–Nagy lifting theorem.[15] More precisely, we shall give here an outline of the proofs which can be found in their full extent in Appendix A. As the construction of the function $\chi_1^0(\zeta)$, once the numerical value of ϵ_0 is known, is quite similar to that of all other $\chi_1(\zeta)$'s satisfying (2.17)–(2.19), this second problem will be postponed until the next section.

Let the $\chi(e^{i\theta})$ be different functions, defined on the unit circle Γ, whose *negative-frequency* Fourier coefficients coincide with the *negative-frequency* coefficients of the weighted histogram $\tilde{h}(e^{i\theta})$:

$$Q_-\chi(e^{i\theta})=Q_-h(e^{i\theta}), \tag{3.2a}$$

where Q_- is the projection operator on the space $L^2\ominus H^2$ of functions with negative frequencies only. We shall write similarly to (2.16),

$$\chi(e^{i\theta})=\chi_1(e^{i\theta})+\chi_2(e^{i\theta}), \tag{3.3a}$$

with

$$\chi_1(e^{i\theta})=(1-Q_-)\chi(e^{i\theta}) \quad\text{and}\quad \chi_2(e^{i\theta})=Q_-\chi(e^{i\theta}), \tag{3.3b}$$

so that (3.2a) is nothing but

$$\chi_2(e^{i\theta})\equiv\tilde{h}_2(e^{i\theta}). \tag{3.2b}$$

Let us now remark that, as the L^∞ norm of a function,

$$\|\chi(\zeta)\|_{L^\infty}\equiv\operatorname*{ess.\,sup}_{\zeta\in\Gamma}|\chi(\zeta)|, \tag{3.4}$$

is always greater or equal to its L^2 norm, we have

$$\begin{aligned}\epsilon_0=\|\chi^0(\zeta)\|_{L^\infty}&\geq\|\chi^0(\zeta)\|_{L^2}\equiv\|\chi_1^0(\zeta)\|_{L^2}+\|\chi_2^0(\zeta)\|_{L^2}\\ &\geq\|\chi_2^0(\zeta)\|_{L^2}.\end{aligned} \tag{3.5}$$

Owing to (3.2), we get

$$\epsilon_0\geq\|\tilde{h}_2(\zeta)\|_{L^2}\equiv\left(\sum_{-1}^{-\infty}|c_n|^2\right)^{-1/2}, \tag{3.6}$$

which, by the way, proves the existence of ϵ_0 as a positive, not vanishing constant.

Now, as it is less easy to handle the L^∞ norm than the L^2 one, following C. Foiaş, one can transpose the whole problem into an L^2-norm problem for the suitable chosen *operators* Y_χ closely connected to the functions $\chi(e^{i\theta})$. Namely, if $\varphi(e^{i\theta})$ is some general L^2 function defined on Γ, we define

$$Y_\chi\varphi(e^{i\theta})\equiv\check{\chi}(e^{i\theta})^*U^+\check{\varphi}(e^{i\theta}), \tag{3.7a}$$

where we have denoted by $\check{\varphi}$ the "θ-reflected" function

$$\check{\varphi}(e^{i\theta})\equiv\varphi(e^{-i\theta}), \tag{3.7b}$$

and by U the multiplication operator with the factor $e^{i\theta}$:

$$\begin{aligned}U\varphi'(e^{i\theta})&=e^{i\theta}\varphi'(e^{i\theta}),\\ U^+\varphi'(e^{i\theta})&=e^{-i\theta}\varphi'(e^{i\theta}).\end{aligned} \tag{3.7c}$$

As neither of the two operations (3.7b) or (3.7c) change the norms, it is clear that the L^2 norm of *the operator* Y_χ coincides with the maximum of the function $\chi(e^{i\theta})$ on Γ, i.e.,

$$\|Y_\chi\|_{L^2}\equiv\|\chi(e^{i\theta})\|_{L^\infty}; \tag{3.8}$$

The additional unit-norm operators contained in Y_χ apart from the function $\chi(e^{i\theta})$ were taken precisely to secure the commutation relation

$$Y_\chi U^+=UY_\chi, \tag{3.9}$$

which is essential for the use of the Foiaş–Nagy lifting theorem (see Appendix A).

The crucial points of the proof (see Appendix A) are now the following two:

Firstly, the special form (3.7a) of Y_χ enables us to show that the restriction X of the operator[16] $Q_-Y_\chi^+$ on the subspace $L^2\ominus H^2$ of the negative frequency functions $\varphi_2=Q_-\varphi$,

$$X\varphi_2=Q_-Y_\chi^+\varphi_2=Q_-\chi(e^{i\theta})U^+\check{\varphi}_2, \tag{3.10}$$

depends solely [see further Eq. (3.15)] *on the negative frequency part* $\chi_2(e^{i\theta})$ of the function $\chi(e^{i\theta})$; its L^2 norm—which, as we shall show, coincides with ϵ_0 and which, as X itself, is also completely determined by the Fourier coefficients $c_{-1},c_{-2},\ldots$ of $\tilde{h}_2(e^{i\theta})$—is obviously[17] smaller than the L^2 norms of all the operators Y_χ corresponding to functions $\chi(e^{i\theta})$ with $\chi_2\equiv\tilde{h}_2$:

$$\|X\|_{L^2}\leq\|Y_\chi\|_{L^2}. \tag{3.11}$$

The second important point is that, using Foiaş–Nagy lifting theorem,[15,18,19] we can show *that there exists a* *function* $\chi = \chi^0(e^{i\theta})$ *satisfying* (3.2), *for which equality is reached in the inequality* (3.11); this enables us to define ϵ_0 as the norm of X:

$$\epsilon_0 \equiv \|X\|_{L^2} = \|Y_{\chi^0}\|_{L^2} \equiv \|\chi^0(e^{i\theta})\|_{L^\infty}. \tag{3.12a}$$

Indeed, combining (3.11) with (3.8), we get

$$\epsilon_0 < \text{all other } \|\chi(e^{i\theta})\|_{L^\infty} \text{ with } \chi_2(e^{i\theta}) \equiv \tilde{h}_2(e^{i\theta}), \tag{3.12b}$$

which corresponds to the previous definition (3.1) of ϵ_0 as being the maximum of the error produced by the best holomorphic approximant $-\chi_1^0$ of $\tilde{h}_2$ on Γ. The Foiaş–Nagy lifting theorem which made this point possible asserts, indeed, that if we have two isometric operators T and T' operating in the spaces K and K', respectively, and if S and S' are the *restrictions of their adjoints operators* T^+ and T'^+ on the *invariant* subspaces $H \subseteq K$ and $H' \subseteq K'$, and if the operator X transforms the subspace H into H' and satisfies the commutation relation $XS = S'X$, then there exists an extension Y_0 of X transforming K into K' and satisfying $Y_0 T^+ = T'^+ Y_0$, *whose norm coincides with that of* X. Now, to prove (3.12a), one has to take $T = U$ and $T' = U^+$ [see Eq. (3.9)] and apply the Foiaş–Nagy theorem twice (see Appendix A), once with $K = L^2$ and $H = H' = K' = L^2 \ominus H^2$ and once with $H = L^2 \ominus H^2$ and $K = H' = K' = L^2$. Further, one defines $\check{\chi}^0(e^{i\theta})^*$ to be that function one would get if one applied the minimal-norm operator Y_0 to the constant function 1 (and multiplied it by $e^{+i\theta}$):

$$\check{\chi}^0(e^{i\theta})^* = e^{i\theta} Y_0 \mathbf{1} \equiv U Y_0 \mathbf{1}.$$

Reciprocally, one can express then the minimal operator Y_0 in terms of χ^0 in the way of Eq. (3.7a). Indeed, if $\varphi(e^{i\theta})$ is any L^2 function defined on Γ,

$$\varphi(e^{i\theta}) = \sum_{-\infty}^{+\infty} a_n e^{in\theta} = \sum_{-\infty}^{+\infty} a_{-n} (U^+)^n \mathbf{1},$$

owing to (3.9):

$$Y_0 \varphi = \sum_{-\infty}^{+\infty} a_{-n} Y_0 (U^+)^n \mathbf{1} = \sum_{-\infty}^{+\infty} a_{-n} U^n Y_0 \mathbf{1} = \sum_{-\infty}^{+\infty} a_{-n} U^n U^{-1} \check{\chi}^{0*}$$
$$\equiv \check{\chi}^{0*} U^+ \check{\varphi}.$$

Hence,

$$\|\chi^0\|_{L^\infty} = \|Y_0\|_{L^2} = \|X\|_{L^2} \leq \|Y_\chi\|_{L^2} \equiv \|\chi\|_{L^\infty}$$

and the proof of relations (3.12) is now complete.

Once the identity between the L^∞ norm of the optimal function χ^0 and the L^2 norm of the operator X has been established, one can evaluate ϵ_0 numerically as the square root of the greatest eigenvalue of the Hermitian operator XX^+, that is,

$$\epsilon_0 = \lim_{N' \to \infty} (\mathrm{Tr}[(XX^+)^{N'}])^{1/2N'}. \tag{3.13}$$

A matrix representation for X to be used in (3.13) can readily be found in the basis spanned by the eigenvectors $e^{-i\theta}, e^{-2i\theta}, e^{-3i\theta}, \cdots$ of the $L^2 \ominus H^2$ subspace. Indeed, by taking $\varphi_2 = e^{-ik\theta}$ from (3.10), it follows that

$$X e^{-ik\theta} = Q_- \chi(e^{i\theta}) U^+ e^{+ik\theta} = Q_- \sum_{n=-\infty}^{n=\infty} c_n \exp[i(n+k-1)\theta], \tag{3.14}$$

so that, putting $|k\rangle \equiv \exp(-ik\theta)$, $k = 1, 2, 3, \cdots$, we have

$$\langle j | X | k \rangle = c_{-(k+j-1)}. \tag{3.15}$$

Thus,

$$\mathbf{X} = \begin{Bmatrix} c_{-1} & c_{-2} & c_{-3} & \cdot \\ c_{-2} & c_{-3} & \cdot & \cdot \\ c_{-3} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Bmatrix}, \tag{3.16a}$$

where $c_{-1}, c_{-2}, \cdots$, are the negative frequency Fourier coefficients (2.16d) of the weighted histogram. (Such a matrix is usually called a *Hankel-matrix*). In practical calculations one could set to zero in (3.16) all c_{-n} with n greater than a certain N, sufficiently so that

$$\tilde{h}_2^{(N)}(e^{i\theta}) = \sum_{n=1}^{N} c_{-n} \exp(-in\theta) \tag{3.17a}$$

would approximate sufficiently well $\tilde{h}_2(e^{i\theta})$; usually a fairly good approximation of ϵ_0 is reached with (3.12) in few N'-steps at a computer. However, one could spare computer loading [by smaller (3.16) matrices], using in (3.16) a quicker converging trigonometric series (for instance, the Fejer series $c'_{-k} = (1 - 1/k) c_{-k}$ or, better yet, the Chebysheff approximation), instead of the Fourier coefficients c_{-n}. Indeed, if

$$\tilde{h}_2'^{(N)}(e^{i\theta}) = \sum_{n=1}^{N} c'_{-n} \exp(-in\theta), \tag{3.17b}$$

with

$$\sup_{0 \leq \theta < 2\pi} |\tilde{h}_2(e^{i\theta}) - \tilde{h}_2'^{(N)}(e^{i\theta})| = \eta^{(N)}, \tag{3.18}$$

[for a given N, $\eta^{(N)}$ is the smallest when (3.17b) is the Chebysheff approximation of order N to $\tilde{h}_2$], we have quick estimate of the accuracy with which $\epsilon_0^{(N)}$, computed with the help of the truncated matrix

$$X^{(N)} = \begin{Bmatrix} c'_{-1} & c'_{-2} & \cdots & c'_{-N} \\ c'_{-2} & c'_{-3} & \cdots & 0 \\ \vdots & & & \\ c'_{-N} & 0 & \cdots & 0 \end{Bmatrix}, \tag{3.16b}$$

represents the true ϵ_0. Indeed, from the previous theory it follows that if there exists a holomorphic function $-\chi_1^{0(N)}$ which approximates in the best way on Γ the negative frequency function $\tilde{h}_2'^{(N)}$, we have

$$\epsilon_0^{(N)} \equiv |\chi_1^{0(N)} + \tilde{h}_2'^{(N)}| \leq \sup |\chi_1^0 + \tilde{h}_2 + \tilde{h}_2'^{(N)} - \tilde{h}_2| \leq \epsilon_0 + \eta^{(N)}, \tag{3.19}$$

as well as

$$\epsilon_0 \equiv |\chi_1^0 + \tilde{h}_2| < \sup |\chi_1^{0(N)} + \tilde{h}_2'^{(N)} + \tilde{h}_2 - \tilde{h}_2'^{(N)}| \leq \epsilon_0^{(N)} + \eta^{(N)}.$$

Hence,

$$\epsilon_0^{(N)} - \eta^{(N)} < \epsilon_0 < \epsilon_0^{(N)} + \eta^{(N)}, \tag{3.20}$$

so that, if the series (3.17b) approximates fairly well $\tilde{h}_2$, the constant $\epsilon_0^{(N)}$ computed via (3.13) and (3.16b) represents well the actual value of ϵ_0.

From (2.16d) and from (3.16) it is clear that ϵ_0 is a functional of the histogram $h(\xi)$ and a function of the ratio M/ϵ:

$$\epsilon_0[\tilde{h}] \equiv \epsilon_0[h; M/\epsilon]. \tag{3.21}$$

In general, ϵ_0 *is not* the smallest ϵ for which there still exist holomorphic functions bounded by M on Γ_2 (1.2) and satisfying (1.1). Nevertheless, $\epsilon_0[h; M/\epsilon]$ being the smallest possible deviation *on* $\Gamma_1+\Gamma_2$ of an analytic function $\tilde{f}_0(\zeta)$ *from the* M/ϵ Carleman *weighted histogram* $\tilde{h}(\zeta)$, we obviously have

$$\epsilon_{k+1} \overset{\mathrm{DEF}}{=} \epsilon_0[h;M/\epsilon_k] < \epsilon_k, \tag{3.22}$$

so that the decreasing series $\epsilon_1 > \epsilon_2 > \epsilon_3 > \cdots$ defines a minimal ϵ,

$$\epsilon_{00} = \lim_{n\to\infty} \epsilon_n \quad (>0), \tag{3.23}$$

satisfying the transcedent equation

$$\epsilon_0[h;M/\epsilon_{00}] = \epsilon_{00}. \tag{3.24}$$

The corresponding $\tilde{f}_0(\zeta)$ amplitude

$$|\tilde{f}_{00}(\zeta) - h(\zeta)C_0(M/\epsilon_{00};\zeta)|_\Gamma = \epsilon_{00} \tag{3.25a}$$

defines the minimal function

$$f_{\epsilon_0}(\zeta) = \tilde{f}_{00}(\zeta)/C_0(M/\epsilon_{00};\zeta),$$
$$|f_{\epsilon_0}(\zeta)|_{\Gamma_2} < M, \tag{3.25b}$$
$$|f_{\epsilon_0}(\zeta) - h(\zeta)|_{\Gamma_1} \le \epsilon_{00},$$

the smallest value of ϵ for which such an analytic function still exists. The value ϵ_{00} can be found either as the limit (3.23), computing step-by-step ϵ_k (3.22), or directly, solving the trancedent equation on a computer, combining (3.24) with (2.16d), (3.13), and (3.16).

The second external problem consists in finding the amplitude $f_{M_0}(\zeta)$ of least module $M=M_0$ on Γ_2, ϵ being now fixed. As in (3.24), one would first have to determine M_0 from the equation [for a proof see (5.34)]

$$\epsilon_0[h;M_0/\epsilon] = \epsilon \quad (\epsilon \text{ being given}) \tag{3.26}$$

and then, using the methods of the next section, build the function $f_0(\zeta)$ corresponding to M_0 and ϵ:

$$f_{M_0}(\zeta) = \tilde{f}_0(\zeta)/C_0(M_0/\epsilon;\zeta). \tag{3.27}$$

We shall come back to this problem in Sec. 5.

IV. CONSTRUCTION OF THE SET OF ALL ADMISSIBLE AMPLITUDES

Once we have computed the constant ϵ_0, we turn back to the effective construction of all analytic functions $\tilde{f}(\zeta)$ satisfying (2.7):

$$|\tilde{f}(\zeta) - \tilde{h}(\zeta)|_{\Gamma_1+\Gamma_2} < \epsilon. \tag{4.1}$$

Of course, (2.19), ϵ has to be greater than ϵ_0, which is the smallest value of ϵ for which the set of admissible functions is not void; all the extremal functions $\tilde{f}_0(\zeta$ $\tilde{f}_0(\zeta), \tilde{f}_{\epsilon_0}(\zeta)$, and $\tilde{f}_{M_0}(\zeta)$ are then constructed in a similar way to all other $\tilde{f}(\zeta)$ [there is still a difference, as will be seen, namely, that the function $\psi_N(\,)$ for $f_0(\zeta)$ reduces to zero, but this happens in an automatic way if ϵ is set equal to ϵ_0], by simply replacing ϵ by the corresponding ϵ_0 [by ϵ_{00} if the Carleman weight was $C_0(M/\epsilon_{00},\zeta)$, or by ϵ if one has used $C_0(M_0/\epsilon,\zeta)$].

According to (2.13) each admissible function $\tilde{f}(\zeta)$ contains besides the holomorphic part $\tilde{h}_1(\zeta)$ of the weighted amplitude,

$$\tilde{h}_1(\zeta) = \frac{1}{2\pi i}\int_{\Gamma_1+\Gamma_2} \frac{\tilde{h}(\zeta')}{\zeta'-\zeta}\, d\zeta' \quad (\zeta\in D), \tag{4.2}$$

a supplementary holomorphic part, also, $-\chi_1(\zeta)$, which (2.17) approximates the negative frequency part $\tilde{h}_2(\zeta)$ of the weighted histogram

$$\tilde{f}(\zeta) = \tilde{h}_1(\zeta) - \chi_1(\zeta),$$
$$|\chi_1(\zeta) + \tilde{h}_2(\zeta)|_{\zeta\in\Gamma} \le \epsilon. \tag{4.3}$$

[In contradistinction with the L^2 norm problem (see Sec. 5), where, owing to the orthogonality of the positive and negative Fourier components, one *cannot* approximate $\tilde{h}_2$ by analytic functions in D.] In what follows we shall suppose for the sake of simplicity that the nonholomorphic part $\tilde{h}_2(e^{i\theta})$ of the weighted histogram contains only a finite number, N, of negative frequency Fourier coefficients ($\tilde{h}_2 = \tilde{h}_2^{(N)}$ so that the function

$$\begin{aligned}\psi_0(\zeta) &= \frac{\zeta^N}{\epsilon}(\chi_1(\zeta) + \tilde{h}_2^{(N)}(\zeta)) \equiv \frac{\zeta^N}{\epsilon}\chi(\zeta)\\ &= \frac{c_{-N}}{\epsilon} + \frac{c_{-N+1}}{\epsilon}\zeta + \cdots + \frac{c_{-1}}{\epsilon}\zeta^{N-1}\\ &\quad + \frac{\zeta^N}{\epsilon}\chi_1(\zeta)\end{aligned} \tag{4.4}$$

is *holomorphic* inside D. Moreover, according to (4.3) we have

$$|\psi_0(\zeta)| \le 1 \tag{4.5}$$

for every $|\zeta| \le 1$. Finding all amplitudes satisfying (4.1) reduces thus to finding all holomorphic functions satifying (4.5), and having N preassigned Taylor coefficients:

$$\psi_{0,0} = c_{-N}/\epsilon, \quad \psi_{0,1} = c_{-(N-1)}/\epsilon, \quad \cdots\ \psi_{0,N-1} = c_{-1}/\epsilon. \tag{4.6}$$

This problem can be solved in a simple recurrent way. Indeed, if the unity-bounded function $\psi_{k-1}(\zeta)$,

$$|\psi_{k-1}(\zeta)| \le 1, \quad |\zeta| \le 1, \tag{4.7}$$

has $N-(k-1)$ preassigned Taylor coefficients

$$\frac{1}{k!}\left.\frac{d^j\psi_{k-1}(\zeta)}{d\zeta^j}\right|_{\zeta=0} = \psi_{k-1,j} \quad (\text{given for all } 0\le j\le N-k), \tag{4.8}$$

where

$$|\psi_{k-1,0}| \lneq 1, \tag{4.9}$$

the function

$$\psi_k(\zeta) = \frac{1}{\zeta}\,\frac{\psi_{k-1}(\zeta) - \psi_{k-1,0}}{1-\psi_{k-1}(\zeta)\psi^*_{k-1,0}} \tag{4.10}$$

is also unity-bounded,

$$|\psi_k(\zeta)| \le 1, \quad |\zeta| \le 1, \tag{4.11}$$

analytic in D, and its $N-k$ first Taylor coefficients are completely determined by the first $N-k+1$ coefficients (4.7) of $\psi_{k-1}(\zeta)$. (The reverse statement being also valid):

$$\psi_{k,n} = \sum \frac{(\sum_{i=1}^{n+1}k_i)!\,\psi^{*(\sum_1^{n+1}k_i-1)}_{k-1,0}}{k_1!\ldots k_{n+1}!(1-|\psi_{k-1,0}|^2)}\psi^{k_1}_{k-1,1}\cdots\psi^{k_{n+1}}_{k-1,n+1}, \tag{4.12}$$

where the sum $\sum$ is extended to all combinations of nonnegative integer k_i with

$$k_1 + 2k_2 + \cdots + (n+1)k_{n+1} = n+1. \tag{4.13}$$

The inequality (4.11) does not assure that $\psi_{k,0}$ satisfies the inequality (4.9) too, but this really happens for $\epsilon > \epsilon_0$.

Going further, one finally gets a holomorphic function $\psi_N(\zeta)$, which, beside the inequality

$$|\psi_N(\zeta)|<1, \quad |\zeta|\le 1, \tag{4.14}$$

is completely free of supplementary conditions.

Hence, using the inverse of (4.10),

$$\psi_{k-1}(\zeta)=\frac{\zeta\psi_k(\zeta)+\psi_{k-1,0}}{1+\psi^*_{k-1,0}\zeta\psi_k(\zeta)}, \tag{4.15}$$

one finally gets the whole family of functions $\psi_0(\zeta)$ satisfying (4.4) and (4.5), in terms of a general, unity-bounded function $\psi_N(\zeta)$, and, of course, depending on the negative Fourier coefficients $c_{-1},\ldots, c_{-N}$, (2.16d), of the weighted histogram [via Eqs. (4.6), (4.12), and (4.15)]. Once one knows the functions $\psi_{0[\psi_N]}(\zeta)$, the most general holomorphic function satisfying the conditions (1.1) and (1.2) is given in the approximation

$$\tilde{h}_2(\zeta)\approx\tilde{h}_2^{(N)}(\zeta)=\sum_{k=1}^{N}c_{-k}\zeta^{-k} \tag{4.16}$$

by

$$f_{[\psi_N]}(\zeta)=[\tilde{h}_1(\zeta)+\tilde{h}_2^{(N)}(\zeta)-(\epsilon/\zeta^N)\psi_{0[\psi_N]}(\zeta)]/C_0(M/\epsilon,\zeta), \tag{4.17}$$

where $\tilde{h}_1$ is given by (4.2) [see also (2.1), (2.5), and (2.16d)]. Of course, the negative powers of ζ cancel out identically, owing to Eq. (4.4).

If in Eq. (4.4) one puts $\epsilon=\epsilon_0[h;M/\epsilon]$, at least the last $(k=N-1)$ of the constants $\psi_{k,0}$ has to be unimodular. If this were not the case, one would find a condition-free N unity-bounded iterate $\psi_N(\zeta)$, and this would be possible by continuity also for an ϵ slightly smaller than ϵ_0, when, by definition there are no more solutions. Thus, if $\psi_{k,0}$ infringes inequality (4.9),

$$|\psi_{k,0}|=1 \tag{4.18a}$$

from (4.11) and from the principle of the maximum of the modulus $[\psi_{k,0}=\psi_k(0)!]$, one gets

$$\psi_k(\zeta)\equiv\psi_{k,0}\quad(\epsilon=\epsilon_0), \tag{4.18b}$$

and the extremal weighted function $\tilde{f}_0(\zeta)$ can then be built again in a recursive (4.15) way; but, starting from the last not vanishing function $\psi_k(\zeta)$, given by (4.18b), rather than from $\psi_N(\zeta)$,

$$f_0(\zeta)=\tilde{h}_1(\zeta)+\tilde{h}_2^N(\zeta)-(\epsilon_0/\zeta^N)\psi_0(\zeta)C_0(M/\epsilon_0,\zeta), \tag{4.19}$$

where in, contradistinction to (4.17), $\psi_0(\zeta)$ is completely determined in terms of the Fourier coefficients c_{-n} $(1\le n\le N)$ of $\tilde{h}_2(\zeta)$. Unless $\epsilon=\epsilon_{00}$ [and hence $f_0(\zeta)=f_{\epsilon_0}(\zeta)$] or $M=M_0$ [see (3.26): $f_0(\zeta)=f_{M_0}(\zeta)$], the amplitudes $f_0(\zeta)$ have no special physical meaning; nevertheless, one should notice that for every weighted function $\tilde{f}_0(\zeta)$ one has the interesting property

$$|\tilde{f}_0(\zeta)-\tilde{h}(\zeta)|_{\Gamma_1+\Gamma_2}=\epsilon_0[h,M/\epsilon]\,(=\text{const}), \tag{4.20a}$$

and hence

$$\begin{aligned}&|f_0(\zeta)-h(\zeta)|_{\Gamma_1}=\epsilon_0,\\&|f_0(\zeta)|_{\Gamma_2}=M\epsilon_0/\epsilon.\end{aligned} \tag{4.20b}$$

Equation (4.20a) is a consequence of the fact that if $\psi_k(\zeta)=\psi_{k,0}$, the modules of all the functions $\psi_j(\zeta)$, $j\le k$, are equal to one for the boundary points $|\zeta|=1$ and thus, owing to Eq. (4.4), with ϵ replaced by ϵ_0,

$$|\tilde{f}_0(\zeta)-\tilde{h}(\zeta)|_{\zeta\in\Gamma}=\epsilon_0|\psi_0(\zeta)/\zeta^N|_{\zeta\in\Gamma}=\epsilon_0,$$

which proves (4.20a).

In practical extrapolation problems it is perhaps more important to know the value $\hat{f}(\zeta)$ and $\hat{\eta}(\zeta)=\eta(\zeta)/C_0(\zeta)$ of the center and of the radius of the set of values of all the admissible functions $f_{[\psi_N]}(\zeta)$ in a given point ζ, rather than the admissible functions themselves. One can readily prove that, in every point ζ, this set of values fills densely a circle. Indeed, if in the recurrence formula (4.15) with fixed ζ and $\psi_{k-1,0}$ the possible values of $\psi_k(\zeta)$ fill a circle of center γ_k and radius η_k

$$\psi_k(\zeta)=\gamma_k+\alpha\eta_k e^{i\beta},\quad 0\le\alpha\le1,\ 0\le\beta\le2\pi, \tag{4.21}$$

then the values of $\psi_{k-1}(\zeta)$ will fill a circle too, whose center and radius are given by

$$\gamma_{k-1}=\frac{1}{\psi^*_{k-1,0}}\left(1-\frac{(1-|\psi_{k-1,0}|^2)(1+\psi_{k-1,0}\zeta^*\gamma_k^*)}{|1+\psi_{k-1,0}\zeta^*\gamma_k^*|^2-|\eta_k\psi_{k-1,0}\zeta|^2}\right), \tag{4.22a}$$

$$\hat{\eta}(\zeta)=\eta_{k-1}=\frac{\eta_k|\zeta|(1-|\psi_{k-1,0}|^2)}{|1+\psi_{k-1,0}\zeta^*\gamma_k^*|^2-|\eta_k\psi_{k-1,0}\zeta|^2}. \tag{4.22b}$$

Since the values of the arbitrary functions $\psi_N(\zeta)$ just fill—at fixed ζ—a circle of radius 1 and with the center in the origin, we have

$$\gamma_N\equiv0,\quad \eta_N\equiv1. \tag{4.22c}$$

Using (4.22a) and (4.22b) in a recurrent way, one gets finally $\gamma_0(\zeta)$ and $\eta_0(\zeta)$, as well as the center and radius of the (unweighted) admissible functions:

$$\hat{f}(\zeta)\equiv\hat{\tilde{f}}(\zeta)/C_0(M/\epsilon,\zeta)=[\tilde{h}_1(\zeta)+\tilde{h}_2^{(N)}(\zeta)-(\epsilon/\zeta^N)\gamma_0(\zeta)]/C_0(M/\epsilon,\zeta), \tag{4.22d}$$

$$\hat{\eta}(\zeta)=\eta(\zeta)/|C_0(M/\epsilon,\zeta)|=\epsilon\eta_0(\zeta)/|\zeta^N C_0(M/\epsilon,\zeta)|. \tag{4.22e}$$

Again, no poles appear in (4.20d), (4.20e) at $\zeta=0$, as one can see from (4.4) and from the denominator of (4.22b). One should also notice that for $|\zeta|=1$

$$\eta(\zeta)=\epsilon,\quad \zeta\in\Gamma \tag{4.23a}$$

$$\hat{\tilde{f}}(\zeta)=\tilde{h}(\zeta),\quad \zeta\in\Gamma, \tag{4.23b}$$

(see Fig. 3), but for $|\zeta|<1$, $\eta(\zeta)$ is usually much smaller than ϵ (see Fig. 4). This phenomenon is especially apparent when ϵ is only slightly different from ϵ_0, when some of the $|\psi_{k-1,0}|$ are close to 1 and, thus, the numerator of the formulas (4.22b) is small.

If $\epsilon=\epsilon_0$, η_k vanishes identically for $\zeta\in D$ [(4.22)], while for the boundary points we get the equations (4.20), i.e.,

$$|\tilde{f}_0(\zeta)-\tilde{h}(\zeta)|_\Gamma=\epsilon_0=\text{const}. \tag{4.24}$$

V. CONCLUDING REMARKS AND THE L^2 PROBLEM

A. Review of results

It has been shown in the previous sections that, given a data function $h(z)$ ("the histogram") along some finite parts Γ_1 of the cuts Γ of the amplitude, together with an error channel of width ϵ (1.1) and with a bound M (1.2) for the amplitude on the remaining parts $\Gamma_2=\Gamma\setminus\Gamma_1$ of the cuts, one can effectively construct the set of all analytic functions compatible with this input information. As it was shown in Sec. 4, the set of the values of all these possible ("admissible") amplitudes fill at each z

a disk (see Fig. 4) whose center $\hat{f}(\zeta)=\hat{\tilde{f}}(\zeta)/C_0(\zeta)$ and radius $\hat{\eta}(\zeta)=\eta(\zeta)/|C_0(\zeta)|$ can be computed [see Eqs. (4.22d) and (4.22e) at each point $\zeta=\zeta(z)$ (see caption of Fig. 2) in terms of ϵ and M and of the negative Fourier coefficients $c_{-1}, c_{-2}, \cdots$ in the $\zeta(z)$ complex plane of the weighted [see Eqs. (2.1)–(2.3)] histogram $\tilde{h}(\zeta)$. There is no holomorphic amplitude at all, compatible with the initial conditions (1.1) and (1.2), if ϵ is smaller than the important constant ϵ_0—the norm of the matrix X (3.16), defined in terms of the negative Fourier coefficients $c_{-1}, c_{-2}, \cdots$. In the limit $\epsilon=\epsilon_0$, there remains a unique admissible amplitude $f_0(\zeta)=\tilde{f}_0(\zeta)/C_0(\zeta)$, as in this case the radius $\eta(\zeta)\to 0$ and $\hat{\tilde{f}}(\zeta)\to\tilde{f}_0(\zeta)$. Generally, $\tilde{f}_0(\zeta)$ is the "minimal" (weighted) amplitude, i.e., that holomorphic function which approximates best (4.24) the weighted data function $\tilde{h}(\zeta)$ on the boundary Γ.

As was emphasized in Sec. 2, in the case when $\epsilon_0 \ll \epsilon$, the center $\hat{f}(\zeta)$ of the set of values in ζ of all admissible amplitudes $f(\zeta)$ differs little from the best weighted dispersion relation extrapolated function $\hat{h}(\zeta)$.[7] Conversely, when ϵ_0 is close to ϵ, most of the circle of the Nevanlinna error bound $\epsilon/|C(\zeta)|$ (corresponding to the circle of radius ϵ of Fig. 4) is empty, the admissible amplitudes being clustered in a small circle around $\hat{f}(\zeta)$ which, in general, may differ considerably from $\hat{h}(\zeta)$; in this latter case, the techniques developed in Secs. 3 and 4 represent a net improvement over the Poisson weighted dispersion relations of Refs. 7 and 8. Nevertheless, in both cases the function $\hat{f}(\zeta)$ [although nonanalytic, in contrast to the function $f_0(\zeta)$ which, being itself an admissible amplitude, is holomorphic] is for each value of z the *most unbiased* estimate one can find for an admissible amplitude taken at random.

B. Variable error channel

So far the error-channel width ϵ was regarded to be a constant. The physically important variable error case can be readily reduced to the former one using the techniques of Sec. 4 of Ref. 7, namely by introducing a supplementary weight function.[20]

$$C_1(\zeta)=\exp\{-[w(\zeta)+i\tilde{w}(\zeta)]\}, \tag{5.1}$$

where

$$w(\zeta)+i\tilde{w}(\zeta)=\frac{1}{2\pi}\int_{-\pi/2}^{\pi/2}\frac{e^{i\theta'}+\zeta}{e^{i\theta'}-\zeta}\ln\frac{\epsilon(\theta')}{\epsilon(\pi/2)}\,d\theta'. \tag{5.2}$$

Thus, the variable error channel conditions ($\zeta=e^{i\theta}$),

$$\begin{aligned} |f(\zeta)-h(\zeta)| &\le \epsilon(\theta) \quad \text{for } -\pi/2<\theta<\pi/2,\\ |f(\zeta)| &\le M \qquad\quad \text{for } \pi/2<\theta<3\pi/2, \end{aligned} \tag{5.3}$$

reduce to the constant error ones for the weighted functions $C_1(\zeta)f(\zeta)$ and $C_1(\zeta)h(\zeta)$:

$$|C_1(\zeta)f(\zeta)-C_1(\zeta)h(\zeta)|\le\epsilon(\pi/2) \quad \text{for } \zeta\in\Gamma_1, \tag{5.4}$$

$$|C_1(\zeta)f(\zeta)|\le M \qquad \text{for } \zeta\in\Gamma_2 \tag{5.5}$$

and one then proceeds as in Sec. 4.

C. A probabilistic approach: L^∞ versus L^2 problems

We should like now to outline how the similar, but much simpler, L^2-problem can be solved. This problem can be connected in a natural way[3] to the χ^2 test if one makes the assumption that the data have a normal (Gaussian) distribution around the true amplitude. [It is well known that if ξ_i are independent random variables *of class* $N(0,\sigma)$ (centered Gaussian distributions of dispersion σ), the random variable $\eta=\Sigma_{i=1}^{s}\xi_i^2$ follows the usual χ^2 distribution, being of class $H(S,\sigma)$.] However, we should like to stress that the L^2 norm *does not* exhaust the possible connections with statistics; moreover, one can relax the normality assumption by using *parameter free* test of the Kolmogoroff type,[21] which lead to L^∞ norm problems, but which are much less known among physicists than the χ^2 one. For instance, if one takes a sample of volume n from an ordered population $\{\xi\}$ subjected to a certain repartition law[22] $F(\xi)$ yet unspecified) and if $\xi^{(1)}$ and $\xi^{(n)}$ are the minimal and the maximal value $\xi^{(j)}$ of the sample so that

$$\xi^{(n)}-\xi^{(1)}=\sup|\xi^{(i)}-\xi^{(j)}|, \tag{5.6}$$

then the probability P of finding a value ξ outside the range $(\xi^{(1)},\xi^{(n)})$ with a probability greater than α/n, equals asymptotically a universal function of the parameter α. Indeed, it can be proved[23] that

$$P=\int_0^\alpha h(x)\,dx,$$

where, asymptotically, $h(x)$ is the convolution of two $\gamma(1)$ (pure exponential) distributions

$$g(x)=e^{-x} \quad \text{(asymptotically)},$$

$$h(x)=\begin{cases}\int_0^x g(x-y)g(y)dy=xe^{-x} & \text{for } x\ge 0\\ 0 & \text{for } x<0.\end{cases}$$

Hence,

$$\begin{aligned}&\lim_{n\to\infty}P\{1-[F(\xi^{(n)})-F(\xi^{(1)})]<\alpha/n\}\\ &=\int_0^\alpha h(x)\,dx=1-e^{-\alpha}(1+\alpha).\end{aligned} \tag{5.7}$$

The fact that Eq. (5.7) is irrespective of the actual form of the repartition law, $F(\xi)$ has a great theoretical importance, especially when the values for the random variable ξ are not obtained by measurements of the characteristics of some palpable object, but, rather, are inferred themselves from some empirical repartition laws, as in the case of the scattering amplitude. Moreover, one could also write down the exact (nonasymptotic) form of Eq. (5.7), so that one could find the exact value for the volume n of the sample, in order to have all ξ inside the range $\xi^{(1)}, \xi^{(n)}$ with a probability greater than a given number, to a specified confidence level. For example, if the confidence level is 0.05 ($P=0.95$) and if the probability of ξ lying between $\xi^{(1)}$ and $\xi^{(n)}$ is 99%, we find $n=473$. It is apparent that Eq. (5.6) leads to the L^∞ problems of the previous sections, but this question needs more elaborations and will be treated elsewhere.

One of the first questions one could ask in connection with the L^2 problem might be that of finding those "amplitudes" which minimalize the L^2 norm (*over the* Γ_1 *cut*) of their difference to the histogram $h(\zeta)$, i.e., those functions which minimalize the lhs of the inequality

$$\|f-h\|^2_{L^2(\Gamma_1)}\equiv\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\rho(\theta)|h(e^{i\theta})-f(e^{i\theta})|^2d\theta<\epsilon, \tag{5.8}$$

where $\rho(\theta)$ is a suitable weight function. Obviously, this hardly could be the correctly posed physical problem, since, for instance, if one had to solve this problem for a finite sum (the discret points case) instead of the integral of the lhs of (5.8), one could reduce the sum to zero taking high enough polynomials; but, of course, the higher the degree of the polynomial, the stronger it would blow up outside the range Γ_1 of energies in which the data points were given! Therefore, as it was discussed in the Introduction, one would have to add to (5.8) supplementary stabilizing physical information, which, in the context of the L^2 norm problems could be of the form

$$\|f\|^2_{L^2(\Gamma_2)}=\frac{1}{\pi}\int_{\pi/2}^{3\pi/2}\rho_2(\theta)|f(e^{i\theta})|^2d\theta<\mathcal{M}^2. \tag{5.9}$$

D. Simplified L^2 problem

An incomplete but very simple way of treating this problem is the following: Divide (5.9) by $\mathcal{M}^2/\epsilon^2$ and add it to (5.8) and try to find the Carleman weighted function $\tilde{f}=C_0f$ which minimalizes the L^2 norm integral over the whole unit circle $\Gamma_1+\Gamma_2$:

$$\frac{1}{2\pi}\int_{\Gamma_1+\Gamma_2}|h-f|^2|C_0|^2\rho(\theta)d\theta=\left\|\tilde{f}-\tilde{h}\right\|^2_{L^2(\Gamma_1+\Gamma_2)}, \tag{5.10}$$

where, as in Sec. 2, $\tilde{h}(e^{i\theta})$ was settled equal to zero on Γ_2. Obviously this new formulation of the problem departs from the previous one as the minimum of (5.10) *by no means* implies the minimum of the lhs of (5.8) under condition (5.9); furthermore, we shall show how[24] the initial problem can be answered in a correct way.

In contrast to the L^∞ problem studied in Secs. 2 and 3, the solution of this simplified L^2 problem is immediate: First, define a new external ("Carleman") function $C_\rho(\zeta)$ satisfying on the unit circle the condition

$$|C_\rho(\zeta)|_\Gamma=\sqrt{\rho(\theta)}, \tag{5.11}$$

so that [see (5.1) and (5.2)]

$$C_\rho(\zeta)=\exp\left(\frac{1}{2}\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\theta}+\zeta}{e^{i\theta}-\zeta}\log[\rho(\theta)]d\theta\right). \tag{5.12}$$

By introducing the weighted functions

$$\bar{h}(\zeta)=C_0(\zeta)C_\rho(\zeta)h(\zeta)\quad(\zeta\in\Gamma), \tag{5.13}$$

$$\bar{f}(\zeta)=C_0(\zeta)C_\rho(\zeta)f(\zeta)\quad(\zeta\in D), \tag{5.14}$$

our problem reduces to finding that analytic function $\bar{f}$ which minimalizes the unweighted L^2 norm on $\Gamma=\Gamma_1+\Gamma_2$;

$$\|\bar{f}-\bar{h}\|_{L^2(\Gamma)}=\left[\frac{1}{2\pi}\int_\Gamma|\bar{f}(e^{i\theta})-\bar{h}(e^{i\theta})|^2d\theta\right]^{1/2}<\epsilon, \tag{5.15}$$

with the obvious solution

$$\bar{f}_{\min}(\zeta)=\bar{h}_1(\zeta)\equiv\frac{1}{2\pi i}\int_\Gamma\frac{h(\zeta')C_0(\zeta')C_\rho(\zeta')}{\zeta'-\zeta}d\zeta', \tag{5.16}$$

which is the direct consequence of the orthogonality of the positive and negative parts of $\bar{f}-\bar{h}$ on Γ:

$$\|\bar{f}-\bar{h}\|^2_{L^2}=\|\bar{f}-\bar{h}_1\|^2_{L^2}+\|\bar{h}_2\|^2_{L^2}. \tag{5.17}$$

In contrast to the L^∞ norm problem where the minimal function $\tilde{f}_0$ differs in general from the optimal approximation $\tilde{f}$, the center of the whole set of functions satisfying (5.15) coincides with $\bar{f}_{\min}(\zeta)$ defined by (5.16); indeed, owing again the orthogonality of the positive and negative frequencies on the unit circle Γ, a general L^2 admissible function can always be written in the form

$$\bar{f}(\zeta)=\bar{f}_{\min}(\zeta)+\bar{l}_1(\zeta)\quad[\bar{f}_{\min}(\zeta)\equiv\bar{h}_1(\zeta)], \tag{5.18}$$

where $\bar{l}_1(\zeta)$ is an arbitrary holomorphic function whose L^2 norm is smaller than

$$\|\bar{l}_1\|_{L^2(\Gamma)}<(\epsilon^2-\epsilon^2_{\bar{h}_2})^{1/2}, \tag{5.19}$$

where

$$\epsilon_{\bar{h}_2}=\|\bar{h}_2\|_{L^2(\Gamma)}. \tag{5.20}$$

Indeed,

$$\|\bar{f}-\bar{h}\|^2_{L^2(\Gamma)}=\|\bar{f}-\bar{h}_1\|^2_{L^2}+\|\bar{h}_2\|^2_{L^2}=\|\bar{l}_1\|^2_{L^2}+\epsilon^2_{\bar{h}_2}. \tag{5.21}$$

of course, using L^2 conditions on Γ, one loses information about the behavior of $l(\zeta)$ in the special points ζ, so that the radius.

$$R(\zeta)=\sup|\bar{l}_1(\zeta)|. \tag{5.22}$$

Of the set of values of the possible admissible L^2 weighted function around $\bar{f}_{\min}(\zeta)$ exceeds considerably the L^∞ one, being equal to

$$R(\zeta)=(\epsilon^2-\epsilon^2_{\bar{h}_2})^{1/2}/(1-|\zeta|^2)^{1/2} \tag{5.23}$$

(see Appendix B), and blows up when ζ goes to the boundary Γ.

E. Complete L^2 problem

The logical drawback of this simplified approach is connected to the fact that the error channel condition (5.8) and the stabilizing condition (5.9) "mix" in (5.10) in an uncontrollable way. This mixing can be changed by changing the weight function $\rho(\theta)$ on the unknown cut by some given multiplicative factor—this amounts to the introduction in (5.10) of a supplementary Carleman function $C_0(\zeta)$—but, nevertheless, the mixing will subsist. The proper way to handle this problem is to look to all holomorphic functions satisfying (5.8) and (5.9) separately. This question was solved by Sabba Stefănescu[24] orthogonalizing the first N (N= sufficiently large) powers of ζ on both Γ_1 and Γ_2. For instance, one could first find, in a progressive way, the first N polynomials $P_n^1(\zeta)$ *of degree* n, which are *normal* and orthogonal on Γ_1 to all other polynomials of degree less than n. $P_n^{(1)}(\zeta)$ are nothing but the Legendre polynomials of the curve Γ_1 and, of course, are *not* orthogonal also on Γ_2, so that we can write

$$\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}P^{(1)}_{n_1}(e^{i\theta})P^{(1)*}_{n_2}(e^{i\theta})\,d\theta=\delta_{n_1n_2}, \tag{5.24}$$

$$\frac{1}{\pi}\int_{\pi/2}^{3\pi/2}P^{(1)}_{n_1}(e^{i\theta})P^{(1)*}_{n_2}(e^{i\theta})\,d\theta=B_{n_1n_2}, \tag{5.25}$$

where $B_{n_1n_2}$ is an $N\times N$ Hermitian matrix. Sabba-Stefănescu then diagonalizes this matrix through a suitable basis change,

$$\mathcal{P}_m(\zeta)=\sum_1^N U_{mn}P_n^{(1)}(\zeta), \tag{5.26}$$

so that,

$$\frac{1}{\pi}\int_{\pi/2}^{3\pi/2}\mathcal{P}_{m_1}(\zeta)\mathcal{P}^*_{m_2}(\zeta)\,d\theta=\sum_{n_1n_2}U_{m_1n_1}B_{n_1n_2}U^+_{n_2m_2}=\lambda_{m_1}\delta_{m_1m_2}$$

$$(\text{all }\lambda_m>0). \tag{5.26}$$

In contradistinction to the Legendre polynomials $P_n^{(1)}$ on Γ_1, *all* $\mathcal{P}_n(\zeta)$ are of *degree* N and their coefficients change when N is changed (to simplify the notations, we have dropped the index N on which they depend). Specific convergence problems arise in the infinite dimensional Hilbert, but they are carefully discussed in Ref. 24. If the $h_n^{(N)}$ are the expansion coefficients of the histogram in terms of $\mathcal{P}_n(\zeta)$ on Γ_1,

$$h_n^{(N)} = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} h(e^{i\theta})\mathcal{P}_n^*(e^{i\theta})\,d\theta, \tag{5.27}$$

the conditions (5.8) and (5.9), for an admissible polynomial $f^{(N)}(\zeta)$ of degree N, are

$$\sum_i |h_i^{(N)} - f_i^{(N)}|^2 < \epsilon_N^2 \quad \text{where } \epsilon_N^2 = \epsilon^2 - \sum_{n=N+1}^{\infty}\left|\frac{1}{\pi}\int_{-\pi/2}^{\pi/2} P_n^{(1)}(e^{i\theta}) \times h(e^{i\theta})\,d\theta\right|^2 \tag{5.28}$$

and

$$\sum_i \lambda_i |f_i^{(N)}|^2 < \mathfrak{M}^2, \tag{5.29}$$

so that the set of the admissible functions $f^{(N)}(\zeta)$ is the N dimensional Hilbert space region common to the hypersphere (5.28) centered around $h^{(N)}(\zeta)$ and the hyperelipsoid (5.29) with the center in the origin of the Hilbert space. If, as usually, $\|h\|_{L^2(\Gamma_1)} > \epsilon$, i.e., the origin is not contained in the sphere (5.28), there exists a smallest $\mathfrak{M} = \mathfrak{M}_0$ below which the ellipsoids no longer intersect the spheres:

$$\mathfrak{M}_0 = \lim_{N\to\infty} \mathfrak{M}_N^0, \tag{5.30}$$

where the $\mathfrak{M}_N^0$ are that $\mathfrak{M}$ for which the N-dimensional spheres (5.28) and ellipsoids (5.29) are tangent. It is a simple geometrical matter to find also the components of the tangent point vector $f_{0,n}^{(N)}$, the set $f_0^{(N)}(\zeta) = \sum f_{0,n}^{(N)}\mathcal{P}_n^N(\zeta)$ converging to the minimal function $f_{\mathfrak{M}_0}(\zeta)$. It is quite obvious (if two $N+1$ dimensional bodies are tangent in a N dimensional subspace, they have certainly at least one common point also in the $N+1$ dimensional space) that

$$\mathfrak{M}_{N+1}^0 < \mathfrak{M}_N^0, \tag{5.31}$$

so that the $\mathfrak{M}_N^0$ represent upper valued estimates for $\mathfrak{M}_0$; a lower bound for $\mathfrak{M}_0$ is yielded by the minimal $\mathfrak{M}$ of the simplified L^2 problem approach (5.10), namely lowering the parameter of the Carleman function C_0 until there remains only a single admissible amplitude (5.18), i.e., until

$$\|\tilde{h}\|_{L^2(\Gamma_1)} = \epsilon, \tag{5.32}$$

such that $\|l_1(\zeta)\|_{L^2(\Gamma_1)} \equiv 0$ see (5.19).

F. The analogous M_0 problem for the L^∞ case

As we have already shown at the end of the Sec. 3, there exists a minimal value M_0 also in the L^∞ norm problem, under which there are no more analytic functions satisfying (1.1) and (1.2). Let $M_2 < M_1$ and let ϵ_{0i} $(i = 1, 2)$ be the smallest (L^∞) deviations of a function holomorphic in D from the $(i = 1, 2)$ weighted histograms $\tilde{h}^{(i)}(\zeta) = h(\zeta)C_0(M_i/\epsilon, \zeta)$. Since[25]

$$\left|\frac{C_0(M_1/\epsilon,\zeta)}{C_0(M_2/\epsilon,\zeta)}\right|_\Gamma = \begin{cases} 1, & \text{on } \Gamma_1 \\ M_2/M_1 < 1, & \text{on } \Gamma_2, \end{cases}$$

we get

$$\left|\tilde{f}^{(2)}\frac{C_0(M_1/\epsilon,\zeta)}{C_0(M_2/\epsilon,\zeta)} - \tilde{h}^{(2)}\frac{C_0(M_1/\epsilon,\zeta)}{C_0(M_2/\epsilon,\zeta)}\right|_\Gamma = \begin{cases} \epsilon_{02}, & \text{on } \Gamma_1 \\ < \epsilon_{02}, & \text{on } \Gamma_2. \end{cases}$$

Since $\tilde{h}^{(2)}(\zeta)C_0(M_1/\epsilon,\zeta)/C_0(M_2/\epsilon,\zeta) \equiv \tilde{h}^{(1)}(\zeta)$, from the definition of ϵ_{01} we have to have $\epsilon_{01} < \epsilon_{02}$; moreover, as it was shown in Ref. 25 this inequality is strict. Hence we get the important monotony property

$$\epsilon_0[h; M_1/\epsilon] < \epsilon_0[h; M_2/\epsilon] \quad \text{if } M_1 > M_2. \tag{5.33}$$

Thus, decreasing M, we see that ϵ_0 increases until it reaches, at $M = M_0$, the actual value of ϵ:

$$\epsilon_0[h; M_0/\epsilon] = \epsilon \quad [= (3.26)]. \tag{5.34}$$

There are no admissible amplitudes for M less than M_0, since, owing to (5.33), $\epsilon_0(M)$ would have to be greater than the channel error width ϵ. The numerical value of M_0 can be determined from Eq. (5.34), computer programming being much facilitated by the monotony property (5.33) (computer programmes are available at request, both for M_0 and for $\mathfrak{M}_0$). The corresponding minimal amplitude,

$$f_{M_0}(\zeta) = f_0(\zeta)\big|_{\epsilon_0 = \epsilon}, \tag{5.35}$$

is of course unique *and does not depend on M* (the stabilizing lever of this L^∞ problem), being a characteristic function of all the (M-dependent) sets of admissible amplitudes (4.17). If no information exists about the possible range of the true value of M, the minimal function $f_{M_0}(\zeta)$ could be used as a first reference for the amplitude. Owing to the fact that for every

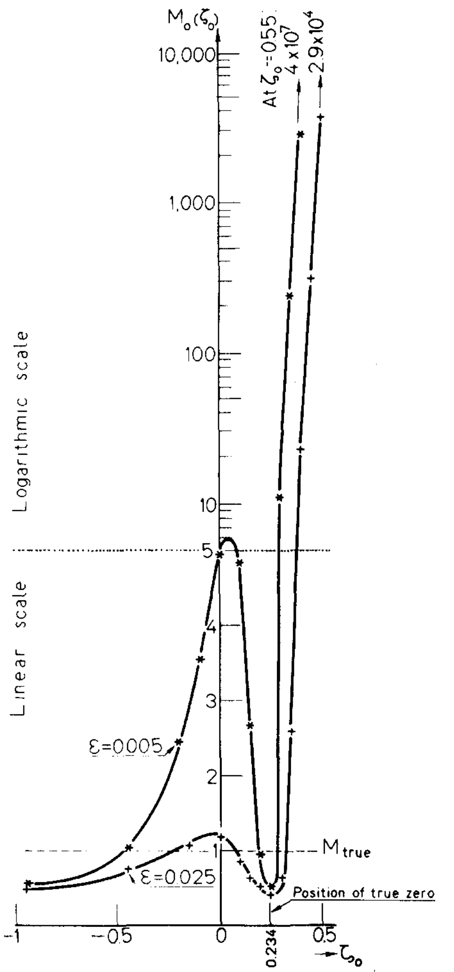

FIG. 5. Typical dependence (see Ref. 26) of M_0 versus the location of the artificial pole $\zeta_0 = \zeta(z_0)$ {here, $\zeta(z) = [30 - z + i\sqrt{195z(z-4)}]/(14z_0 - 30)$} of the function $F_1(z)/[(\zeta - \zeta_0)/(1 - \zeta_0\zeta)]$, introduced in order to locate the zero of the *model* amplitude $F_1(z) = (1 - \sqrt{4-z})/(1 + \sqrt{4-z})$. The dip corresponds to $z_0 = 3$, where the artificial pole disappears identically. Upper curve corresponds to 1% errors, while lower curve to 5% ones.

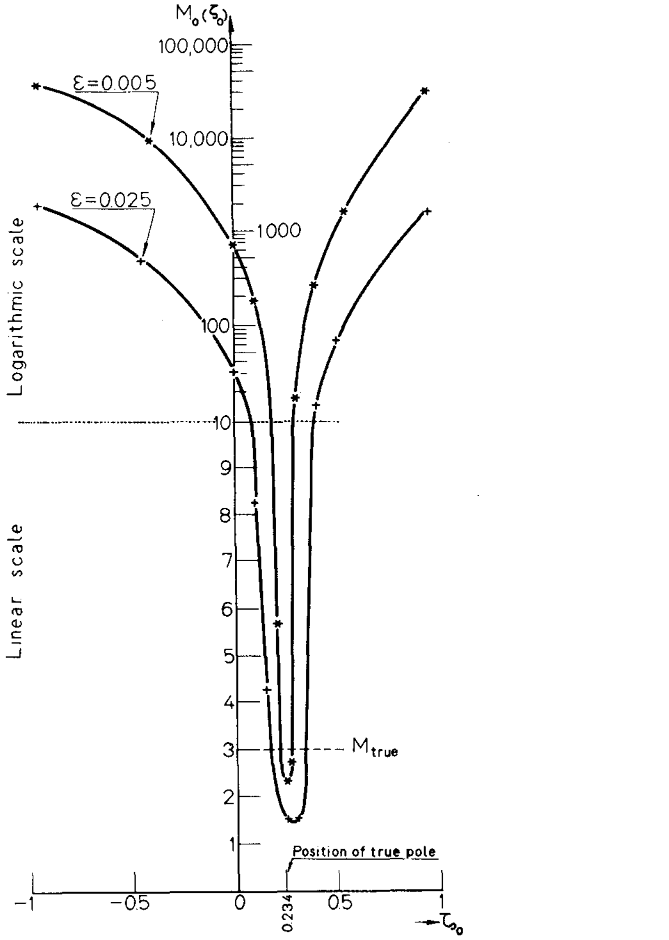

FIG. 6. Typical dependence of M_0 versus the location of the artificial zero $\zeta_0 = \zeta(z_0)$ of the function $F_2(z)(\zeta - \zeta_0)/(1 - \zeta_0\zeta)$, where $F_2(z) = 1/F_1(z)$ of Fig. 5.

L^∞-minimal function, the module $|\tilde{h} - \tilde{f}_0|$ is constant along $\Gamma_1 + \Gamma_2$ and equal to ϵ_0 [see (4.20)], $|f_{M_0}(\zeta)| = M_0$ on Γ_2 and, hence, its L^2 norm on Γ_2 coincides with M_0. Thus we get the inequality

$$\mathfrak{M}_0 \leq M_0. \tag{5.36}$$

G. Detection of singularities

On the other hand, M_0, as well as $\mathfrak{M}_0$, could be useful tools in a great variety of problems. For instance, they provide a sensitive device in the location of the zeros or poles of the amplitude.[26] For instance, if the amplitude $A(\zeta)$ has some zero in D, the function

$$f(\zeta) = \frac{A(\zeta)}{(\zeta - \zeta_0)/(1 - \zeta_0^*\zeta)} \tag{5.37}$$

whose "histogram" $h(\zeta)$ can be constructed simply on Γ_1 from the data function for the amplitude, would be nonholomorphic in D unless the parameter ζ_0 has exactly the value of the zero of A. The curve $M_0(\zeta_0)$ (see Figs. 5 and 6), and Ref. 26) is very sensitive to that, especially when the error corridor is not too large; indeed, if ϵ is small, it would be very hard, i.e., M_0 would be very high—to find holomorphic function approximating the histogram of a nonanalytic function! If some theoretical information is available (unitarity, Froissart bound, etc.) limiting the upper value of M (the dashed lines $M = M_{\text{true}}$ of Figs. 5 and 6), then the only possible values of ζ_0 are those for which $M_0(\zeta_0) < M_{\text{true}}$, the distance between the M_0 curve and the line $M = M_{\text{true}}$ defining a sort of probability distribution for the location of the zero (poles) of the amplitude. One should notice in Fig. 5 the extremely steep wall on the right of the dip corresponding to the true position [for $\zeta_0 = 0.55$, the two M_0 values are 2.9×10^4 and 4×10^7 respectively!] of the zero, which is due to the artificial pole moving towards the physical cut Γ_1. In Fig. 6 the two branches of the curve are more symmetric, although also high, since here the pole is fixed, the artificial zero, here, being moving.

ACKNOWLEDGMENTS

This paper was begun at the end of 1969 when one of us (S.C.) was at Nordita, Copenhagen; we should like to express our gratitude to Professor J. Hamilton for the hospitality extended to him. G.N. is grateful to Dr. Jan Fischer for the opportunity offered to him to present this paper at the Prague Colloquium of Elementary particles,[27] in September 1970. We should like to thank our friend Professor C. Foiaş for many long discussions and for his constant mathematical assistance. Last, but not least, we thank our colleagues I. Caprini, A. Pomponiu, and I. Sabba-Stefănescu for their interest in this problem and for their daily consultation.

APPENDIX A BY C. FOIAS

The aim of this short note is to give a comprehensive general foundation in the frame of the nowadays abstract operator theory for some of the mathematical questions considered in the present paper; it is hoped that this treatment might be useful also in other researches in the analytic theory of strong interactions.

Let K and K' be two Hilbert complex spaces. Let T and T' be two isometric operators in K, resp. K', i.e., linear operators such that

$$\|T\varphi\| = \|\varphi\|, \quad \|T'\varphi'\| = \|\varphi'\|$$

for all $\varphi \in K$, $\varphi' \in K'$, where $\|\cdot\|$ denote the norms in K and K'. We recall that a (closed linear) subspace $H \subset K$ (resp. $H' \subset K'$) is said to be invariant to T^+ (resp. T'^+), the adjoint operator of T (resp. T') if $T^+H \subset H$ (resp. $T'^+H' \subset H'$). Denote now by S (resp. S') the restriction of T^+ (resp. T'^+) to an invariant subspace H (resp. H'). Let Y be an operator from K in K' intertwining T^+ and T'^+, i.e.,

$$YT^+ = T'^+Y, \tag{A1}$$

and verifying

$$YH \subset H'. \tag{A2}$$

Let X denote the restriction $Y|_H$ of Y to H. This is obviously an operator from H to H' verifying

$$XS = S'X \quad \text{and} \quad \|X\| \leq \|Y\|. \tag{A3}$$

[where, let us recall, the norm of an operator, say X, is defined by

$$\|X\| = \sup\|X\varphi\|,$$

where the supremum is taken over all φ in the domain of X (i.e., $\varphi \in H$), verifying $\|\varphi\| = 1$.]

Conversely, the following is valid (see Ref. 15):

(I) Let X be any operator from H in H' such that

$S'X = XS$.

Then there exists an operator Y_0 from K in K' verifying

$$Y_0T^+ = T'^+Y_0, \quad Y_0|_H = X, \quad \|Y_0\| = \|X\|. \tag{A4}$$

Remark: Let us remark that $\|Y_0\|$ is the infimum of the norms of all operators Y such that $Y|_H = X$.

We introduce now some notation of functional spaces. Let L^2 be the Hilbert space of function defined on $\theta \in [0, 2\pi]$, $\varphi(e^{i\theta})$, with the following scalar product:

$$(\varphi, \psi) = \frac{1}{2\pi}\int_0^{2\pi} \varphi(e^{i\theta})\psi^*(e^{i\theta})\, d\theta,$$

representable as $\varphi(e^{i\theta}) = \sum_{-\infty}^{\infty} a_n e^{in\theta}$ with $\sum_{-\infty}^{\infty} |a_n|^2 < \infty$.

Let H^2 be the subspace of functions representable as

$$\varphi(e^{i\theta}) = \sum_0^{\infty} a_n e^{in\theta},$$

with positive frequencies only.

It is clear that every function $\varphi(e^{i\theta}) = \sum_0^{\infty} a_n e^{i\theta} \in H^2$ may be extended in a natural way to an analytic function $\varphi(z) = \sum_0^{\infty} a_n z^n$, whose boundary values $\lim_{r\to 1}\varphi(\mathrm{R}e^{i\theta})$ coincides a.e. with $\varphi(e^{i\theta})$ (see Ref. 16, Chap. III).

We shall denote by U the unitary operator in L^2 defined by $U\varphi(e^{i\theta}) = e^{i\theta}\varphi(e^{i\theta})$. Obviously $UH^2 \subset H^2$ (i.e., H^2 is invariant to U) and shall denote by $U|_{H^2}$ the restriction to H^2 of U. This is an isometric operator in H^2. The space of all functions $\chi_1(z)$ bounded and analytic for $|z| \leq 1$ is denoted by $\mathcal{H}^\infty$ and it is included in a natural way in H^2, but it is endowed with the L^∞ norm:

$$\|\chi_1(z)\|_{L^\infty} = \operatorname*{ess.\,sup}_{0 \leq \theta < 2\pi} |\chi_1(e^{i\theta})|. \tag{A5}$$

(Essential superior means the superior on a given set, modulo a set of measure zero.)

Obviously $U^+ = U^{-1}$ is given by $U^+\varphi(e^{i\theta}) = e^{-i\theta}\varphi(e^{i\theta})$ and $U^{-1}(L^2\ominus H^2) \subset L^2\ominus H^2$ (where $L^2\ominus H^2$ denotes the orthogonal supplement of H^2 in L^2). Let us denote by Q_+ and Q_- the orthogonal projections of L^2 into H^2, resp. $L^2\ominus H^2$, i.e., for $\varphi = \sum_{-\infty}^{+\infty} a_n e^{i\theta}$,

$$Q_+ \sum_{-\infty}^{\infty} a_n e^{in\theta} = \sum_0^{\infty} a_n e^{in\theta}, \quad Q_- \sum_{-\infty}^{\infty} a_n e^{i\theta} = \sum_{-\infty}^{-1} a_n e^{in\theta}. \tag{A6}$$

Moreover, let us put $U_- = U^+|_{L^2\ominus H^2}$. (As $L^2\ominus H^2$ is not invariant to U, we *cannot* write $U_-^+ = U|_{L^2\ominus H^2}$, but $U_-^+ = Q_-U|_{L^2\ominus H^2}$.) Now let

$$\chi(e^{i\theta}) = \sum_{-\infty}^{\infty} c_n e^{in\theta} \equiv \tilde{h}(e^{i\theta}) \tag{A7}$$

be bounded, i.e.,

$$\operatorname*{ess.\,sup}_{0 \leq \theta < 2\pi} |\tilde{h}(e^{i\theta})| = \|\tilde{h}\|_{L^\infty} < \infty$$

be given; denoting by "$\check{\ }$" the inversion $\check{\varphi}(e^{i\theta}) = \varphi(e^{-i\theta})$, we define X for every $\varphi_2 \in L^2\ominus H^2$ by

$$X\varphi_2 = Q_-(\chi U^+\check{\varphi}_2). \tag{A8}$$

It is clear that X depends only on the negative index coefficients, c_{-k}, $k = 1, 2, \ldots$, of $\chi(e^{i\theta})$. Then

$$XU_-\varphi_2 = Q_-(\chi U^+\overset{\vee}{\overline{U^{-1}\varphi_2}}) = Q_-(\chi U^+U\check{\varphi}_2)$$

$$= Q_-(\chi\check{\varphi}_2) = Q_-U(1 - Q_-)U^+\chi\check{\varphi}_2$$

$$+ Q_-UQ_-U^+\chi\check{\varphi}_2 = Q_-UQ_-(\chi U^+ \cdot \check{\varphi}_2)$$

$$= Q_-UX\varphi_2 = U_-^+X\varphi_2,$$

where we used, in order,

$$Q_-UQ_+ = 0, \quad U^+(\chi\varphi) = \chi U^+\varphi$$

(where $\varphi \in L^2$ and χ is bounded) and

$$Q_-U|_{L^2\ominus H^2} = U_-^+,$$

relations which can be easily verified. In this manner, if X is given by (A8), then

$$XU_- = U_-^+X. \tag{A9}$$

It is obvious that

$$\|X\| \leq \|\chi\|_{L^\infty}, \tag{A10}$$

and if X is defined in the same manner as X but with $\bar{\chi}$ instead of χ, then

$$Q_-(\chi - \bar{\chi}) = 0 \Rightarrow X = \bar{X}. \tag{A11}$$

Moreover, if we consider in $L^2\ominus H^2$ the orthogonal basis $\{e^{-in\theta}\}_{n=1}^{n=\infty}$, then X corresponds to the matrix

$$\mathbf{X} = \begin{pmatrix} c_{-1} & c_{-2} & c_{-3} & c_{-4} & \cdot & \cdot & \cdot \\ c_{-2} & c_{-3} & c_{-4} & \cdot & \cdot & & \cdot \\ c_{-3} & c_{-4} & \cdot & \cdot & \cdot & & \cdot \\ c_{-4} & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \end{pmatrix}. \tag{A12}$$

It is an easy matter to see that if such a matrix is given (this is a Hankel matrix) and if the operator X defined by it in $L^2\ominus H^2$ by the intermediate of the basis $\{e^{-in\theta}\}_{n=1}^{n=\infty}$, then X verifies (A9). Apply now Theorem I in Sec. 1 with $K = L^2$, $T = U$,

$$H = L^2\ominus H^2, \ S = U_- \text{ and } K' = H' = L^2\ominus H^2, T' = U_-, \ S' = U_-^+;$$

it results that there exists an operator Z_0 from L^2 in $L^2\ominus H^2$ such that

$$Z_0U^+ = U_-^+Z_0, \tag{A13}$$

$$Z_0|_{L^2\ominus H^2} = X, \tag{A14}$$

$$\|Z_0\| = \|X\|. \tag{A15}$$

Now apply again Theorem I with $K = L^2$, $T = U$, $H = L^2\ominus H^2$, $S = U_-$, $K' = H' = L^2$, $T' = U^+$, $S' = U$ and with X replaced by Z_0^+: $Z_0^+U_- = UZ_0^+$. It results that there exists an operator Y_0 from L^2 in L^2 such that

$$Y_0U^+ = UY_0, \tag{A16}$$

where

$$Y_0|_{L^2\ominus H^2} = Z_0^+, \tag{A17}$$

$$\|Y_0\| = \|Z_0^+\|. \tag{A18}$$

Then (A15) and (A18) give

$$\|Y_0\| = \|X\|. \tag{A19}$$

(A16) reads also

$$\overline{UY_0^+} = Y_0^+U^+ \tag{A20}$$

and with (A14) and (A15) imply (for φ_2, $\psi_2 \in L^2\ominus H^2$)

$$(X\varphi_2, \psi_2) = (Z_0\varphi_2, \psi_2) = (\varphi_2, Q_- Z_0^+\psi_2) = (\varphi_2, Y_0\psi_2)$$
$$= (Y_0^+\varphi_2, \psi_2) = (Q_- Y_0^+\varphi_2, \psi_2). \tag{A21}$$

The projection operator Q_- is redundant in the last equality of (A21), but we need it in order to have an operator under which the space $L^2 \ominus H^2$ is invariant. Hence,

$$X = Q_- Y_0^+ |_{L^2\ominus H^2}, \quad \|X\| = \|Y_0\|. \tag{A22}$$

Now denoting by **1** the constant function e^0, and putting $\chi^0(e^{i\theta}) \equiv e^{i\theta} Y_0^+ \mathbf{1}$, we have for any trigonometric polynomial

$$\varphi = \sum_{|n|\le N} a_n e^{in\theta} = \sum_{|n|\le N} (U^+)^{-n}\mathbf{1};$$

hence by (A20)

$$Y_0^+\varphi = \sum a_n U^{-n} Y_0^+ \mathbf{1} = \chi^0 U^+ \check{\varphi}. \tag{A23}$$

By using (A23), it is now an easy matter to verify that χ^0 is essentially bounded and that

$$\|Y_0\|_{L^2} = \|\chi^0\|_{L^\infty}. \tag{A24}$$

Finally, from (A23) and (A22) we obtain (A8). In this manner we have obtained the following extrapolation theorem due to Z. Nehari[13] (see also Ref. 19):

(II) Let

$$\tilde{h}_2 = \sum_{k=1}^{\infty} c_{-k} e^{-ik\theta} \tag{A25}$$

be given and suppose that the matrix corresponding to **X** *given by* (A12) *defines in* $L^2\ominus H^2$ *an operator* X *of norm* $\|X\| < \infty$; *then, if* ϵ_0 *denotes the infimum of the* $\|\chi\|_{L^\infty}$ *of all essentially bounded functions* χ *such that* $\chi_2 \equiv \tilde{h}_2$ (the negative frequency part of the histogram), *where*

$$Q_-\chi = \chi_2, \tag{A26}$$

we have

$$\epsilon_0 = \|X\|; \tag{A27}$$

moreover, this infimum is reached (namely such an optimal function is that constructed by the above successive application of Theorem I).

Remark: If $c_{-1} = 0$ for $k > N$, then the norm of the operator corresponding to the matrix **X** given by (A12) is identical to the norm of the linear transofrmation given by the matrix

$$\mathbf{X}_N = \begin{pmatrix} c_{-1} & c_{-2} & c_{-1} & \cdots & c_{-N+1} & c_{-N} \\ c_{-2} & c_{-3} & c_{-4} & \cdots & c_{-N} & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ c_{-N+1} & c_{-N} & 0 & \cdots & 0 & 0 \\ c_{-N} & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{A28}$$

in the N-dimensional complex euclidian space. Since this norm can be computed easily as the greatest eigenvalues of $\sqrt{\mathbf{X}_N^+ \mathbf{X}_N}$, it is important to know if we may neglect the remainder of the terms expansion (A24).

We shall give some properties related to the proceding theorem.

First, let us remark that if $\tilde{h}_2$ is a polynomial in $e^{-in\theta}$, say $\sum_1^N c_{-k} e^{-ik\theta}$, then, since

$$\mathbf{X} = \begin{Bmatrix} \mathbf{X}_N & 0 \\ 0 & 0 \end{Bmatrix},$$

it is easy to see that the operator X is with finite-dimensional range. Suppose now that χ is continuous, and let

$$\chi_2^{(N)} = \sum_{k=1}^{N} \left(1 - \frac{k-1}{N}\right) c_{-k} e^{-ik\theta} \tag{A29}$$

be the Fejer—Cesaro sequence for its Fourier expansion. Then

$$\|\chi_2^{(N)} - \chi\|_{L^\infty} \to 0.$$

Therefore, if $X^{(N)}$ denotes the operator corresponding to $\chi_2^{(N)}$, then instead of χ_2 we have

$$\|X^{(N)} - X\| \to 0 \quad \text{for } N \to \infty. \tag{A30}$$

In particular, (A30) implies

$$\|X^{(N)}\| \to \|X\|. \tag{A30$'$}$$

On the other hand, by the remark made above, $X^{(N)}$ is with finite-dimensional range, so that (A30) implies that X is completely continuous. But then there exists a $\varphi_2^0 \in L^2\ominus H^2$, $\|\varphi_2^0\| = 1$, such that $\|X\varphi_2^0\| = \|X\|$ (take an eigenvector for the greatest eigenvalues of X^+X). But then, using (A8), (A20), and (A24), we have

$$\|X\| = \|X\varphi_2^0\| = \|Q_-\chi^0 U^+ \check{\varphi}_2^0\| \le \|\chi^0 U^+ \check{\varphi}_2^0\|$$
$$= \|\chi^0 \check{\varphi}_2^0\| \le \|\chi^0\|_{L^\infty} = \|Y_0\| = \|X\|,$$

so that

$$\|\chi^0 \check{\varphi}_2^0\|_{L^2} = \|\chi^0\|_{L^\infty}.$$

Hence, since ($\|\varphi_2^0\|_{L^2} = 1$)

$$\|\chi^0 \check{\varphi}_2^0\|_{L^2} \le \|\chi^0\|_{L^2} \|\varphi_2^0\|_{L^2} = \|\chi^0\|_{L^2} \le \|\chi^0\|_{L^\infty},$$

we have $\|\chi^0\|_{L^2} = \|\chi^0\|_{L^\infty}$, i.e., we have

$$\frac{1}{2\pi}\int_0^{2\pi} (\|\chi^0\|^2_{L^\infty 1} - |\chi^0(e^{i\theta})|^2)\, d\theta = 0, \quad \text{i.e.,} \quad |\chi^0|$$
$$= \|\chi^0\|_{L^\infty} \text{ a.e.}$$

Thus, we obtain the following supplementary properties to the Theorem II:

(III) *Suppose that* $\tilde{h}_2$ *is continuous*; *then the norm of the matrix* [see (A29)],

$$\mathbf{X}_N' = \begin{Bmatrix} c_{-1} & (1-1/N)c_{-2} & \cdots & [1-(N-1)/N]c_{-N} \\ (1-1/N)c_{-2} & (1-2/N)c_{-3} & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ [1-(N-1)/N]c_{-N} & \cdot & \cdots & 0 \end{Bmatrix} \tag{A31}$$

in E^N *tends for* $N \to \infty$ *to* ϵ_0. *Moreover, there exists a unique* χ^0 *such that*

$$Q_-\chi^0 = \tilde{h}_2 \quad \text{and} \quad \|\chi^0\|_{L^\infty} = \epsilon_0, \tag{A32}$$

and this verifies

$$\|\chi^0\|_{L^\infty} = |\chi^0(e^{i\theta})| \quad \text{a.e. on } [0, 2\pi]. \tag{A33}$$

We have only to prove the uniticity, knowing that any optimal χ [i.e., verifying (A32)] satisfies (A33). To this purpose, let χ^1 to be another optimal function. Then,

$$\chi' = \tfrac{1}{2}(\chi^1 + \chi^0)$$

is still an optimal function, thus

$$|\chi'(e^{i\theta})| = \epsilon_0 \text{ a.e.},$$

i.e.,

$$\tfrac{1}{2}|\chi^0(e^{i\theta}) + \chi^1(e^{i\theta})| = \epsilon_0 \text{ a.e.}$$

But this obviously implies

$$\chi^0(e^{i\theta}) = \chi^1(e^{i\theta}) \text{ a.e.},$$

by the strict convexity of the modulus.

Remarks:

(1) In the hypotheses of Theorem III we have for the optimal function χ^0,

$$\|\chi^0\|_{L^\infty} = \|\chi^0\|_{L^2} \geq \|\tilde{h}_2\|_{L^2} = \sum_1^\infty |c_{-k}|^2. \tag{A34}$$

(2) Let us denote by $\epsilon_0[\tilde{h}_2]$ the ϵ defined in Theorem II. The relations (A32) and (A34) show

$$\epsilon_0[\tilde{h}_2] \geq \|\tilde{h}_2\|_{L^2}. \tag{A35}$$

Let us remark that we have quite different behavior with respect to the L^∞ norm, namely,

$$\inf \epsilon_0[\tilde{h}_2]/\|\tilde{h}_2\|_{L^\infty} = 0, \tag{A36}$$

the infimum being taken for the continuous $\tilde{h}_2$.

To see this, let us suppose the contrary, that is,

$$\epsilon_0[\tilde{h}_2] \geq \delta . \|\tilde{h}_2\|_{L^\infty} \tag{A37}$$

for all continuous $\tilde{h}_2$ and a fixed δ. Take $\tilde{h}'$ essentially bounded, i.e., $\tilde{h}' \in L^\infty$, and put $\tilde{h}_2' = Q_-\tilde{h}'$. Also let σ_n' denote the Fejer–Cesaro sequence for the Fourier expansion of $\tilde{h}'$.

Then since the $Q_-\sigma_n'$ are continuous, we have

$$\|\tilde{h}'\|_{L^\infty} \geq \|\sigma_n\|_{L^\infty} \geq \epsilon_0[Q_-\sigma_n'] \geq \delta\|Q_-\sigma_n\|_{L^\infty}, \tag{A38}$$

where the first inequality follows from the well-known properties of the Fejer kernel. It is obvious that

$$Q_-\sigma_n' \to Q_-\tilde{h}' = \tilde{h}_2' \text{ in } L^2.$$

Therefore, (A38) implies easily that

$$\|Q_-\tilde{h}'\|_{L^\infty} \leq (1/\delta)\|\tilde{h}'\|_{L^\infty}, \tag{A39}$$

for all $\tilde{h}' \in L$. Or this is impossible, since, for instance, for $\tilde{h}' = \sum_1^\infty (1/n)\sin(n\theta)$ we have $Q_-\tilde{h}' = (1/2i)\sum_1^\infty (e^{-in\theta}/n)$, which does not belong to L^∞!

(3) Theorem III is a particular (though sufficient!) case of the results V.M. Adamjan, D.Z. Arov, and M.G. Krein have published in their papers (see, for instance, Sec. 3 in Ref. 14a and Secs. 2 and 4 in Ref. 14b). Moreover, we recommend the paper 14b for its explicit formula concerning the optimal function and its complete and definite study of such extrapolation questions.

APPENDIX B

We shall remind the reader here of a simple theorem about the supremum of the module of a function $\bar{l}(\zeta)$ holomorphic in the unit disk D, whose L^2 norm on the boundary $\Gamma(|\zeta|=1)$ equals $\epsilon_l = (\epsilon^2 - \epsilon_{h_2}^2)^{1/2}$. As

$$\|\bar{l}\|_{L^2}^2 = \sum_0^\infty |a_n|^2 < \epsilon_l^2, \tag{B1}$$

where a_n are the Fourier coefficients of $\bar{l}(e^{i\theta})$, [i.e., $\bar{l}(\zeta) = \sum_0^\infty a_n \zeta^n$], one readily finds that the value of the module of $\bar{l}(\zeta)$ in the origin

$$|\bar{l}(\zeta = 0)| \equiv |a_0| \tag{B2}$$

cannot exceed ϵ_l.

A similar inequality can be derived for each interior point $\zeta_0 \in D$. Indeed, performing the usual transformation which leaves the unit circle invariant and brings the point $\zeta = \zeta_0$ into the origin,

$$\zeta' = (\zeta - \zeta_0)/(1 - \zeta_0^* \zeta), \tag{B3}$$

we get $d\theta = |1 - \zeta_0^*\zeta|^2/(1 - |\zeta_0|^2)d\theta'$, so that the unweighted problem on the $|\zeta|=1$ circle becomes a weighted problem in the $|\zeta'|=1$ circle. The weight

$$\rho = |1 - \zeta_0\zeta|^2/(1 - |\zeta_0|^2) \tag{B4}$$

can be absorbed by the exterior function

$$C_\rho = (1 - \zeta_0\zeta)/(1 - |\zeta_0|^2)^{1/2} \tag{B5}$$

[the function $C_\rho(\zeta')$ defined by (B5) has no singularities inside D!], so that one gets for the weighted function

$$\tilde{l}(\zeta) = C_\rho(\zeta)\bar{l}(\zeta), \tag{B6}$$

$$\frac{1}{2\pi}\int_0^{2\pi} |\bar{l}(\zeta')|^2 \rho(\zeta')\,d\theta' = \frac{1}{2\pi}\int_0^{2\pi} |\bar{l}(\zeta')C_\rho(\zeta')|^2\,d\theta' = \frac{1}{2\pi}\int_0^{2\pi} |\tilde{\bar{l}}(\zeta')|^2\,d\theta' < \epsilon_l^2. \tag{B7}$$

so, one gets in analogy to (B2),

$$|\tilde{l}(\zeta'=0)| \equiv |\bar{l}(\zeta_0)C_\rho(\zeta_0)| < \epsilon_l. \tag{B8}$$

Hence, as $C_\rho(\zeta_0) = (1 - |\zeta_0|^2)^{1/2}$, one finally gets

$$|\bar{l}(\zeta_0)| < \epsilon_l/(1 - |\zeta_0|^2)^{1/2}. \tag{B9}$$

[1] R. E. Cutkosky and B. B. Deo, Phys. Rev. **174**, 1859 (1968).
[2] S. Ciulli, Nuovo Cimento A **61**, 787 (1969); Nuovo Cimento A 62, 301 (1969).
[3] R. E. Cutkosky, Ann. Phys. (N.Y.) **54**, 110 (1969).
[4] J. E. Bowcock and J. John, Nucl. Phys. B **11**, 659 (1969).
[5] R. E. Cutkosky, "Partial Wave Phenomenology", Carnegie Mellon Univ. preprint talk at the ZGS Users Conference, Argonne National Laboratory, May, 1970.
[6] S. Ciulli and G. Nenciu, "Cut to cut extrapolation," talk at the Lund Conference, May 1969, and Commun. Math. Phys. **26**, 237 (1972).
[7] S. Ciulli and J. Fischer, Nucl. Phys. B **24**, 537 (1970).
[8] G. Nenciu, Nuovo Cimento Lett. **4**, 96 (1970).
[9] G. Calucci, L. Fonda, and G. C. Ghiraldi, Phys. Rev. **166**, 1719 (1968).
[10] A. N. Tihonov, Dokl. Akad. Nauk SSSR **159**, 501 (1963).
[11] Usually one maps the whole cut z plane into the ζ-unit circle; nevertheless, if information on the position of the singularities in the second sheet is available, one could map also parts of the superior Riemann sheets into the unit disk, enhancing in a substantial way [see problem (i), Sec. 3 of Ref. 2] the stability of the extrapolation.
[12] I. Schur, J. Reine Angew. Math. **147**, 205 (1917).
[13] Z. Nehari, Ann. Math. **65**, 153 (1957).
[14] V. M. Adamjan, D. Z. Arov, M. G. Krein (in Russian), Funk. Anal. ego Prid. **2**, (a) (4), 1 (1968); see also Funk. Anal. ego Prid. **2**, (b) (1), (1968).
[15] C. Foias and B. Sz. Nagy, C. R. Acad. Sci. (Paris) **266**, 493 (1968); B. Sz. Nagy and C. Foias, *Harmonic Analysis of Hilbert Space Operators* (North-Holland, Amsterdam, 1970).

[16]To find Y_χ^+ [see Eq. (3.10)] from Eq. (3.7a) one first notices that the θ-reflection operator $T\varphi \equiv \check{\varphi}$ is self-adjoint. Indeed, if u_n are the basis vectors exp (in θ), we have $(T^+u_m u_n) \equiv (u_m, Tu_n) \equiv (u_m, \check{u}_n) = (u_m, u_{-n}) = \delta_{m,-n}$ and hence $T^+u_m = u_{-m}$, i.e., $T^+ = T$. Further, observing that $TU = U^+T$, starting from (3.7a) one gets $Y_\chi^+ = (\check{\chi}^* U^+ T)^+ = T^+ U\check{\chi} = TU\check{\chi} = U^+\chi T = \chi U^+ T$, where the last equality holds because both χ and U^+ are multiplication operators.

[17]This follows directly from the definition of the norm of an operator as being the maximum of the numbers $\|Y\varphi\| / \|\varphi\|$ for all possible functions φ belonging to the space (or subspace) under consideration.

[18]R. Douglas, P. Muhly, and C. Pearcy, Mich. Math. J. 15, 385 (1968).

[19]L. B. Page, Indiana Math. J. 20 (1971).

[20]All these weight functions (C_0, C_1, and so on) defined by the values of their moduli on the boundary and having no zeroes inside D are commonly known to mathematicians under the name of "exterior functions." The name "truncated Carleman kernels" or "Carleman functions" was introduced by us at the seminar at the 1969 Lund Conference [6] in analogy with Lavrentiev's "Carleman kernels", but this appears to have produced much confusion in the literature, since references to Carleman's book on quasianalytic functions are now floating around although there is no trace in it of "Carleman functions" at all! Similar weight functions were used, for instance, in the weighted L^2 norms [see Eq. (5.11)] in the early papers of Szegö and one can certainly say that the weight functions are as old as analysis itself. Therefore, in order to avoid further confusions, it might be better to call them simply "exterior weight functions."

[21]A. N. Kolmogoroff, Giornale Ist. Ital. Attuari 4, 83 (1933).

[22]We recall that the repartition function $F(\xi_0)$ of a random variable ξ, is nothing but the probability of finding a ξ less than ξ_0.

[23]A. Rényi, "On the theory of order statistics," Acta Mathematica Acad. Sci. Hungarial. 4, (1953).

[24]Sabba-Stefǎnescu, Nucl. Phys. B 56, 287 (1973).

[25]S. Ciulli and G. Nenciu, Nuovo Cim. 8, 735 (1972).

[26]I. Capriani, S. Ciulli, A. Pomponiu, and J. Sabba-Stefǎnescu, Phys. Rev. D 5, 1658 (1972).

[27]G. Nenciu, "The Best Analytic Continuation of the Scattering Amplitudes," invited talk at the IV Colloquium on Theoretical Physics and Elementary Particles, Prague, 30 Sept.–2 Oct. 1970.

Nonlinear Lee model. II

F. Calogero

Istituto di Fisica dell'Università, 00185 Roma, Italy

Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy
(Received 31 January 1972)

The nonlinear Lee model is characterized by the Hamiltonian $H = H_0 + f(H_I)$, where $f(x)$ is a largely arbitrary (real) function and H_0 resp. H_I are the free resp. interaction parts of the usual Lee model. The exact solutions of this field-theoretical model, in some sectors of its Hilbert space, are explicitly displayed; they are similar to, but somewhat richer than, the corresponding solutions of the usual Lee model. The $V - N\theta$ and $VN - NN\theta$ sectors are treated in detail.

1. INTRODUCTION

The Lee model[1] is one of the few nontrivial field-theoretical models that can be solved exactly, and it has therefore attracted considerable attention.[2] Recently we noted that a quite general nonlinear extension of it can also be solved.[3] These results are quite interesting *per se*; moreover, they might be useful for phenomenological applications; and they should appeal to those researchers who are currently engaged in the study of field theories characterized by nonpolynomial Lagrangians; indeed the solvable example explicitly displayed here might serve as a convenient testing ground for the (approximate) approaches that have been devised to cope with nonpolynomial field theories.

The Lee model describes 3 kinds of particles, conventionally dubbed V, N, and θ. The V and N particles are superheavy baryons, and have no dynamical degrees of freedom. The θ particle is a scalar meson of mass μ. The Hamiltonian describing these (free) particles is

$$H_0 = m_{V_0} \int d\mathbf{p} V^+(\mathbf{p}) V(\mathbf{p}) + m_{N_0} \int d\mathbf{p} N^+(\mathbf{p}) N(\mathbf{p}) + \int d\mathbf{k} \omega_{\mathbf{k}} a^+(\mathbf{k}) a(\mathbf{k}), \tag{1.1}$$

where we are using Schweber's self-explanatory notation.[2] Here, and throughout this paper, we work in the Schroedinger picture.

The characterizing feature of the Lee model is that the interaction between these particles is assumed to induce the processes $V \rightleftarrows N\theta$, but to forbid the transitions $N \rightleftarrows V\theta$. Specifically, in the usual Lee model, the interaction term is written

$$H_I = \lambda_0 (2\pi)^{-3/2} \int d\mathbf{k} (2\omega_{\mathbf{k}})^{-1/2} f_1(k^2) \times \int d\mathbf{p} \{ V^+(\mathbf{p}) N(\mathbf{p}-\mathbf{k}) a(\mathbf{k}) + N^+(\mathbf{p}-\mathbf{k}) V(\mathbf{p}) a^+(\mathbf{k}) \}, \tag{1.2}$$

where again we report Schweber's notation.[2]

The usual Lee model is characterized by the Hamiltonian

$$H = H_0 + H_I. \tag{1.3}$$

The nonlinear Lee model that is studied in this paper is characterized by the Hamiltonian

$$H = H_0 + f(H_I), \tag{1.4}$$

where $f(x)$ is a real, but otherwise arbitrary, function. To ascribe a definite meaning to the operator $f(H_I)$ we shall use the Taylor expansion

$$f(x) = \sum_{n=1}^{\infty} f_n x^n; \tag{1.5}$$

the assumption that this expansion converges for all values of x, i.e., that the function $f(x)$ is entire, is certainly sufficient, although by no means necessary, for the validity of all the following developments. We shall return to this point below. It is also convenient to introduce separately the even and odd parts of the function $f(x)$, defined by

$$f_e(x) = \sum_{n=1}^{\infty} f_{2n} x^{2n}, \tag{1.6a}$$

$$f_o(x) = \sum_{n=0}^{\infty} f_{2n+1} x^{2n+1}. \tag{1.6b}$$

Note that we have assumed, to eliminate a trivial additive constant, that $f(0)$ vanishes.

The function $f_1(k^2)$ in Eq. (1.2) is a (real) cutoff function whose Fourier transform describes the size of the region over which the interaction is assumed to be smeared; it is conventionally normalized setting $f_1(0) = 1$. The case $f_1(k^2) = 1$, corresponding to a point-like interaction, causes some divergence difficulties; it is the case that has been studied more thoroughly in the usual version of the Lee model, in connection with the renormalization program. In this paper we retain the cutoff function $f_1(k^2)$, and assume that, due to its presence, all integrals are convergent. The simple and explicit nature of the final results would allow an easy discussion of the limit when the cutoff function ceases to guarantee that all integrals converge; but we prefer to defer the discussion of this point to a separate paper.[4]

The Lee model is characterized by the existence of two conserved quantities, the "baryon number" $Q_1 = n_V + n_N$ and the "Lee number" $Q_2 = n_N - n_\theta$, where n_V, n_N resp. n_θ indicate the number of (bare) V, N resp. θ particles. Obviously the quantities Q_1 and Q_2 commute both with H_0, Eq. (1.1), and with H_I, Eq. (1.2); therefore, they also commute with H, Eq. (1.4), i.e., these quantities are also conserved in the nonlinear Lee model. Thus the solution of the problem requires a treatment sector by sector. This we shall do in the following sections, according to the following plan.

In Sec. 2 we introduce a simplified notation appropriate to all sectors with only one baryon present ($Q_1 = 1$), and we establish some preliminary results appropriate to these sectors.

In Sec. 3 we treat the physical V particle, and $\theta - N$ scattering (sector $Q_1 = 1$, $Q_2 = 0$). In the usual Lee model, the V particle may be stable, or show up as a resonance in $\theta - N$ scattering. In the nonlinear Lee model discussed here, there can exist one or two physi-

Copyright © 1973 American Institute of Physics

cal V particles, depending on the parameters of the model; and again they may be stable, or show up as resonances in $\theta - N$ scattering. It should perhaps be emphasized that, in the case when two stable V particles occur, to both of them there correspond perfectly legitimate eigenstates of the Hamiltonian, normalizable and with positive norm. These results reproduce essentially the previous findings by Marr and Shimamoto,[5] who had investigated the model that obtains adding a four-point coupling to the Hamiltonian of the ordinary Lee model. Indeed, in the sector under consideration, the nonlinear Lee model considered in the present paper, in spite of its apparent generality, reduces essentially only to such an extension of the usual Lee model (see below).

In Sec. 4 we present the complete solution, in all sectors with $Q_1=1$, of the simplified model that obtains if also the kinematical degrees of freedom of the θ boson are frozen ("one-mode" model).

In Sec. 5, we introduce a simplified notation appropriate to all sectors with two baryons present ($Q_1=2$) and localized at two points $\mathbf{r}_1$ and $\mathbf{r}_2$.

In Sec. 6, we treat the case of one V and one N particles localized at a distance r from one another, and the scattering of a boson on two N particles localized at $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively (sector $Q_1=2$, $Q_2=1$). This scattering problem does not appear to have been treated previously, even for the usual Lee model. It is quite interesting, as an example of a nonspherically symmetrical (elastic) scattering process, whose scattering amplitude can be explicitly displayed.

Finally in Sec. 7 we collect some concluding remarks, and we mention a number of open problems for further study.

Certain mathematical developments have been confined to 3 Appendixes.

2. SECTORS WITH ONE BARYON PRESENT ($Q_1=1$). NOTATION AND PRELIMINARIES

For sectors of the Hilbert space with only one baryon present ($Q_1=1$), a simplified notation[6] can be used. This obtains representing the two (bare) baryon states (N or V) by a Pauli spinor, and noting that only S-wave mesons are coupled. All reference to the momentum variable can therefore be dropped, and the creation and annihilation operators of S-wave θ mesons can be introduced setting $a(\omega)=(4\pi k\omega)^{1/2}a(\mathbf{k})$ so that,

$$[a(\omega),a(\omega')]=\delta(\omega-\omega'). \tag{2.1}$$

The free part of the Hamiltonian, Eq. (1.1), can then be written in the form

$$H_0=m_{N_0}+E(1+\sigma_3)+\int d\omega\omega a^+(\omega)a(\omega), \tag{2.2}$$

where σ_3 is the third Pauli matrix and

$$2E=m_{V_0}-m_{N_0}. \tag{2.3}$$

Here, and always in the following, the integration over ω extends from μ, the θ meson mass, to infinity. The interaction part of the Hamiltonian, Eq. (1.2), can accordingly be written

$$H_I=\sigma_-\alpha^+ +\sigma_+\alpha, \tag{2.4}$$

where σ_- and σ_+ are the lowering and raising Pauli operators $[\sigma_\pm|\pm\rangle=0,\ \sigma_\pm|\mp\rangle=|\pm\rangle]$, and of course the states $|+\rangle$ resp. $|-\rangle$ represent the bare V resp. N states (and the vacuum as far as the boson field is concerned). The quantities α and α^+ are defined by

$$\alpha=\int d\omega\lambda(\omega)a(\omega),\quad \alpha^+=\int d\omega\lambda(\omega)a^+(\omega), \tag{2.5}$$

with

$$\lambda(\omega)=[\lambda_0/(2\pi)]k^{1/2}f_1(k^2),\quad \omega=(k^2+\mu^2)^{1/2}. \tag{2.6}$$

Note that, consistently with this definition and the normalization condition $f_1(0)=1$, $\lambda(\mu)$ vanishes.

We now introduce the two important quantities

$$\Lambda^2=\int d\omega\lambda^2(\omega)=[\alpha,\alpha^+] \tag{2.7}$$

and

$$F(\omega)=P\int d\omega'\lambda^2(\omega')/(\omega'-\omega). \tag{2.8}$$

As mentioned in the Introduction, throughout this paper we shall assume that the function $\lambda(\omega)$ vanishes sufficiently fast at infinity so that the integral of Eq. (2.7) [and *a fortiori* the integral of Eq. (2.8)] converge. A sufficient condition for that is that

$$\lim_{\omega\to\infty}[\omega^{(1/2)+\epsilon}\lambda(\omega)]=0,\quad \epsilon>0. \tag{2.9}$$

In the point-like case,

$$\lambda(\omega)=[\lambda_0/(2\pi)]k^{1/2}, \tag{2.10}$$

neither one of the two integrals of Eqs. (2.7), (2.8) converges; this case is considered in a separate paper.[4] The function $\lambda(\omega)$ is moreover assumed to be such that, aside from the question of asymptotic convergence just mentioned, the integrals of Eqs. (2.7) and (2.8) be well defined and finite.

The function $F(\omega)$ vanishes as ω diverges to $-\infty$, and it is a monotonically increasing function of ω in the interval $-\infty$, μ:

$$F(-\infty)=0, \tag{2.11a}$$

$$dF(\omega)/d\omega>0\quad\text{for }\omega\leq\mu. \tag{2.11b}$$

These two important properties follow immediately from its definition, Eq. (2.8).

The problem that we are to investigate is characterized by the Hamiltonian of Eqs. (1.4), (2.2), and (2.3). We end this section reporting the important identity

$$\begin{aligned}f(H_I)=\sum_{m=0}^{\infty}(-)^m\Lambda^{-2m}\sum_{s=0}^{m}(-)^s[s!(m-s)!]^{-1}\\ \times\{(\alpha^+)^m\alpha^m[P_+f_e(\Lambda(s+1)^{1/2})+P_-f_e(\Lambda s^{1/2})]\\ +[\sigma_-(\alpha^+)^{m+1}\alpha^m+\sigma_+(\alpha^+)^m\alpha^{m+1}]f_o(\Lambda(s+1)^{1/2})/\\ (\Lambda(s+1)^{1/2})\},\end{aligned} \tag{2.12}$$

where P_+ resp. P_- are the projection operators over the states $|+\rangle$ resp. $|-\rangle$,

$$P_\pm=(1\pm\sigma_3)/2. \tag{2.13}$$

This formula is proved in Appendix A. It implies that, for the whole sector of Hilbert space with $Q_1=1$, the only values of the function $f(x)$ that play any role in the dynamics of the nonlinear Lee model are the values $f(\Lambda n^{1/2})$, $n=0,1,2\cdots$. In particular, if the function $f(x)$, although not identically vanishing, has the property

$$f(\Lambda n^{1/2})=0, \quad n=0,1,2,\cdots, \tag{2.14}$$

then clearly for all sectors of Hilbert space with $Q_1=1$ the Hamiltonian H reduces effectively to the free part H_0. An example of a class of functions $f(x)$ that possess the property (2.14) is

$$f(x)=\bar{f}(x)\sin[\pi(x/\Lambda)^{1/2}], \tag{2.15}$$

with $\bar{f}(x)$ entire.

More generally, if

$$f_e(\Lambda n^{1/2})=f_e, \quad n=0,1,2,\cdots, \tag{2.16a}$$

$$f_o(\Lambda n^{1/2})=f_o, \quad n=1,2,3,\cdots, \tag{2.16b}$$

with f_e and f_o independent on n, then Eq. (2.12), together with Eq. (A.10), yields

$$f(H_I)=f_e+(f_o/\Lambda)H_I, \tag{2.17}$$

so that for this class of functions the nonlinear Lee model reduces essentially to the usual Lee model (in the sectors of Hilbert space with one baryon present).

3. THE PHYSICAL V PARTICLE, AND θN SCATTERING (SECTOR $Q_1=1, Q_2=0$)

In this sector, all states can be represented as a superposition of the bare V state $|+\rangle$, and of the states $a^+(\omega)|-\rangle$ representing a boson of energy ω and a N particle. Therefore, in this sector the Hamiltonian (1.4) reduces simply to[7]

$$H=H_0+f_e(\Lambda)P_+ +[f_o(\Lambda)/\Lambda](\sigma_-\alpha^+ +\sigma_+\alpha) - [f_e(\Lambda)/\Lambda^2]P_-\alpha^+\alpha, \tag{3.1}$$

since all the other operators appearing in the rhs of Eq. (2.12) give a vanishing result when applied to these states.

The field theoretical model characterized by the Hamiltonian (3.1) coincides essentially with the "Lee model with a four-point coupling" introduced several years ago by Marr and Shimamoto.[5] Thus the results reported in the rest of this section reproduce to a large extent the findings of these authors.

We look first of all for normalizable eigenstates of the Hamiltonian H, Eq. (3.1); such states represent physical (stable) V particles. As we shall presently see, depending on the parameters of the model, there can be two, one, or zero such states.

Let us consider the stationary Schrödinger equation

$$H|V\rangle=m_V|V\rangle, \tag{3.2}$$

the state $|V\rangle$ being represented by

$$|V\rangle=\mathcal{N}_V(|+\rangle+\int d\omega u(\bar{\omega},\omega)a^+(\omega)|-\rangle). \tag{3.3}$$

The normalization constant $\mathcal{N}_V$ is clearly expressed, in terms of the function $u(\bar{\omega},\omega)$, by the formula

$$\mathcal{N}_V=[1+\int d\omega|u(\bar{\omega},\omega)|^2]^{-1/2}. \tag{3.4}$$

The quantity $\bar{\omega}$ is introduced for convenience; it is related to the eigenvalue m_V by

$$m_V=m_N+\bar{\omega}, \tag{3.5}$$

where we have written m_N in place of m_{N_0} since obviously the bare and physical masses of the N baryon coincide,[7] as do the corresponding states. It is also convenient to introduce the four constants[7]

$$c=f_o(\Lambda)/\Lambda, \tag{3.6}$$

$$b=f_e(\Lambda)/\Lambda^2, \tag{3.7}$$

$$\omega_0=m_{V_0}-m_N+f_e(\Lambda)=2E+f_e(\Lambda), \tag{3.8}$$

$$\Omega=\omega_0-(c^2/b). \tag{3.9}$$

In the usual Lee model, $c=1$, $b=0$, $\omega_0=m_{V_0}-m_N=2E$, Ω diverges and $b\Omega=-1$.

Using Eq. (3.1), we immediately get from Eq. (3.2) the relations

$$\omega_0-\bar{\omega}+c\int d\omega\lambda(\omega)u(\bar{\omega},\omega)=0, \tag{3.10}$$

$$(\omega-\bar{\omega})u(\bar{\omega},\omega)+c\lambda(\omega)+b\lambda(\omega)\int d\omega'\lambda(\omega')u(\bar{\omega},\omega')=0. \tag{3.11}$$

From these equations we obtain

$$u(\bar{\omega},\omega)=\lambda(\omega)[b(\bar{\omega}-\Omega)]/[c(\bar{\omega}-\omega)], \tag{3.12}$$

and the eigenvalue condition

$$-(\bar{\omega}-\omega_0)/[b(\bar{\omega}-\Omega)]=F(\bar{\omega}), \tag{3.13}$$

with $F(\bar{\omega})$ defined by Eq. (2.8).

Clearly any solution $\bar{\omega}$ of Eq. (3.13) yields a normalizable eigenstate, provided the stability condition

$$\bar{\omega}<\mu \tag{3.14a}$$

or, equivalently,

$$m_V<m_N+\mu, \tag{3.14b}$$

holds. If on the other hand a solution of Eq. (3.13) occurs for $\bar{\omega}>\mu$, it does not yield a normalizable eigenstate, because the integral over ω in Eq. (3.3) becomes singular, as it is apparent from the explicit form of $u(\bar{\omega},\omega)$, Eq. (3.12). Such solutions represent unstable V states, and show up as resonances in θN scattering (see below).

In Figs. 1 and 2 we have provided a graphical representation of Eq. (3.13) in the two cases $b>0$ and $b<0$.

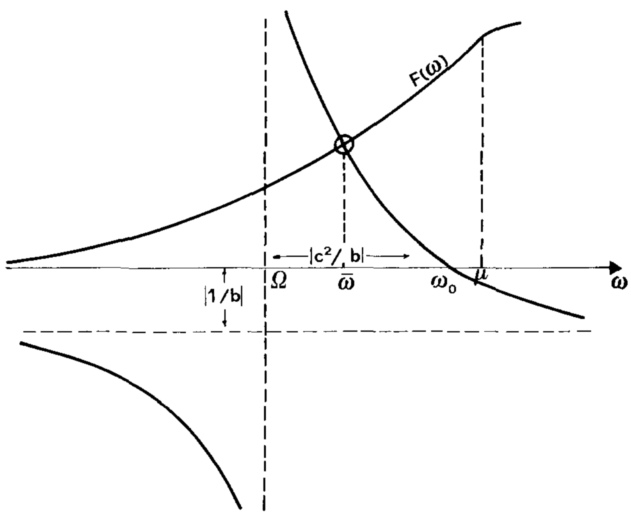

FIG. 1. Graphical display of Eq. (3.13) for positive b.

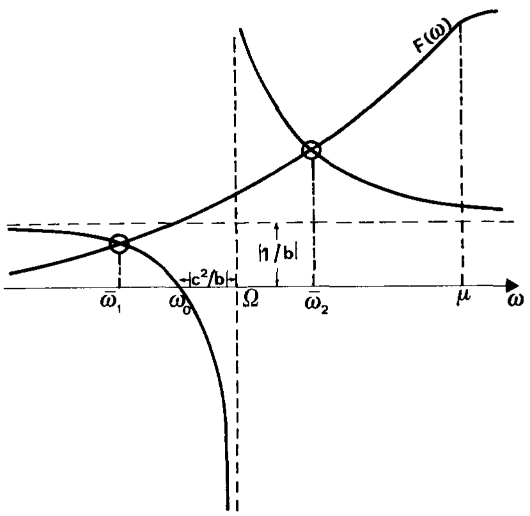

FIG. 2. Graphical display of Eq. (3.13) for negative b.

The graph of $F(\omega)$ has been drawn taking account of the properties of Eqs. (2.11). It is clear from Fig. 1 that, if b is positive, there can be at most one solution of Eq. (3.13) for $\bar{\omega} < \mu$; and indeed there will be one solution, if the condition $\omega_0 < \mu$, or, equivalently,

$$m_{V_0} < m_N + \mu - f_e(\Lambda) \tag{3.15}$$

holds. Clearly a less stringent necessary condition for the existence of one solution of Eq. (3.13) with $\bar{\omega} < \mu$ is provided, in this case $b > 0$, by the inequality

$$\Omega < \mu, \tag{3.16a}$$

or, equivalently,

$$m_{V_0} < m_N + \mu - f_e(\Lambda) + c^2/b, \tag{3.16b}$$

and, if this inequality holds, a necessary and sufficient condition for the existence of one solution of Eq. (3.13) in the range $\bar{\omega} < \mu$ is

$$(\omega_0 - \mu)/[b(\mu - \Omega)] < F(\mu) \tag{3.17a}$$

or, equivalently,

$$m_{V_0} < m_N + \mu - f_e(\Lambda) + (c^2/b)\{1 + [bF(\mu)]^{-1}\}^{-1}. \tag{3.17b}$$

On the other hand, as is apparent from Fig. 2, if b is negative, there can be two, one, or zero solutions of Eq. (3.13), in the stability region $\bar{\omega} < \mu$. Clearly the condition (3.16) is in this case sufficient to guarantee the existence of at least one solution, and necessary for the existence of two solutions. Another independent condition, that, in this case $b < 0$, is also both sufficient to guarantee the existence of at least one stable V particle, and necessary for the existence of two stable V particles, is the inequality

$$-1/b < F(\mu). \tag{3.18}$$

And if both inequalities, (3.17) and (3.18), hold, then validity of the inequality (3.17) is, in this case $b < 0$, necessary and sufficient for the existence of two stable V particles.

If b vanishes (as is the case in the usual Lee model), Eq. (3.13) becomes

$$-c^2(\bar{\omega} - \omega_0) = F(\bar{\omega}), \tag{3.19}$$

and it clearly admits at most one solution in the stability region $\bar{\omega} < \mu$. A necessary and sufficient condition for the occurrence of this solution is provided by the inequality

$$\omega_0 < \mu + F(\mu)/c^2 \tag{3.20a}$$

or, equivalently,

$$m_{V_0} < m_N + \mu + F(\mu)/c^2. \tag{3.20b}$$

The most remarkable difference between the nonlinear Lee model and the usual Lee model is the possibility that two stable V particles appear in the former case. It should be emphasized that these two states are sound physical states, and in particular their norm is positive.

Let us turn now to a discussion of (S-wave) scattering. We consider again the stationary Schrödinger equation,

$$H|\omega, {}^{\text{in}}_{\text{out}}) = (m_N + \omega)|\omega, {}^{\text{in}}_{\text{out}}), \tag{3.21}$$

with the (incoming and outcoming) states represented by

$$|\omega, {}^{\text{in}}_{\text{out}}) = a^\dagger(\omega)|-\rangle + \int d\omega' \bar{u}_{\substack{\text{in}\\\text{out}}}(\omega, \omega') a^\dagger(\omega')|-\rangle + \eta_{\substack{\text{in}\\\text{out}}}(\omega)|+\rangle. \tag{3.22}$$

Now the function $\bar{u}(\omega, \omega')$ is singular at $\omega' = \omega$ (see below), and the prescription to treat the singularity is the distinguishing feature of the ingoing and outgoing states.

Inserting the ansatz (3.22) into the Schrödinger equation (3.21) and proceeding as above, one easily obtains the following explicit expressions for $\bar{u}_{\substack{\text{in}\\\text{out}}}(\omega, \omega')$ and $\eta_{\substack{\text{in}\\\text{out}}}(\omega)$:

$$\eta_{\substack{\text{in}\\\text{out}}}(\omega) = c\lambda(\omega)\{(\omega - \omega_0)[1 + bF_{\substack{\text{in}\\\text{out}}}(\omega)] + c^2 F_{\substack{\text{in}\\\text{out}}}(\omega)\}^{-1}, \tag{3.23}$$

$$\bar{u}_{\substack{\text{in}\\\text{out}}}(\omega, \omega') = (\omega - \omega' \pm i\epsilon)^{-1}(b/c)(\omega - \Omega)\lambda(\omega')\eta_{\substack{\text{in}\\\text{out}}}(\omega), \tag{3.24}$$

with

$$F_{\substack{\text{in}\\\text{out}}}(\omega) = \int d\omega' \lambda^2(\omega')/(\omega' - \omega \mp i\epsilon) \tag{3.25a}$$

$$= F(\omega) \pm i\pi\lambda^2(\omega). \tag{3.25b}$$

In Eq. (3.25b), $F(\omega)$ is of course the quantity defined by Eq. (2.8).

From these expressions, and the standard definition of the (reduced) S matrix (for S-wave scattering),

$$(\omega', \text{out}|\omega, \text{in}) = \delta(\omega - \omega') S_0(\omega) = \delta(\omega - \omega') \exp[2i\delta_0(\omega)], \tag{3.26}$$

one easily obtains the expression

$$S_0(\omega) = J(\omega - i\epsilon)/J(\omega + i\epsilon), \tag{3.27}$$

with the "Jost function" $J(z)$ defined by

$$J(z) = (z - \omega_0)/[b(z - \Omega)] + \int d\omega \lambda^2(\omega)/(\omega - z). \tag{3.28}$$

Equation (3.27) is of course written for real ω. The analytic continuation of $S_0(\omega)$ to complex ω can be performed in the usual way, i.e., evaluating the Jost function $J(\omega)$ appearing in the denominator of Eq. (3.27) on the "physical sheet" of its Riemann surface, and the

Jost function of the numerator on the "unphysical sheet" (connected to the physical sheet through the cut from $\omega=\mu$ to $\omega=\infty$). Using this analytic continuation, it is easy to ascertain that the analytically continued S matrix has no poles on the physical sheet, except for those occurring for real $\omega<\mu$, that correspond to stable physical states and that have already been discussed. In fact setting $\omega=x+iy$ and looking for zeros of the denominator of the S matrix,

$$\mathcal{J}(x+iy+i\epsilon)=0, \tag{3.29}$$

we get from Eq. (3.28) the two (real) equations

$$(x-\omega_0)(x-\Omega)+y^2=-b[(x-\Omega)^2+y^2] \times \int d\omega\lambda^2(\omega)(\omega-x)/[(\omega-x)^2+(y+\epsilon)^2], \tag{3.30a}$$

$$c^2y=-b^2(y+\epsilon)[(x-\Omega)^2+y^2] \times \int d\omega\lambda^2(\omega)/[(\omega-x)^2+(y+\epsilon)^2]. \tag{3.30b}$$

Here we have retained the variable ϵ, although of course the limit $\epsilon\to 0$ is always implicit. Clearly Eq. (3.30b) implies that y vanishes, i.e., the only poles of $S_0(\omega)$ on the physical sheet can occur for real ω; and Eq. (3.30a) implies that they can occur only in the stability region $\omega<\mu$ and that they are indeed determined by the same eigenvalue equation that yields the mass of the physical V particle, Eq. (3.13). Note that the requirement that $S_0(\omega)$ have the analyticity properties associated with a correct causal behavior implies no additional restrictions on the parameters of the nonlinear Lee model. It is also evident that to every solution of the eigenvalue Eq. (3.13) occurring in the instability region $\omega>\mu$ there corresponds a resonance in (S-wave) $\theta-N$ scattering. This is most vividly displayed by the formula for the total cross section for elastic $\theta-N$ scattering, that follows immediately from Eq. (3.27), and reads

$$\sigma(\omega)=4\pi(\omega^2-\mu^2)^{-1}[\pi\lambda^2(\omega)]^2\{[(\omega-\omega_0)/b(\omega-\Omega)+F(\omega)]^2 + [\pi\lambda^2(\omega)]^2\}^{-1} \tag{3.31a}$$

$$=4\pi[\lambda_0^2f_1^2(k^2)/(4\pi)]^2\{[(\omega-\omega_0)/b(\omega-\Omega)+F(\omega)]^2 + [k\lambda_0^2f_1^2(k^2)/(4\pi)]^2\}^{-1}. \tag{3.31b}$$

Note that the total cross section vanishes for $\omega=\Omega$; this value is of course physically accessible only if $\Omega>\mu$.

In the previous discussion we have implicitly assumed that the quantity c does not vanish. If it does vanish, then Ω coincides with ω_0, and in place of the eigenvalue equation (3.13) we have the equation

$$-1/b=F(\overline{\omega}), \tag{3.32}$$

that clearly admits no solution in the range $\overline{\omega}<\mu$ if b is positive, and one solution if b is negative and larger in modulus than $1/F(\mu)$. The corresponding eigenstate is

$$|V\rangle=[\int d\omega\lambda^2(\omega)/(\omega-\overline{\omega})^2]^{-1/2}\int d\omega\lambda(\omega)(\omega-\overline{\omega})^{-1}a^+(\omega)|-\rangle. \tag{3.33}$$

In addition, the bare V state $|+\rangle$ is a (normalized) eigenstate of H, with eigenvalue $m_V=m_N+\omega_0$ (even if $\omega_0>\mu$). In fact, as it is easily seen from Eq. (3.1), if $c=0$, i.e., if $f_o(\Lambda)=0$, the interaction part of the Hamiltonian cannot induce transitions between the (bare) N and V states (in the sector of Hilbert space under present consideration). As for the scattering problem, the formulas given above continue to be valid even if $c=0$, with the obvious simplification implied by the equality of ω_0 with Ω.

4. "ONE-MODE" LEE MODEL (SECTORS $Q_1=1$)

In this section we consider a simplified model, that obtains from the nonlinear Lee model considered in the previous sections if also the kinematical degrees of freedom of the θ boson are frozen. The corresponding case had been previously discussed, for the ordinary Lee model, by Barton.[8]

The model is characterized by the Hamiltonian

$$H=H_0+f(H_I) \tag{4.1}$$

with

$$H_0=E\sigma_3+\omega a^+a \tag{4.2}$$

and

$$H_I=\lambda(a^+\sigma_-+a\sigma_+), \tag{4.3}$$

where now ω and λ are two constants and the creation and annihilation operators a^+, a, obey the commutation relations

$$[a,a^+]=1. \tag{4.4}$$

The other symbols are defined as above.

The (complete) spectrum of this Hamiltonian is of course discrete, and it can be given explicitly together with the corresponding eigenstates. The formulas are

$$H|n,\pm\rangle=W_{n,\pm}|n,\pm\rangle, \tag{4.5}$$

$$W_{n,\pm}=(n-\tfrac{1}{2})\omega+f_e(\lambda n^{1/2})\pm[(E-\tfrac{1}{2}\omega)^2+f_o^2(\lambda n^{1/2})]^{1/2}, \tag{4.6}$$

$$|n,\pm\rangle=\mathcal{N}_{n,\pm}\{n^{1/2}f_o(\lambda n^{1/2})(a^+)^{n-1}|+\rangle+[W_{n,\pm}-E -f_e(\lambda n^{1/2})-(n-1)\omega](a^+)^n|-\rangle\}; \tag{4.7}$$

$$\mathcal{N}_{n,\pm}=\{2\cdot n![(E-\tfrac{1}{2}\omega)^2+f_o^2(\lambda n^{1/2})]^{1/2}([(E-\tfrac{1}{2}\omega)^2 +f_o^2(\lambda n^{1/2})]^{1/2}\pm 1)\}^{-1/2}. \tag{4.8}$$

Here $n=1,2,3\cdots$; clearly this quantum number is directly related to the Lee number Q_2 by

$$Q_2=1-n. \tag{4.9}$$

In addition, the state $|-\rangle$ is also an eigenstate of H, with eigenvalue $-E$.[7] These results can be easily verified by direct substitution; they have actually been obtained using a technique[9] that guarantees that these states constitute the complete set of eigenstates of the Hamiltonian (4.1).

The explicit spectrum (4.6) displays the conditions that the function $f(x)$ must satisfy in order that the Hamiltonian H of Eqs. (4.1–3) possess a ground state; this condition coincides with the requirement that the spectrum (4.6) possess a finite minimum. Clearly if the function $f(x)$ is finite for finite real x, as we always assume, a necessary and sufficient condition for this to happen is that there exist a finite constant M such that

$$\lim_{x\to\infty}[f_e(x)\pm f_o(x)]>-M. \tag{4.10}$$

5. SECTORS WITH TWO BARYONS PRESENT ($Q_1 = 2$). NOTATION AND PRELIMINARIES

In this section we introduce a simplified notation that is appropriate to the treatment of sectors of Hilbert space with two baryons present ($Q_1 = 2$). It is constructed in clear analogy to the notation introduced in Sec. 2 for sectors with one baryon present. It is also clear how this notation should be extended in order to treat sectors with more than two baryons present; a problem, however, that is not considered in this paper.

The free part of the Hamiltonian is written as

$$H_0 = 2m_N + \sum_{j=1}^{2} E[\sigma_3(j)+1] + \int d\mathbf{k}\,\omega_k a^+(\mathbf{k})a(\mathbf{k}), \qquad (5.1)$$

where

$$[a(\mathbf{k}), a^+(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}'), \qquad (5.2)$$

$$\omega_k = (k^2+\mu^2)^{1/2}. \qquad (5.3)$$

The integrals over $d\mathbf{k}$, here and always in the following, extend over the whole space. The Pauli matrices $\sigma_3(j)$ act on a two-spinor space according to the following notation:

$$\sigma_3(1)\,|\,\alpha, \beta\rangle = \alpha\,|\,\alpha, \beta\rangle, \qquad (5.4a)$$

$$\sigma_3(2)\,|\,\alpha, \beta\rangle = \beta\,|\,\alpha, \beta\rangle, \qquad (5.4b)$$

where α and β stand for + or −.

The operator H_I is written as

$$H_I = \sum_{j=1}^{2} \int d\mathbf{k}\gamma(k)[a(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r}_j)\sigma_+(j) + a^+(\mathbf{k}) \times \exp(-i\mathbf{k}\cdot\mathbf{r}_j)\sigma_-(j)]; \qquad (5.5)$$

we are clearly assuming that the two baryons are localized at the positions $\mathbf{r}_1$ and $\mathbf{r}_2$, so that, for instance, the state $|+,-\rangle$ represents a (bare) V particle localized at $\mathbf{r}_1$ and a N particle localized at $\mathbf{r}_2$. The (real) form factor $\gamma(k)$ is related to the form factor of Eq. (1.2) by

$$\gamma(k) = \lambda_0(2\pi)^{-3/2}(2\omega_k)^{-1/2}f_1(k^2) \qquad (5.6a)$$

and to the form factor of Eq. (2.6) by

$$\gamma(k) = (2\pi)^{-1/2}(2k\omega_k)^{-1/2}\lambda(\omega). \qquad (5.6b)$$

The raising and lowering operators $\sigma_+(j)$, $\sigma_-(j)$ are defined by

$$\sigma_\alpha(1)\,|\,\alpha, \beta\rangle = 0, \quad \sigma_\alpha(1)\,|-\alpha, \beta\rangle = |\,\alpha, \beta\rangle, \qquad (5.7a)$$

$$\sigma_\beta(2)\,|\,\alpha, \beta\rangle = 0, \quad \sigma_\beta(2)\,|\,\alpha, -\beta\rangle = |\,\alpha, \beta\rangle, \qquad (5.7b)$$

where again α stands for + and −, and so does β.

We now introduce the important quantities

$$\Lambda_\pm^2(r) = \int d\mathbf{k}\gamma^2(k)[1 \pm \exp(i\mathbf{k}\cdot\mathbf{r})] \qquad (5.8a)$$

$$= 4\pi\int_0^\infty dk k^2\gamma(k)[1\pm\sin(kr)/(kr)] \qquad (5.8b)$$

$$= \int_\mu^\infty d\omega\lambda^2(\omega)[1\pm\sin(kr)/(kr)] \qquad (5.8c)$$

and

$$F_\pm(\omega;r) = P\int d\mathbf{k}'\ \gamma^2(k')[1\pm\exp(i\mathbf{k}'\cdot\mathbf{r})]/(\omega_{k'}-\omega) \qquad (5.9a)$$

$$= 4\pi P\int_0^\infty dk'k'^2\gamma^2(k')[1\pm\sin(k'r)/(k'r)]/(\omega_{k'}-\omega) \qquad (5.9b)$$

$$= P\int_\mu^\infty d\omega'\lambda^2(\omega')[1\pm\sin(k'r)/(k'r)]/(\omega'-\omega). \qquad (5.9c)$$

As already mentioned, we assume that these integrals are convergent. A comparison with the notation of Sec. 2 implies

$$\Lambda_\pm^2(\infty) = \tfrac{1}{2}[\Lambda_+^2(r) + \Lambda_-^2(r)] = \Lambda^2, \qquad (5.10a)$$

$$F_\pm(\omega;\infty) = \tfrac{1}{2}[F_+(\omega;r) + F_-(\omega;r)] = F(\omega), \qquad (5.10b)$$

$$\Lambda_+^2(0) = 2\Lambda^2, \qquad (5.10c)$$

$$F_+(\omega;0) = 2F(\omega), \qquad (5.10d)$$

$$\Lambda_-(0) = 0, \qquad (5.10e)$$

$$F_-(0) = 0, \qquad (5.10f)$$

with Λ and $F(\omega)$ defined by Eqs. (2.7) and (2.8).

Note that the functions $F_\pm(\omega;r)$ have the same properties as the function $F(\omega)$:

$$F_\pm(-\infty;r) = 0, \qquad (5.11a)$$

$$dF_\pm(\omega;r)/d\omega > 0, \quad \text{for } \omega \le \mu. \qquad (5.11b)$$

The first property is directly implied by the definition (5.9) and the assumed finite existence of $\Lambda_\pm(r)$, Eq. (5.8); the second property obtains differentiating Eq. (5.9) and noting that the modulus of $(\sin x)/x$ never exceeds unity.

6. NV STATES, AND θNN SCATTERING (SECTOR $Q_1 = 2$, $Q_2 = 1$)

In this section, the physical states are superpositions of the bare VN states $|+,-\rangle$ and $|-,+\rangle$ and of the states $a^+(\mathbf{k})|-,-\rangle$ representing a boson of momentum k and two N particles. Of course the two baryons are localized at the fixed positions $\mathbf{r}_1$ and $\mathbf{r}_2$.

We look first of all for normalizable eigenstates of the Hamiltonian H of Eqs. (1.4), (5.1), and (5.5). We indicate such states as $|W(r);\mathbf{r}_1,\mathbf{r}_2\rangle$, where $W(r)$ indicates the corresponding eigenvalue of H:

$$H\,|\,W(r);\mathbf{r}_1,\mathbf{r}_2\rangle = W(r)\,|\,W(r);\mathbf{r}_1,\mathbf{r}_2\rangle. \qquad (6.1)$$

It is convenient to introduce a quantity $\bar\omega(r)$ through

$$W(r) = 2m_N + \bar\omega(r). \qquad (6.2)$$

As we shall presently see, the quantity $\bar\omega(r)$ must satisfy the stability condition $\bar\omega(r) < \mu$ in order that the state $|W(r);\mathbf{r}_1,\mathbf{r}_2\rangle$ be normalizable. The physical meaning of this condition is obvious.

In these equations, and always below,

$$\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2 \qquad (6.3)$$

is the (fixed) distance between the baryons.

For the state $|W(r);\mathbf{r}_1,\mathbf{r}_2\rangle$ we have the ansatz

$$|\,W(r);\mathbf{r}_1,\mathbf{r}_2\rangle = \mathcal{N}(r)\{\eta_1(r)\,|+,-\rangle + \eta_2(r)\,|-,+\rangle + \int d\mathbf{k}u[\bar\omega(r),\mathbf{k};\mathbf{r}]a^+(\mathbf{k})\,|-,-\rangle\}. \qquad (6.4)$$

The normalization constant $\mathcal{N}(r)$ is clearly given by

$$\mathcal{N}(r) = \{|\eta_1(r)|^2 + |\eta_2(r)|^2 + \int d\mathbf{k}\,|u[\bar\omega(r),\mathbf{k};\mathbf{r}]|^2\}^{-1/2}. \qquad (6.5)$$

From the definitions of H_0 and $|W(r);\mathbf{r}_1,\mathbf{r}_2)$ we get

$$[H_0 - W(r)]\,|W(r);\mathbf{r}_1,\mathbf{r}_2) = \mathcal{N}(r)\{[2E-\overline{\omega}(r)][\eta_1(r)\,|+,-\rangle + \eta_2(r)\,|-,+\rangle] + \int d\mathbf{k}[\omega_{\mathbf{k}}-\overline{\omega}(r)] \times u[\overline{\omega}(r),\mathbf{k};\mathbf{r}]a^+(\mathbf{k})\,|-,-\rangle\}. \quad (6.6)$$

There remains to evaluate the effect of the application of the operator $f(H_I)$, with H_I defined by Eq. (5.5), to $|W(r);\mathbf{r}_1,\mathbf{r}_2)$. This is an easy task, once the following formulas (whose detailed proof is given in Appendix C) are established:

$$f_e(H_I)\,|\alpha,-\alpha\rangle = \tfrac{1}{2}\{(f_e[\Lambda_+(r)]+f_e[\Lambda_-(r)])\,|\alpha,-\alpha\rangle + (f_e[\Lambda_+(r)]-f_e[\Lambda_-(r)])\,|-\alpha,\alpha\rangle\}, \quad (6.7)$$

$$f_o(H_I)\,|\alpha,-\alpha\rangle = \tfrac{1}{2}\int d\mathbf{k}\gamma(k)\{\exp(-i\mathbf{k}\cdot\mathbf{r}_1)[c_+(r)+\alpha c_-(r)] + \exp(-i\mathbf{k}\cdot\mathbf{r}_2)[c_+(r)-\alpha c_-(r)]\}a^+(\mathbf{k})\,|-,-\rangle, \quad (6.8)$$

$$f_e(H_I)\int d\mathbf{k}u(\mathbf{k})a^+(\mathbf{k})\,|-,-\rangle = \tfrac{1}{2}\int d\mathbf{k}\gamma(k)\{[\exp(-i\mathbf{k}\cdot\mathbf{r}_1)\chi(\mathbf{r}_1) + \exp(-i\mathbf{k}\cdot\mathbf{r}_2)\chi(\mathbf{r}_2)][b_+(r)+b_-(r)] + [\exp(-i\mathbf{k}\cdot\mathbf{r}_1)\chi(\mathbf{r}_2)+\exp(-i\mathbf{k}\cdot\mathbf{r}_2)\chi(\mathbf{r}_1)] \times[b_+(r)-b_-(r)]\}a^+(\mathbf{k})\,|-,-\rangle, \quad (6.9)$$

with

$$\chi(\mathbf{r}) = \int d\mathbf{k}\gamma(k)\exp(i\mathbf{k}\cdot\mathbf{r})u(\mathbf{k}), \quad (6.10)$$

$$f_o(H_I)\int d\mathbf{k}u(\mathbf{k})a^+(\mathbf{k})\,|-,-\rangle = \tfrac{1}{2}(\{\chi(\mathbf{r}_1)[c_+(r)+c_-(r)]+\chi(\mathbf{r}_2)[c_+(r)-c_-(r)]\}\,|+,-\rangle + \{\chi(\mathbf{r}_1)[c_+(r)-c_-(r)]+\chi(\mathbf{r}_2)[c_+(r)+c_-(r)]\}\,|-,+\rangle). \quad (6.11)$$

The quantities $b_\pm(r)$ and $c_\pm$ that appear in these equations are defined, in close analogy to the constants introduced in Sec. 4, by

$$b_\pm(r) = f_e[\Lambda_\pm(r)]/\Lambda_\pm^2(r), \quad (6.12)$$

$$c_\pm(r) = f_o[\Lambda_\pm(r)]/\Lambda_\pm(r). \quad (6.13)$$

It is also convenient to introduce the quantities

$$\omega_{0\pm}(r) = 2E + f_e[\Lambda_\pm(r)], \quad (6.14)$$

$$\Omega_\pm(r) = \omega_{0\pm}(r) - [c_\pm^2(r)/b_\pm(r)]. \quad (6.15)$$

In writing all these equations we are assuming that $f(0)$ vanishes. Here of course $\Lambda_\pm(r)$ is defined by Eqs. (5.8). Note that these definitions, together with Eqs. (5.10), imply

$$b_\pm(\infty) = b\,, \quad (6.16a)$$

$$c_\pm(\infty) = c\,, \quad (6.16b)$$

$$\omega_{0\pm}(\infty) = \omega_0, \quad (6.16c)$$

$$\Omega_\pm(\infty) = \Omega, \quad (6.16d)$$

the quantities appearing in the rhs of these equations being those introduced in Sec. 3, Eqs. (3.6–3.9).

After this preparation, it is easy to write explicitly the conditions that the Schrödinger equation (6.1) implies for the quantities $\eta_j(r)$ and $u[\overline{\omega}(r),\mathbf{k};\mathbf{r}]$. We find

$$[\tfrac{1}{2}(\omega_{0+}+\omega_{0-})-\overline{\omega}]\eta_1 + \tfrac{1}{2}[f_{e+}-f_{e-}]\eta_2 + \tfrac{1}{2}[c_++c_-]\chi_1 + \tfrac{1}{2}[c_+-c_-]\chi_2 = 0, \quad (6.17a)$$

$$\tfrac{1}{2}[f_{e+}-f_{e-}]\eta_1 + [\tfrac{1}{2}(\omega_{0+}+\omega_{0-})-\overline{\omega}]\eta_2 + \tfrac{1}{2}[c_+-c_-]\chi_1 + \tfrac{1}{2}[c_++c_-]\chi_2 = 0, \quad (6.17b)$$

$$(\omega-\overline{\omega})u(\mathbf{k}) + \tfrac{1}{2}\gamma(k)\{[(c_++c_-)\eta_1+(c_+-c_-)\eta_2+(b_++b_-)\chi_1 + (b_+-b_-)\chi_2]\exp(-i\mathbf{k}\cdot\mathbf{r}_1) + [(c_+-c_-)\eta_1+(c_++c_-)\eta_2 + (b_+-b_-)\chi_1+(b_++b_-)\chi_2]\exp(-i\mathbf{k}\cdot\mathbf{r}_2)\} = 0. \quad (6.18)$$

Here we have omitted to indicate explicitly the dependence of all quantities on $\mathbf{r}$, and we have introduced some additional self-explanatory notational simplifications [such as writing χ_j for $\chi(\mathbf{r}_j)$, ω for $\omega_{\mathbf{k}}$, etc.].

From Eq. (6.18) we can immediately obtain an explicit expression for $u(\mathbf{k})$ in terms of η_1, η_2, χ_1, χ_2, and it is clear that the condition

$$\overline{\omega} < \mu \quad (6.19)$$

must be satisfied in order that $u(\mathbf{k})$ contain no singularity in the integration range. If this condition is satisfied, we can multiply Eq. (6.18) by $\gamma(k)\exp(i\mathbf{k}\mathbf{r}_j)/(\omega-\overline{\omega})$, with $j=1,2$, and then integrate over $d\mathbf{k}$.

In this manner [and using the definitions (6.10) and (5.9)] we get a (homogeneous) system of four linear equations for the determination of the four quantities η_1, η_2, χ_1, χ_2, namely,

$$d_{11}\eta_1 + d_{12}\eta_2 + d_{13}\chi_1 + d_{14}\chi_2 = 0, \quad (6.20a)$$

$$d_{21}\eta_1 + d_{22}\eta_2 + d_{23}\chi_1 + d_{24}\chi_2 = 0, \quad (6.20b)$$

$$d_{31}\eta_1 + d_{32}\eta_2 + d_{33}\chi_1 + d_{34}\chi_2 = 0, \quad (6.20c)$$

$$d_{41}\eta_1 + d_{42}\eta_2 + d_{43}\chi_1 + d_{44}\chi_2 = 0, \quad (6.20d)$$

with the coefficients d_{ij} defined by

$$d_{11} = d_{22} = \tfrac{1}{2}(\omega_{0+}+\omega_{0-}) - \overline{\omega}, \quad (6.21a)$$

$$d_{12} = d_{21} = \tfrac{1}{2}(f_{e+}-f_{e-}), \quad (6.21b)$$

$$d_{13} = d_{24} = \tfrac{1}{2}(c_++c_-), \quad (6.21c)$$

$$d_{14} = d_{23} = \tfrac{1}{2}(c_+-c_-), \quad (6.21d)$$

$$d_{31} = d_{42} = \tfrac{1}{2}(c_+F_+ + c_-F_-), \quad (6.21e)$$

$$d_{32} = d_{41} = \tfrac{1}{2}(c_+F_+ - c_-F_-), \quad (6.21f)$$

$$d_{33} = d_{44} = 1 + \tfrac{1}{2}(b_+F_+ + b_-F_-), \quad (6.21g)$$

$$d_{34} = d_{43} = \tfrac{1}{2}(b_+F_+ - b_-F_-), \quad (6.21h)$$

where, again for simplicity, we have written $F_\pm$ in place of $F_\pm[\overline{\omega}(r);r]$.

It is now convenient to introduce the variables

$$\eta_\pm = \eta_1 \pm \eta_2, \quad (6.22a)$$

$$\chi_\pm = \chi_1 \pm \chi_2, \quad (6.22b)$$

so that

$$\eta_1 = \tfrac{1}{2}(\eta_+ + \eta_-), \quad \eta_2 = \tfrac{1}{2}(\eta_+ - \eta_-), \quad (6.23a)$$

$$\chi_1 = \tfrac{1}{2}(\chi_+ + \chi_-), \quad \chi_2 = \tfrac{1}{2}(\chi_+ - \chi_-). \quad (6.23b)$$

In fact, adding and subtracting the first two and the last two equations of the system (6.20), we get two decoupled systems of two (homogeneous) equations in two variables:

$$(\omega_{0\pm}-\overline{\omega})\eta_\pm + c_\pm \chi_\pm = 0, \tag{6.24a}$$

$$c_\pm F_\pm \eta_\pm + (1 + b_\pm F_\pm = 0, \tag{6.24b}$$

where one must take either the plus sign or the minus sign wherever the double sign appears.

Of course to have a nonvanishing solution for η_1, η_2, χ_1, χ_2 either the determinant D_+ or the determinant D_- must vanish, with the definitions

$$D_\pm = (\omega_{0\pm}-\overline{\omega})(1 + b_\pm F_\pm) - c_\pm^2 F_\pm \tag{6.25a}$$

$$= \omega_{0\pm} - \overline{\omega} + b_\pm(\Omega_\pm - \overline{\omega})F_\pm. \tag{6.25b}$$

Of course if D_+ vanishes (and D_- does not), η_- and χ_- vanish, and vice versa.

Two separate equations for the determination of $\overline{\omega}$ have therefore been obtained, either one of which must be satisfied. They read

$$(\omega_{0\pm}-\overline{\omega})(1 + b_\pm F_\pm) = c_\pm^2 F_\pm, \tag{6.26a}$$

or, equivalently,

$$-[\omega_{0\pm}(r) - \overline{\omega}(r)]/\{b_\pm(r)[\Omega_\pm(r) - \overline{\omega}(r)]\} = F_\pm[\overline{\omega}(r);r]. \tag{6.26b}$$

In this last equation the explicit dependence of all quantities upon r has been reinserted; let us recall that $\omega_{0\pm}(r)$, $\Omega_\pm(r)$, and $b_\pm(r)$ are defined by Eqs. (6.14), (6.15), and (6.12), and $F_\pm(\omega, r)$ by Eqs. (5.9).

The two equations (6.26b) are remarkably similar to the Eq. (3.13) that determines the energy of a single physical V particle. Indeed, since the ω dependence of the functions $F_\pm(\omega;r)$ (for fixed r) is, as noted in Sec. 5, analogous to that of the function $F(\omega)$ of Secs. 2 and 3, the discussion of the two equations (6.26b) can be made in complete analogy to that of Eq. (3.13); and in particular, the graphical display of Figs. 1 and 2 remains relevant (with obvious modifications), as well as the conclusions about the number of solutions [that depends primarily on the sign of $b_+(r)$ and $b_-(r)$, and also on the values of the other parameters, as discussed in Sec. 3].

Let us reemphasize that only solutions occurring in the range $\overline{\omega} < \mu$ yield normalizable solutions of the stationary Schrödinger equation (6.1). On the basis of the analysis outlined above, there can be at most four such solutions. The corresponding wave functions (6.14) can be easily computed solving the system (6.24) and then using Eqs. (6.23) to evaluate η_1, η_2, χ_1, χ_2, and Eq. (6.18) to evaluated $u[\overline{\omega}(r), \mathbf{k};\mathbf{r}]$. In this manner one finds that the (normalized) eigenfunction(s) corresponding to the root(s) $\overline{\omega}$ of Eq. (6.26b) with the + sign have the explicit form

$$\begin{aligned}|NV;r;+\rangle = (&2[c_+(r)]^2 + 4\{b_+(r)[\Omega_+(r) - \overline{\omega}(r)]\}^2 \int d\mathbf{k}\gamma^2(k) \\ &\times \cos^2(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{r})/[\omega_k - \overline{\omega}(r)])^{-1/2} \\ &\times (c_+(r)[|+,-\rangle + |-,+\rangle] + 2\{b_+(r)[\Omega_+(r) \\ &- \overline{\omega}(r)]\} \int d\mathbf{k}\gamma(k)\exp[-(i/2)\mathbf{k}\cdot(\mathbf{r}_1 + \mathbf{r}_2)] \\ &\times \cos(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{r})[\omega_k - \overline{\omega}(r)]^{-1}a^+(\mathbf{k})|-,-\rangle),\end{aligned} \tag{6.27a}$$

whereas the (normalized) eigenfunction(s) corresponding to the root(s) ω of Eq. (6.26b) with the − sign have the form

$$\begin{aligned}|NV;r;-\rangle = (&2[c_-(r)]^2 + 4\{b_-(r)[\Omega_-(r) - \overline{\omega}(r)]\}^2 \int d\mathbf{k}\gamma^2(k) \\ &\times \sin^2(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{r})/[\omega_k - \overline{\omega}(r)])^{-1/2} \\ &\times (c_-(r)[|+,-\rangle - |-,+\rangle] - 2i\{b_-(r)[\Omega_-(r) \\ &- \overline{\omega}(r)]\} \int d\mathbf{k}\gamma(k)\exp[-(i/2)\mathbf{k}(\mathbf{r}_1 + \mathbf{r}_2)] \\ &\times \sin(\tfrac{1}{2}\mathbf{k}\cdot\mathbf{r})[\omega_k - \overline{\omega}(r)]^{-1}a^+(\mathbf{k})|-,-\rangle).\end{aligned} \tag{6.27b}$$

Note that the states (6.27a) are symmetrical under the exchange of the coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$ characterizing the location of the two baryons, while the states (6.27b) are antisymmetrical. Thus, only the former resp. latter should be retained if the baryons where assumed to satisfy Bose resp. Fermi statistics; but since the mass of the baryons is being treated as infinite, so that they can be localized as classical point particles, there is no reason to restrict attention to Fermi, or Bose, statistics, in place of the more general case, corresponding to distinguishable (classical) particles (Boltzmann statistics), and including both symmetrical and antisymmetrical states.

The physical interpretation of these solutions is rather transparent. To discuss this, it is convenient to consider how the situation depends on the distance r between the two baryons.

If this distance is very large, so that $\mu r \gg 1$, the asymptotic formulas (6.16) and (5.10) apply, and therefore Eqs. (6.26), that yield the energies of the physical states NV through Eq. (6.2), coincide with Eq. (3.13), that yields the energy of the (isolated) physical V particles through Eq. (3.5). Therefore, we have

$$W(\infty) = m_N + m_V, \tag{6.28}$$

an equation having an obvious physical meaning: When the two baryons are very far apart, they do not interact, and the energy of the state is just the sum of the energies (masses) of the (isolated) N and the (isolated) V particles. Of course there are two such states, corresponding to the "plus" and "minus" versions of Eq. (6.26b), both of which go into Eq. (3.13) in the large r limit; the corresponding eigenstates are explicitly displayed by Eqs. (6.27), with Eqs. (6.16). Because in the limit of large r these two states have the same energy (6.28), any linear combination of them is also an eigenstate of the Hamiltonian; and, in particular, it is easily seen that the two combinations

$$|NV;\infty\rangle = 2^{-1/2}\{|NV;\infty;+\rangle + |NV;\infty;-\rangle\}, \tag{6.29a}$$

$$|NV;\infty\rangle = 2^{-1/2}\{|NV;\infty;+\rangle - |NV;\infty;-\rangle\}, \tag{6.29b}$$

represent, respectively, an (isolated) N particle localized at $\mathbf{r}_2$ and an (isolated) physical V particle localized at $\mathbf{r}_1$, and an (isolated) N particle localized at $\mathbf{r}_1$ and an (isolated) V particle localized at $\mathbf{r}_2$. Of course of these (stable) states there can exist four, two or zero, depending whether two, one or zero isolated physical V particles exist (as discussed in Sec. 3).

If the physical N and V particles are at a distance apart that is not large relative to the range $1/\mu$, then they interact (through the emission and absorption of virtual θ bosons). This interaction gives rise to a potential energy $V(r)$ depending upon the distance r between the two physical particles, whose magnitude is easily computed from Eqs. (6.26) and (6.2):

$$V(r)=\bar{\omega}(r)-\bar{\omega}(\infty). \tag{6.30}$$

Of course, there is generally a different potential in the even and odd states (even and odd, that is, under the exchange of the coordinates of the two baryons); the potential depends moreover on which one of the two V particles is present (if the parameters of the model are such that two physical particles exist). It may also happen that the number of physical V particles, or rather the number of stable NV states, varies with r; an effect that can be ascribed to the NV potential, that can produce a bound state when it is attractive, or it can dissolve an existing bound state when it is repulsive. All these properties are determined by the eigenvalue equations (6.26), that can be analyzed, and graphically displayed, in close analogy to the treatment of Sec. 4. Note that a separate behavior characterizes the even and odd states. Of course the detailed properties depend on the structure of the functions $F_\pm(\omega;r)$ [that depend on the form factor $\gamma(k)$] and of the quantities $b_\pm(r)$, $\omega_{0\pm}(r)$, and $\Omega_\pm(r)$ [that depend on the function $f(x)$ characterizing the nonlinearity of the model, and also on the form factor $\gamma(k)$].

A detailed discussion of the shape of the potential $V(r)$ acting in the various cases would require a more explicit determination of the function $f(x)$ than we have given thus far. Generally one would find that at large r the potential $V(r)$ vanishes as $\exp(-\mu r)$, where μ is the mass of the θ boson; while its short-range behavior in the various states could be easily inferred from Eqs. (5.10c–f). For some discussion of this point (in the case of the usual Lee model) the interested reader is referred to the book by Baz *et al.*, Ref. 2, and to Ref. 10.

Let us proceed to the study of the scattering of a θ boson on two N particles localized at $\mathbf{r}_1$ and $\mathbf{r}_2$. This problem is easily dealt with on the basis of the previous results. The stationary Schrödinger equation now reads

$$H|\mathbf{k};{}^{\rm in}_{\rm out}\rangle=(2m_N+\omega_{\mathbf{k}})|\mathbf{k};{}^{\rm in}_{\rm out}\rangle, \tag{6.31}$$

where of course the energy $\omega_{\mathbf{k}}$ of the scattering θ boson is larger than μ; and for the scattering eigenfunctions we have the representation

$$|\mathbf{k};{}^{\rm in}_{\rm out}\rangle=\eta_{1,{}^{\rm in}_{\rm out}}|+,-\rangle+\eta_{2,{}^{\rm in}_{\rm out}}|-,+\rangle+a^+(\mathbf{k})|-,-\rangle$$
$$+\int d\mathbf{k}'\bar{u}_{{}^{\rm in}_{\rm out}}(\mathbf{k},\mathbf{k}')a^+(\mathbf{k}')|-,-\rangle. \tag{6.32}$$

The functions $\bar{u}(\mathbf{k},\mathbf{k}')$ are now singular at $\omega_{\mathbf{k}'}=\omega_{\mathbf{k}}$, and the treatment of the singularity characterizes the incoming and the outgoing states. Hereafter, we consider, for notational simplicity, only the incoming case (and we omit indicating explicitly the subscript "in"), and only at the end do we give the results for both cases. Also note that we have omitted indicating explicitly the dependence upon $\mathbf{r}_1$ and $\mathbf{r}_2$.

The determination of η_1, η_2, and $\bar{u}(\mathbf{k},\mathbf{k}')$ can be easily effected noting that they must still satisfy the Eqs. (6.17) and (6.18), with $\bar{\omega}$ replaced by $\omega_{\mathbf{k}}$, ω replaced by $\omega_{\mathbf{k}'}$, $\mathbf{k}$ replaced by $\mathbf{k}'$,

$$u(\mathbf{k})=\delta(\mathbf{k}-\mathbf{k}')+\bar{u}(\mathbf{k},\mathbf{k}'), \tag{6.33}$$

and therefore also

$$\chi_j=\gamma(k)\exp(i\mathbf{k}\mathbf{r}_j)+\bar{\chi}_j \tag{6.34}$$

with

$$\bar{\chi}_j=\int d\mathbf{k}'\gamma(k')\exp(i\mathbf{k}'\cdot\mathbf{r}_j)\bar{u}(\mathbf{k},\mathbf{k}'). \tag{6.35}$$

Therefore, in place of Eq. (6.18) one has

$$\begin{aligned}\bar{u}(\mathbf{k},\mathbf{k}')=(\omega_{\mathbf{k}}-\omega_{\mathbf{k}'}+i\epsilon)^{-1}\tfrac{1}{2}\gamma(k')\{&\gamma(k)[(b_++b_-)\\&\times\exp[i(\mathbf{k}-\mathbf{k}')\mathbf{r}_1]+(b_+-b_-)\exp[i(\mathbf{k}\mathbf{r}_2-\mathbf{k}'\mathbf{r}_1)]\\&+(b_+-b_-)\exp[i(\mathbf{k}\mathbf{r}_1-\mathbf{k}'\mathbf{r}_2)]+(b_++b_-)\\&\times\exp[i(\mathbf{k}-\mathbf{k}')\mathbf{r}_2]]+\exp(-i\mathbf{k}'\mathbf{r}_1)[(c_++c_-)\eta_1\\&+(c_+-c_-)\eta_2+(b_++b_-)\bar{\chi}_1+(b_+-b_-)\bar{\chi}_2]\\&+\exp(-i\mathbf{k}'\mathbf{r}_2)[(c_+-c_-)\eta_1+(c_++c_-)\eta_2\\&+(b_+-b_2)\bar{\chi}_1+(b_++b_-)\bar{\chi}_2]\},\end{aligned} \tag{6.36}$$

and in place of the homogeneous system (6.20) one has the inhomogeneous linear system

$$d_{11}\eta_1+d_{12}\eta_2+d_{13}\bar{\chi}_1+d_{14}\bar{\chi}_2$$
$$=-\tfrac{1}{2}\gamma(k)[(c_++c_-)\exp(i\mathbf{k}\mathbf{r}_1)+(c_+-c_-)\exp(i\mathbf{k}\mathbf{r}_2)], \tag{6.37a}$$

$$d_{21}\eta_1+d_{22}\eta_2+d_{23}\bar{\chi}_1+d_{23}\bar{\chi}_2$$
$$=-\tfrac{1}{2}\gamma(k)[(c_+-c_-)\exp(i\mathbf{k}\mathbf{r}_1)+(c_++c_-)\exp(i\mathbf{k}\mathbf{r}_2)], \tag{6.37b}$$

$$d_{31}\eta_1+d_{32}\eta_2+d_{33}\bar{\chi}_1+d_{34}\bar{\chi}_2$$
$$=-\tfrac{1}{2}\gamma(k)[(b_+F_++b_-F_-)\exp(i\mathbf{k}\mathbf{r}_1)+(b_+F_+-b_-F_-)$$
$$\times\exp(i\mathbf{k}\mathbf{r}_2)], \tag{6.37c}$$

$$d_{41}\eta_1+d_{42}\eta_2+d_{43}\bar{\chi}_1+d_{44}\bar{\chi}_2$$
$$=-\tfrac{1}{2}\gamma(k)[(b_+F_+-b_-F_-)\exp(i\mathbf{k}\mathbf{r}_1)+(b_+F_++b_-F_-)$$
$$\times\exp(i\mathbf{k}\mathbf{r}_2)], \tag{6.37d}$$

where the quantities d_{ij} are defined by Eqs. (6.21), but now $F_\pm$ stands for $F_{\pm,\rm in}(\omega;r)$, defined by

$$F_{\pm,\rm in}(\omega;r)=\int d\mathbf{k}'\gamma^2(k')[1\pm\exp(i\mathbf{k}'\mathbf{r})]/(\omega_{\mathbf{k}'}-\omega-i\epsilon) \tag{6.38a}$$

$$=F_\pm(\omega;r)+i\pi\lambda^2(\omega)\{1\pm[\sin(kr)]/(kr)\}. \tag{6.38b}$$

Here $F_\pm(\omega;r)$ is defined by Eqs. (5.9), and of course $k=(\omega^2-\mu^2)^{1/2}$.

The nonhomogeneous system of linear equations (6.37) allows the computations of η_1, η_2, $\bar{\chi}_1$, and $\bar{\chi}_2$. Once these quantities have been computed, $\bar{u}(\mathbf{k},\mathbf{k}')$ is also known, from Eq. (6.36), and therefore the (incoming) scattering eigenfunction (6.32) is completely and explicitly determined.

To solve the system (6.37) it is convenient to introduce the quantities $\eta_\pm$ and $\bar{\chi}_\pm$, as it was done previously [see Eqs. (6.22–23)]. Then in place of the system (6.37) we get two decoupled systems of two linear equations:

$$(\omega_{0\pm}-\omega_{\mathbf{k}})\eta_\pm+c_\pm\bar{\chi}_\pm=-\gamma(k)c_\pm[\exp(i\mathbf{k}\mathbf{r}_1)\pm\exp(i\mathbf{k}\mathbf{r}_2)], \tag{6.39a}$$

$$c_\pm F_\pm\eta_\pm+(1+b_\pm F_\pm)\bar{\chi}_\pm=-\gamma(k)b_\pm F_\pm[\exp(i\mathbf{k}\mathbf{r}_1)\pm\exp(i\mathbf{k}\mathbf{r}_2)]. \tag{6.39b}$$

Here one must take either the plus sign or the minus sign wherever the double sign appears.

The solution of this system is

$$\eta_\pm = -\gamma(k)c_\pm[\exp(i\mathbf{kr}_1) \pm \exp(i\mathbf{kr}_2)]/D_\pm, \tag{6.40a}$$

$$\bar{\chi}_\pm = -\gamma(k)[b_\pm F_\pm(\omega_{0\pm} - \omega_\mathbf{k}) - c_\pm F_\pm][\exp(i\mathbf{kr}_1) \pm \exp(i\mathbf{kr}_2)]/D_\pm, \tag{6.40b}$$

with $D_\pm$ defined by Eq. (6.25) (with the substitutions previously mentioned). Thus we get

$$\eta_1 = \tfrac{1}{2}\gamma(k)\{[(c_+/D_+) + (c_-/D_-)]\exp(i\mathbf{kr}_1) + [(c_+/D_+) - (c_-/D_-)]\exp(i\mathbf{kr}_2)\}, \tag{6.41a}$$

$$\eta_2 = -\tfrac{1}{2}\gamma(k)\{[(c_+/D_+) - (c_-/D_-)]\exp(i\mathbf{kr}_1) + [(c_+/D_+) + (c_-/D_-)]\exp(i\mathbf{kr}_2)\}, \tag{6.41b}$$

and

$$\bar{u}_{\text{in}}(\mathbf{k},\mathbf{k}') = \bar{v}_{\text{in}}(\mathbf{k},\mathbf{k}')/(\omega_\mathbf{k} - \omega_{\mathbf{k}'} + i\epsilon)$$

with

$$\bar{v}(\mathbf{k},\mathbf{k}') = \gamma(k)\gamma(k')\exp[(i/2)(\mathbf{k}-\mathbf{k}')(\mathbf{r}_1+\mathbf{r}_2)] \times (\{\cos(\tfrac{1}{2}\mathbf{kr})\cos(\tfrac{1}{2}\mathbf{k}'\mathbf{r})]/\{[(\omega_{0+} - \omega_\mathbf{k})/[b_+(\Omega_+ - \omega_\mathbf{k})] + F_{+,\text{in}}(\omega_\mathbf{k};r)\} + [\sin(\tfrac{1}{2}\mathbf{kr})\sin(\tfrac{1}{2}\mathbf{k}'\mathbf{r})]/\{[(\omega_{0-} - \omega_\mathbf{k})/[b_-(\Omega_- - \omega_\mathbf{k})] + F_{-,\text{in}}(\omega_\mathbf{k};r)\}). \tag{6.42}$$

Insertions of these formulas in Eq. (6.32) yields an explicit expression for the incoming scattering state. The outgoing scattering eigenfunction obtains from the incoming one changing everywhere ϵ into $-\epsilon$.

Once the incoming and outgoing scattering states are explicitly known, the scattering amplitude is easily computed from the formula

$$(\mathbf{k}', \text{out}\,|\,\mathbf{k}, \text{in}) = \delta(\mathbf{k}-\mathbf{k}') - 2\pi i\delta(\omega_{\mathbf{k}'} - \omega_\mathbf{k})T(\mathbf{k}',\mathbf{k}). \tag{6.43}$$

Here $T(\mathbf{k}',\mathbf{k})$ is the scattering amplitude from an initial state characterized by the boson momentum $\mathbf{k}$ into the final state characterized by the momentum $\mathbf{k}'$. It is connected to the differential cross section by

$$d\sigma/d\Omega' = (2\pi)^4\omega^2|T(\mathbf{k}',\mathbf{k})|^2, \tag{6.44}$$

where, of course,

$$k = k' = (\omega^2 - \mu^2)^{1/2}. \tag{6.45}$$

From Eq. (6.43), using Eq. (6.41) together with the companion equation for $\bar{u}_{\text{out}}(\mathbf{k},\mathbf{k}')$, one gets

$$T(\mathbf{k}',\mathbf{k}) = \tfrac{1}{2}[\bar{v}_{\text{in}}(\mathbf{k},\mathbf{k}') + \bar{v}^*_{\text{out}}(\mathbf{k}',\mathbf{k})]. \tag{6.46}$$

Inserting in this formula the explicit expression of $\bar{v}_{\text{in}}(\mathbf{k},\mathbf{k}')$, Eq. (6.42), and the analogous expression for $\bar{v}_{\text{out}}(\mathbf{k},\mathbf{k}')$ [obtained replacing $F_{\pm,\text{in}}(\omega;r)$ in Eq. (6.42) with $F_{\pm,\text{out}}(\omega;r) = F^*_{\pm,\text{in}}(\omega;r)$], one gets finally

$$T(\mathbf{k}',\mathbf{k}) = \exp[(i/2)(\mathbf{k}-\mathbf{k}')(\mathbf{r}_1+\mathbf{r}_2)] \times [t_+(\omega;r)\cos(\tfrac{1}{2}\mathbf{kr})\cos(\tfrac{1}{2}\mathbf{k}'\mathbf{r}) + t_-(\omega;r)\sin(\tfrac{1}{2}\mathbf{kr})\sin(\tfrac{1}{2}\mathbf{k}'\mathbf{r})], \tag{6.47}$$

with

$$t_\pm(\omega;r) = \gamma^2(k)/(\{[\omega_{0\pm}(r) - \omega]/[b_\pm(r)(\Omega_\pm(r) - \omega)] + F_\pm(\omega;r) + i\pi\lambda^2(\omega)\{1 \pm [\sin(kr)/(kr)]\}). \tag{6.48}$$

In this formula we have explicitly displayed the dependence of all quantities upon r. Let us recall that $b_\pm(r)$, $\omega_{0\pm}(r)$, and $\Omega_\pm(r)$ are defined by Eqs. (6.12–15), that $F_\pm(\omega;r)$ is defined by Eqs. (6.9), and that $\lambda(\omega)$ is related to the form factor $\gamma(k)$ by Eq. (5.6b).

From the explicit expressions (6.47) and (6.48) one can verify that the scattering amplitude $T(\mathbf{k}',\mathbf{k})$ has all the canonical properties (time reversal, unitarity, causality, i.e., analiticity, correspondence between poles and bound states). It depends of course on the positions $\mathbf{r}_1$ and $\mathbf{r}_2$ of the two N baryons, but only through the difference $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ (except for a phase factor, that does not affect the scattering cross section). Particularly remarkable is the analogy between the present expression of the amplitude describing the scattering of one θ boson over two (fixed) N baryons, with that describing the scattering of one boson over a single N particle (see Sec. 3). Indeed we report here only remarks that are specific to the two-baryon case, referring to the analogy with the treatment of Sec. 3 for all other considerations (including, in particular, any mention of the relations between poles of the scattering amplitudes and bound states and resonances).

The nonspherically symmetrical nature of the scattering target, constituted by the two N baryons located at $\mathbf{r}_1$ and $\mathbf{r}_2$, justifies the remarkable angular dependence of the differential cross section (6.44) implied by the scattering amplitude (6.47); note that there is a dependence both upon the direction of the incident beam, i.e., the direction of $\mathbf{k}$ (relative to the vector $\mathbf{r}$ characterizing the target) and upon the direction of observation, i.e., the direction of $\mathbf{k}'$. Of course if the wavelength $1/k$ associated with the scattering particles is much larger than r, then the nonisotropic nature of the target is not felt by the scattering beam; in fact, in this case the scattering turns out to be spherically symmetrical, namely only S waves contribute to it, and the total cross section reads

$$\sigma(\omega) = 4\pi(\omega^2 - \mu^2)^{-1}[\pi\lambda^2(\omega)]^2 \times [(\{\omega - \omega_{0+}(r)]/\{b_+(r)[\omega - \Omega_+(r)]\} + F(\omega;r))^2 + [2\pi\lambda^2(\omega)]^2]^{-1}; \tag{6.49}$$

an expression that is quite similar to the expression (3.31) of the total cross section for scattering on a single N baryon. Note however that even if the position of the two N baryons coincide, i.e., if $r=0$ (or rather, $\mu r \ll 1$), Eq. (6.49) [where $\omega_{0+}(0)$, $b_+(0)$, and $\Omega_+(0)$ could be evaluated from their definitions (6.12–15) and from Eq. (5.10c), and $F(\omega;0)$ from Eq. (5.10d)] does not quite coincide with Eq. (3.31).

The treatment given in this section applies of course also to the usual Lee model, in which case

$$\omega_{0\pm}(r) = 2E, \quad b_\pm(r) = 0, \quad b_\pm(r)\Omega_\pm(r) = -1. \tag{6.50}$$

The bound state problem in this case had been treated in the literature (see the textbook by Baz *et al.*, Ref. 2), but the scattering case does not appear to have been discussed previously (except in the special case $r = 0$ [10]).

7. CONCLUDING REMARKS AND OUTLOOK

In this paper we have given the explicit exact solutions of the nonlinear Lee model in some sectors of its Hilbert space. The motivations for studying the nonlinear Lee model have been outlined in the Introduction. Here we collect some remarks relative to the results reported above, and we list a number of further problems that are suggested by these findings.

The most remarkable feature of the model under consideration is the possibility of solving it exactly (at least in some sectors of its Hilbert space), in spite of the arbitrariness of its (nonlinear, possibly nonpolynomial) interaction. A characteristic feature is the fact that only the values that the function $f(x)$ (that characterizes the structure of the interaction) takes for certain values of its argument play any role in the solutions; indeed, the results of Sec. 2 [in particular, Eq. (2.12)] imply that, for all sectors with only one baryon present, only the values of $f(x)$ [or rather of its even and odd parts, $f_e(x)$ and $f_o(x)$] at the denumerable and discrete set of points $x=\Lambda s^{1/2}$, s being a nonnegative integer, play any role. This result is confirmed by the explicit complete solution of the simplified model of Sec. 4. However, as soon as sectors of Hilbert space with more than one baryon present are considered, then values of $f_e(x)$ and $f_o(x)$ for other arguments become relevant; indeed, values of these functions for a continuum of determinations of the argument x become relevant, depending on the relative positions of the baryons (that are by assumption fixed in space, but whose relative position is arbitrary).

An interesting open problem is that of ascertaining the conditions that the function $f(x)$ should satisfy in order that the Hamiltonian (1.4) of the nonlinear Lee model be physically sound (and, in particular, possess a spectrum with a lower bound) for all sectors of the Hilbert space, including those with an unlimited number of baryons and/or bosons. It is plausible to conjecture that the conditions found for the simplified model of Sec. 4 are valid also for the general case.

Another interesting problem is to study the nonlinear Lee model if the functions $f_e(x)$ and $f_o(x)$ are not entire, and possibly not even analytic at $x=0$, even though they are finite and well defined for all real values of x. While it is plausible to conjecture that all the results given in this paper would remain valid in this case, the question of giving a precise mathematical definition to $f(H_I)$, and of proving the results, might not be an entirely trivial one.

Of the problems whose study is suggested by the findings reported in this paper, the most obvious one is the treatment of other sectors of the Hilbert space. Particularly interesting is the sector with $Q_1=1$, $Q_2=-1$ ($\theta V - \theta\theta N$ sector), that, already in the case of the usual Lee model, is quite complex and phenomenologically rich.[2] The solution of the nonlinear Lee model in this sector has been obtained in collaboration with A. Degasperis and will be reported in a separate paper.[11]

Another interesting sector that should be solvable after the fashion of Sec. 6 is that characterized by $Q_1=3$, $Q_2=2$ ($NNV-\theta NNN$ sector). One interesting physical problem is the (presumable) presence of three-body forces acting between the baryons, in addition to two-body forces, that should be displayed by the solution in this sector. A more general case worthy of study and that might also be solvable in closed form, is that characterized by $Q_1=n$, $Q_2=n-1$, with n an arbitrary positive integer. The limit of large n would then be particularly interesting. In the case of the usual Lee model, these problems have been investigated by Scarfone,[12] who considered also the $2V$ sector ($Q_1=2$, $Q_2=0$).[13]

Another area that would be interesting to explore, in connection with the solutions given in this paper, and also with those mentioned above, is the version of the nonlinear Lee model in which also the baryons have the proper kinematics, i.e., the model that obtains from that considered here if the assumption that the baryon mass is infinite is dropped. The usual Lee model with this generalization has been discussed (in some sectors of its Hilbert space),[14] and the nonlinear Lee model should also be treatable. Sectors with more than one baryon present, such as that considered in Sec. 6 above, should be particularly interesting. An intermediate approximate approach to this problem, in the spirit of the Born—Oppenheimer approximation, would be to use the results of Sec. 6 of this paper to compute the potential between the N and V baryons, and then study the dynamics of these particles (now with a finite mass) under the effect of this interaction.

On the same line, and certainly very interesting, although probably much too difficult to hope for explicit solutions, would be the study (both from the dynamical and from the statistical—mechanical points of view) of the many-baryon problem, as outlined above, (i.e., considering, to begin with, the sector $Q_1=n$, $Q_2=n-1$).

As we indicated in the introduction, in this paper we have been mainly concerned with the physiological aspects of the Lee model, or rather of its nonlinear generalization. Thus we have introduced a cut off in the interaction term, with the stated purpose to avoid all difficulties with divergent integrals. On the other hand, the Lee model has been mostly studied in connection with the renormalization approach to (ultraviolet) divergences; moreover, the current interest in nonpolynomial field theories is mainly motivated by the hope that the nonpolynomial nature of the interaction will eliminate the divergences that plague relativistic quantum field theory. Thus it is certainly worthwhile to investigate whatever happens to the nonlinear Lee model when the cut off in momentum space is suppressed, and ultraviolet divergences appear. The most interesting question in this connection is: Is it possible that the nonlinear, possibly nonpolynomial, nature of the model be such as to compensate the infinities, so that a finite result emerges even in the local limit, i.e., when no form factor is introduced? This question is answered in a separate paper, written in collaboration with A. Degasperis, where a generalized version of the nonlinear Lee model is discussed from this point of view.[4]

APPENDIX A

In this appendix we derive Eq. (2.12).

It is convenient to consider separately the even and odd parts of $f(H_I)$. Beginning with the even part, let us evaluate

$$(H_I)^{2n} = \alpha\alpha^+\alpha\alpha^+\cdots\alpha\alpha^+P_+ + \alpha^+\alpha\alpha^+\alpha\cdots\alpha^+\alpha P_-, \qquad (\mathrm{A.1})$$

where the operators α and α^+ enter n times and P_+ resp. P_- are the projection operators over the states $|+\rangle$ resp. $|-\rangle$, Eq. (2.13). This formula has been obtained from the definition of H_I, Eq. (2.4), and the remark that

$$\sigma_+^2 = \sigma_-^2 = 0, \quad \sigma_-\sigma_+ = P_-, \quad \sigma_+\sigma_- = P_+. \qquad (\mathrm{A.2})$$

It is now convenient to introduce the coefficients $c_{n,m}$ setting

$$\alpha\alpha^+\alpha\alpha^+\cdots\alpha\alpha^+ = \sum_{m=0}^{n} c_{n,m}\Lambda^{2n-2m}(\alpha^+)^m\alpha^m, \qquad (\mathrm{A.3})$$

where on the left-hand side the operators α and α^+ enter again n times each and where Λ is the quantity defined by Eq. (2.7).

We now note the formula

$$\alpha\alpha^+(\alpha^+)^m\alpha^m = (m+1)\Lambda^2(\alpha^+)^m\alpha^m + (\alpha^+)^{m+1}\alpha^{m+1}, \qquad (\mathrm{A.4})$$

that follows easily from Eq. (2.7).

Using this formula, we obtain from Eq. (1.3) the recursion relations

$$c_{n+1,m} = (m+1)c_{n,m} + c_{n,m-1}, \qquad (\mathrm{A.5})$$

with the conditions

$$c_{n,0} = c_{n,n} = 1. \qquad (\mathrm{A.6})$$

As can be easily shown by direct substitution, these recursion relations are solved by the formula

$$c_{n,m} = (-)^m \sum_{s=0}^{m} \gamma_s(s+1)^n/(m-s)!, \qquad (\mathrm{A.7})$$

the coefficients γ_s being determined by the triangular system of linear equations

$$\sum_{s=0}^{n} \gamma_s(s+1)^n/(n-s)! = (-)^n. \qquad (\mathrm{A.8})$$

This system, in its turn, is solved by the simple formula

$$\gamma_s = (-)^s/s!, \qquad (\mathrm{A.9})$$

as implied by the identity

$$\sum_{s=0}^{n} (-)^s(s+A)^m/[s!(n-s)!] = \begin{cases} 0 & \text{for } m = 0, 1, 2, \ldots, n-1, \\ (-)^n & \text{for } m = n. \end{cases} \qquad (\mathrm{A.10})$$

In this equation A is an arbitrary constant; in our case $A = 1$. Although this identity must be well known, we have not been able to find it in the usual compilations of mathematical formulas, and therefore we provide an explicit proof of it in the following Appendix B.

Thus we may conclude that

$$c_{n,m} = (-)^m \sum_{s=0}^{m} (-)^s(s+1)/[s!(m-s)!], \qquad (\mathrm{A.11})$$

and inserting this expression in Eq. (A.3) we get

$$\alpha\alpha^+\alpha\alpha^+\cdots\alpha\alpha^+ = \sum_{m=0}^{\infty} (-)^m(\alpha^+)^m\alpha^m\Lambda^{2n-2m} \sum_{s=0}^{m} (-)^s(s+1)^n/[s!(m-s)!]. \qquad (\mathrm{A.12})$$

The sum over m extends effectively only up to $m = n$, because for $m > n$ the sum over s vanishes [as implied by Eq. (A.10)].

In a similar fashion it can be shown that

$$\alpha^+\alpha\alpha^+\alpha\cdots\alpha^+\alpha = \sum_{m=0}^{\infty} (-)^m(\alpha^+)^m\alpha^m\Lambda^{2n-2m} \sum_{s=0}^{m} (-)^s s^n/[s!(m-s)!], \qquad (\mathrm{A.13})$$

where again in the left-hand side α and α^+ enter n times each.

From these expressions, and from Eqs. (1.6a) and (A.1), one obtains immediately

$$f_e(H_I) = \sum_{m=0}^{\infty} (-)^m(\alpha^+)^m\alpha^m\Lambda^{2n-2m} \times \sum_{s=0}^{m} (-)^s[s!(m-s)!]^{-1}[P_+f_e(\Lambda(s+1)^{1/2}) + P_-f_e(\Lambda s^{1/2})]. \qquad (\mathrm{A.14})$$

Let us now turn to the odd part of $f(H_I)$. We must evaluate

$$(H_I)^{2n+1} = \alpha\alpha^+\alpha\cdots\alpha^+\alpha\sigma_+ + \mathrm{h.c.}, \qquad (\mathrm{A.15})$$

where in the right-hand side α appears $n+1$ times and α^+ n times.

To obtain this expression we have again used the definition of H_I, Eq. (2.4), and Eqs. (A.2).

It is now easily seen, proceeding just as above, that

$$\alpha\alpha^+\alpha\cdots\alpha^+\alpha = \sum_{m=0}^{n} c_{n,m}\Lambda^{2n-2m}(\alpha^+)^m\alpha^{m+1}, \qquad (\mathrm{A.16})$$

where in the left-hand side α appears $n+1$ times and α^+ n times, and the coefficients $c_{n,m}$ are those already introduced. Using the explicit form (A.11) of these coefficients and proceeding as above, one gets

$$f_o(H_I) = \sum_{m=0}^{\infty} (-)^m[(\alpha^+)^m\alpha^{m+1}\sigma_+ + (\alpha^+)^{m+1}\alpha^m\sigma_-]\Lambda^{-2m} \times \sum_{s=0}^{m} (-)^s[s!(m-s)!]^{-1}f_o(\Lambda(s+1)^{1/2}/\Lambda(s+1)^{1/2}). \qquad (\mathrm{A.17})$$

This equation, together with Eq. (A.15), reproduces Eq. (2.12), that is therefore proved.

It should be emphasized that the rhs of Eqs. (A.15) and (A.17) [and therefore also of Eq. (2.12)] possess generally a finite limit as $\Lambda \to 0$, because if the function $f(x)$ is holomorphic at $x = 0$, the Taylor expansions of expressions like

$$\sum_{s=0}^{m} (-)^s[s!(m-s)!]^{-1}f(x(s+A))$$

begin with a term of order x^m, since the coefficients of the terms x^m with $n < m$ vanish due to Eq. (A.10).

APPENDIX B

In this appendix we provide a proof of Eq. (A.10).

We start from the binomial theorem

$$\sum_{s=0}^{n} (-)^s x^s/[s!(n-s)!] = (-)^n(n!)^{-1}(x-1)^n. \qquad (\mathrm{B.1})$$

We now apply m times, to both sides of this equality, the operator $x^{1-A}(d/dx)x^A$, getting

$$\sum_{s=0}^{n}(-)^s(s+A)^m x^s/[s!(n-s)!]$$
$$=(-)^n(n!)^{-1}[x^{1-A}(d/dx)x^A]^m(x-1)^n. \qquad (B.2)$$

We then set x equal to unity. Then clearly, if $m<n$, the rhs vanishes, and if $m=n$, the rhs reduces to $(-)^n$, because the only term that does not vanish when x is set to unity is the one in which all differentiations have acted on the term $(x-1)^n$. QED

APPENDIX C

In this appendix we prove Eqs. (6.7—11).

We begin writing

$$H_I=H_I(1)+H_I(2), \qquad (C.1)$$

where

$$H_I(j)=\int d\mathbf{k}\gamma(k)\{\exp(i\mathbf{k}\cdot\mathbf{r}_j)a(\mathbf{k})\sigma_+(j)+\text{h.c.}\}. \qquad (C.2)$$

We then note, by explicit computation, that

$$H_I(1)H_I(1)|+,-\rangle=\tfrac{1}{2}[\Lambda_+^2(r)+\Lambda_-^2(r)]|+,-\rangle, \qquad (C.3a)$$
$$H_I(2)H_I(1)|+,-\rangle=\tfrac{1}{2}[\Lambda_+^2(r)-\Lambda_-^2(r)]|+,-\rangle, \qquad (C.3b)$$
$$H_I(1)H_I(2)|+,-\rangle=0, \qquad (C.3c)$$
$$H_I(2)H_I(2)|+,-\rangle=0, \qquad (C.3d)$$

with $\Lambda_\pm^2(r)$ defined by Eqs. (5.8). These equations imply

$$(H_I)^2|+,-\rangle=\tfrac{1}{2}\{[\Lambda_+^2(r)+\Lambda_-^2(r)]|+,-\rangle$$
$$+[\Lambda_+^2(r)-\Lambda_-^2(r)]|-,+\rangle\}, \qquad (C.4)$$

and, more generally, by symmetry

$$(H_I)^2|\alpha,-\alpha\rangle=\tfrac{1}{2}\{[\Lambda_+^2(r)+\Lambda_-^2(r)]|\alpha,-\alpha\rangle$$
$$+[\Lambda_+^2(r)-\Lambda_-^2(r)]|-\alpha,\alpha\rangle\}. \qquad (C.5)$$

Here, and in the following, α stands for + or −.

This equation implies that

$$(H_I)^{2n}|\alpha,-\alpha\rangle=\gamma_{n,+}(r)|\alpha,-\alpha\rangle+\gamma_{n,-}(r)|-\alpha,\alpha\rangle, \qquad (C.6)$$

with

$$\gamma_{0,+}(r)=1,\quad \gamma_{0,-}(r)=0, \qquad (C.7)$$

and

$$\gamma_{n+1,\alpha}(r)=\tfrac{1}{2}\{[\Lambda_+^2(r)+\Lambda_-^2(r)]\gamma_{n,\alpha}(r)$$
$$+[\Lambda_+^2(r)-\Lambda_-^2(r)]\gamma_{n,-\alpha}(r)\}. \qquad (C.8)$$

To solve these recursion relations we introduce

$$\Gamma_{n,\pm}(r)=\gamma_{n,+}(r)\pm\gamma_{n,-}(r), \qquad (C.9a)$$

so that

$$\gamma_{n,\pm}(r)=\tfrac{1}{2}[\Gamma_{n,+}(r)\pm\Gamma_{n,-}(r)]. \qquad (C.9b)$$

Then the recursion relations (C.8) become simply

$$\Gamma_{n+1,\pm}(r)=\Lambda_\pm^2(r)\Gamma_{n,\pm}(r), \qquad (C.10)$$

and the boundary conditions (C.7) become

$$\Gamma_{0,\pm}(r)=1. \qquad (C.11)$$

The quantities $\Gamma_{n,\pm}(r)$ are therefore given immediately by the simple formula

$$\Gamma_{n,\pm}(r)=[\Lambda_\pm^2(r)]^n, \qquad (C.12)$$

and therefore Eq. (C.9b), together with Eqs. (C.6) and (1.6a), immediately imply Eq. (6.7), namely the first equation we had to prove.

The proof of the subsequent equations is now easy. From Eqs. (C.6) it follows immediately that

$$(H_I)^{2n+1}|+,-\rangle=H_I\{\gamma_{n,+}(r)|+,-\rangle+\gamma_{n,-}(r)|-,+\rangle\} \qquad (C13a)$$
$$=\int d\mathbf{k}\gamma(k)\{\gamma_{n,+}(r)\exp(-i\mathbf{k}\cdot\mathbf{r}_1)$$
$$+\gamma_{n,-}(r)\exp(-i\mathbf{k}\mathbf{r}_2)\}a^+(\mathbf{k})|-,-\rangle, \qquad (C.13b)$$

and using the explicit expression of $\gamma_{n,\pm}$ just found, Eqs. (C.9b) and (C.12), and Eq. (1.6b), one obtains Eq. (6.8) (with $\alpha=+$). In a completely analogous fashion one proves Eq. (6.8) for $\alpha=-$.

To prove Eq. (6.11), one notices by direct computation that

$$H_I\int d\mathbf{k}u(\mathbf{k})a^+(\mathbf{k})|-,-\rangle=\chi(\mathbf{r}_1)|+,-\rangle+\chi(\mathbf{r}_2)|-,+\rangle, \qquad (C.14)$$

with $\chi(\mathbf{r})$ defined by Eq. (6.10). Therefore,

$$(H_I)^{2n+1}\int d\mathbf{k}u(\mathbf{k})a^+(\mathbf{k})|-,-\rangle$$
$$=[\chi(\mathbf{r}_1)\gamma_{n,+}(r)+\chi(\mathbf{r}_2)\gamma_{n,-}(r)]|+,-\rangle$$
$$+[\chi(\mathbf{r}_1)\gamma_{n,-}(r)+\chi(\mathbf{r}_2)\gamma_{n,+}(r)]|-,+\rangle, \qquad (C.15)$$

where we have used Eq. (C.6). Using the explicit form of $\gamma_{n,\pm}(r)$ and Eq. (1.6b), Eq. (6.11) follows.

Finally one uses Eqs. (C.14), (C.13) (with $n-1$ in place of n) and the analogous equation with 1 and 2 exchanged, to get

$$(H_I)^{2n}\int d\mathbf{k}u(\mathbf{k})a^+(\mathbf{k})|-,-\rangle$$
$$=\int d\mathbf{k}\gamma(k)\{\gamma_{n-1,+}(r)[\chi(\mathbf{r}_1)\exp(i\mathbf{k}\mathbf{r}_1)+\chi(\mathbf{r}_2)\exp(-i\mathbf{k}\mathbf{r}_2)]$$
$$+\gamma_{n-1,-}(r)[\chi(\mathbf{r}_1)\exp(-i\mathbf{k}\mathbf{r}_2)$$
$$+\chi(\mathbf{r}_2)\exp(-i\mathbf{k}\mathbf{r}_1)]\}a^+(\mathbf{k})|-,-\rangle. \qquad (C.16)$$

Inserting here the explicit form of $\gamma_{n,\pm}(r)$, and using Eq. (1.6a) [also recalling that $f(0)$ vanishes by assumption], Eq. (6.9) obtains.

[1]T. D. Lee, Phys. Rev. **95**, 1329 (1954).

[2]See the following textbooks, and the references quoted there: S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row Peterson, Evanston, 1961), Sec. 12b; E. M. Henley and W. Thirring, *Elementary Quantum Field Theory* (McGraw-Hill, New York, 1962), Chapts. 13, 14; A. I. Baz, Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decay in Nonrelativistic Quantum Mechanics* (Nauka, Moscow, 1966), Chap. 8 (English translation: Israel Program for Scientific Translation, Jerusalem, 1969; in the second Russian edition, Nauka, Moscow, 1971, the Lee Model is treated in Chap. 11, but with no additions relative to the first edition). The following (more recent) papers focus mainly on problems connected with the renormalization of the Lee model without cutoff (or of various modified versions of it): M. S. Maxon and R. B. Curtis, Phys. Rev. **137**, B996 (1965); F. J. Yndurain, J. Math. Phys. **7**, 1133 (1966); J. M. Lévy-Leblond, Commun. Math. Phys. **4**, 157 (1967); R. A. Brandt and A. Orzalesi, Phys. Rev. **162**, 1747 (1967); R. L. Zimmerman, Nuovo Cimento A **56**, 15 (1968); M. M. Broido, J. Math. Phys. **9**, 510 (1968); R. Schrader,

Commun. Math. Phys. **10**, 155 (1968); C. L. Hammer and T. A. Weber, J. Math. Phys. **10**, 2067 (1969); W. Dittrich, Nuovo Cimento A **63**, 85 (1969); R. Collina and G. Luzzatto, Nuovo Cimento A **65**, 277 (1970); Q. Bui-Duy, Nuovo Cimento A **66**, 313 (1970); Nuovo Cimento A **6**, 163 (1971); D. Grosser, Nuovo Cimento A **70**, 213 (1970); F. Esposito and V. Esposito, Nuovo Cimento A **6**, 277 (1971). The first solution of the problem for the $\theta V - \theta\theta N$ sector (whose solution is not given in the textbooks quoted above, nor is discussed in this paper) is due to R. D. Amado, Phys. Rev. **122**, 696 (1961). A list of subsequent papers dealing with this problem (especially with its mathematical aspects) includes: R. P. Kenshaft and R. D. Amado, J. Math. Phys. **5**, 1340 (1964); A. Pagnamenta, J. Math. Phys. **6**, 955 (1965); Ann. Phys. (N.Y.) **39**, 453 (1966); J. Math. Phys. **7**, 356 (1966); J. C. Houard, Ann. Inst. Henri Poincaré A **2**, 105 (1965); C. M. Sommerfield, J. Math. Phys. **6**, 1170 (1965); E. Kazes, J. Math. Phys. **6**, 1772 (1965); M. T. Vaughn, Nuovo Cimento A **40**, 803 (1965); J. B. Bronzan, Phys. Rev. **139**, B751 (1965); T. Muta, Prog. Theor. Phys. **33**, 666 (1965); P. K. Srivastava and R. S. Choudhoury, Nuovo Cimento **39**, 650 (1965); M. S. Maxon, Phys. Rev. **149**, 1273 (1966); H. M. Fried, J. Math. Phys. **7**, 583 (1966); J. B. Bronzan, M. Cassandro, and M. T. Vaughn, Nuovo Cimento **46**, 128 (1966).

[3]F. Calogero, preprints 340 and 356, Istituto di Fisica, Rome University (unpublished). The present paper is a revised version of the second preprint, mainly to take account of the paper by Marr and Shimamoto, Ref. 5. I wish to thank the JMP referee for pointing out the existence of this paper (and of that by Barton, Ref. 8), that had escaped my notice.

[4]F. Calogero and A. Degasperis, Lett. Nuovo Cimento 4, 661 (1972).

[5]R. B. Marr and Y. Shimamoto, Nuovo Cimento **31**, 599 (1964). Other papers discussing versions of the (usual) Lee model with more than one V particle can be traced from the references given in this paper.

[6]F. Calogero, Nuovo Cimento **24**, 614 (1962).

[7]Recall that we are assuming $f(0) = 0$.

[8]G. Barton, Nuovo Cimento **17**, 864 (1960).

[9]F. Calogero, Nuovo Cimento **30**, 916 (1963).

[10]L. M. Scarfone, J. Math. Phys. **7**, 159 (1966).

[11]F. Calogero and A. Degasperis (to be published).

[12]L. M. Scarfone, Nuovo Cimento **23**, 891 (1962); Nuovo Cimento **24**, 480 (1962).

[13]L. M. Scarfone, J. Math. Phys. **9**, 246 (1968).

[14]L. Van Hove, Physica (The Hague) **21**, 901 (1955); Physica (The Hague) **22**, 343 (1956); Physica (The Hague) **23**, 441 (1957); see also the papers by F. J. Yndurain and by M. M. Broido, Ref. 2.

A comparison of two transformation theories of classical mechanics

J. Rae and R. Davidson*

Faculté des Sciences, Université Libre de Bruxelles, 1050 Bruxelles, Belgium

(Received 1 November 1971; revised manuscript received 22 February 1972)

A comparison is made of two transformation theories which can be used in classical mechanics: the averaging method as generalized by Kruskal and the superoperator transformation theory used in statistical mechanics by Prigogine *et al.* For the class of systems considered, a striking connection is found which, on the one hand, illustrates some of the general features of the superoperator method and, on the other, provides an interesting method for calculating invariants of nearly periodic systems. This latter method is shown to be equivalent to, but more systematic than, that developed by McNamara and Whiteman.

I. INTRODUCTION

In recent years there has been introduced by the Brussels school a transformation theory for use in nonequilibrium statistical mechanics.[1] The theory is expressed in terms of superoperators which act on phase functions or density matrices, and an important role is played by the notion of subdynamics, the time evolution of an appropriate projection of a phase function or density matrix.[1,2] Now it is of some interest to see how such a theory is related to other transformation theories existing in classical or quantum mechanics, and some progress in this direction has been achieved so that connections are now known with the usual quantum–mechanical transformation theory for systems with a discrete spectrum[3–5] and with Hamilton–Jacobi theory for classical systems describable in action angle variables.[6] However, these systems are, in a sense, too simple since in these cases the projection operator for subdynamics projects out constants of the motion, and the subdynamics thereby become trivial. The purpose of the present paper is to establish connection with another transformation theory of classical mechanics, the "averaging method" as generalized by Kruskal,[7] and show that this, while still a special case, has more content than the simple examples mentioned above. In this way, we have a nontrivial realization of the superoperator theory which exemplifies some of the general features and, reciprocally, the results of the Brussels school have implications for the "averaging method" leading, for example, to a particularly convenient way of calculating invariants.

In Sec. 2 below we outline the two transformation theories to be compared and show the possibility of connecting them. This is followed by a proof of the connection by a perturbation series method. Sec. 5 is a brief discussion on the nature of the approximations involved, an important prerequisite for Sec. 6 in which we treat our class of systems as an example of the generalized transformation theory and produce some of the more important superoperators. Finally, in Sec. 7, we describe a simple method of calculating invariants and compare it with methods previously used.[8]

II. THE TRANSFORMATION THEORIES AND THE POSSIBILITY OF CONNECTION

In the transformation theory of Prigogine and his coworkers, mechanical systems are described in terms of phase functions of the canonical coordinates and momenta, viewed as elements of some suitable linear space, which evolve in time according to the Liouville equation

$$i\frac{\partial f}{\partial t}(q_1\cdots q_N, p_1\cdots p_N) = \mathbf{L}f \equiv i\{H, f\}, \tag{2.1}$$

where $\{,\}$ is the Poisson bracket.

Since the operator $\mathbf{L}$ involves all N degrees of freedom, in a macroscopic system it is hopeless to look for a solution of this equation as it stands. One common approach to this problem is to introduce reduced distribution functions, obtain their equations of motion from (2.1), and try to solve the resulting system of equations. This leads rapidly to nonlinear equations and incredibly difficult mathematical problems. The other approach, of interest here, is to attempt to extract particularly relevant parts of (2.1) but always maintaining linear equations. The techniques for doing this are well described elsewhere,[2,9] so we simply mention briefly that one decomposes $\mathbf{L}$ into $\mathbf{L}_0 + \epsilon\mathbf{L}_1$, where $\epsilon\mathbf{L}_1$ is supposed to be small in some sense, and makes a related decomposition of the phase function f by use of the projector $\mathbf{P}$, projecting onto the closure of the nullspace of $\mathbf{L}_0$, along with its complement $\mathbf{Q} = 1 - \mathbf{P}$. One then introduces further linear operators, defined in terms of $\mathbf{L}$ and $\mathbf{P}$, which allow the appropriate part of the time evolution to be described by a kinetic equation[1,2,9]

$$i\frac{\partial \mathbf{P}f}{\partial t} = \mathbf{\Omega}\psi\mathbf{P}f \tag{2.2}$$

for the $\mathbf{P}$ part of the phase funtion. The $\mathbf{Q}$ part is then given by

$$\mathbf{Q}f = \mathbf{C}\mathbf{P}f \tag{2.3}$$

and here we are employing the linear operators $\mathbf{\Omega}\psi$ and $\mathbf{C}$, respectively called the collision and creation operators. It has been shown that certain undesirable features of (2.2) (e.g. nonhermiticity of $\mathbf{\Omega}\psi$) can be removed if one works not with $\mathbf{P}f$ but with a transformed phase function f_R related to $\mathbf{P}f$ by

$$f_R(t) = \chi^{-1}\mathbf{P}f(t), \tag{2.4}$$

in terms of which the evolution equation becomes

$$i\frac{\partial f_R(t)}{\partial t} = \phi f_R(t) \tag{2.5}$$

with a Hermitian collision operator ϕ. The linear operator χ^{-1} which performs this useful transformation

Copyright © 1973 American Institute of Physics

is the solution of the Mandel–Turner equation[1,10] and can be written to any desired order as a formal series in the parameter ϵ.[4]

The second transformation theory we shall use applies to certain types of differential equations and has been developed by Kruskal[7] following the averaging method of Krylov and Bogoliubov. For the simplest case the equations are of the type[8]

$$\frac{d\nu}{dt}=1+\epsilon f(\nu,y,\epsilon),\qquad \frac{dy}{dt}=\epsilon \mathbf{g}(\nu,y,\epsilon), \tag{2.6}$$

where $y=(y_1,\ldots,y_N)$ and f, g are periodic in ν with period τ and $O(1)$ for the small parameter ϵ. Equations (2.6) describe a slow drift imposed on an oscillatory motion and the idea is to transform to new variables ϕ, z:

$$\phi=\Phi(\nu,y),\quad \nu=N(\phi,\mathbf{z}),\qquad \mathbf{z}=\mathbf{Z}(\nu,y),\quad \mathbf{y}=\mathbf{Y}(\phi,\mathbf{z}), \tag{2.7}$$

which separate the oscillation from the drift to give equations of motion

$$\frac{d\phi}{dt}=1+\epsilon\omega(\mathbf{z},\epsilon),\qquad \frac{d\mathbf{z}}{dt}=\epsilon\mathbf{h}(\mathbf{z},\epsilon). \tag{2.8}$$

We also require that ϕ be an anglelike variable and that $\mathbf{z}$ be periodic in ν, i.e.,

$$\Phi(\nu+\tau,\mathbf{y})=\Phi(\nu,\mathbf{y})+\tau,\qquad \mathbf{Z}(\nu+\tau,\mathbf{y})=\mathbf{Z}(\nu,\mathbf{y}). \tag{2.9}$$

Kruskal has shown how starting from ν, $\mathbf{y}$, it is possible to find "nice" variables ϕ, $\mathbf{z}$, satisfying (2.8), (2.9), as formal power series in ϵ. The solutions of (2.8) are asymptotically correct solutions of (2.6) in the following sense. One solves Eqs. (2.8) truncated at order ϵ^n say, and converts the solution to $\nu^{(n)}$, $\mathbf{y}^{(n)}$ in ν, $\mathbf{y}$ space by a correspondingly truncated version of the inverse transformation (2.7). It has been proved[7] that if ν^*, $\mathbf{y}^*$ are solutions of the original Eqs. (2.6) with the same initial conditions (to order ϵ^n) as $\nu^{(n)}$, $y^{(n)}$ then

$$\nu^*-\nu^{(n)}=O(\epsilon^{n+1}),\quad y^*-y^{(n)}=O(\epsilon^{n+1}) \tag{2.10}$$

for times in a range of order $1/\epsilon$. Kruskal has further shown, and we shall exploit this, that if the original Eqs. (2.6) are in Hamiltonian form, the transformation to "nice" coordinates can be taken as a canonical transformation.

The transformations just described are point transformations of R^{N+1} taking ν, $\mathbf{y}$, defined in some domain, to ϕ, $\mathbf{z}$, whereas the transformation used by the Brussels school is of phase functions. For definiteness in making the connection, let us consider only those phase functions defined as follows. We take $\mathcal{D}$ as a definite connected set in R^N and use phase functions $f\in F(\mathcal{D})$, the set of real functions on $R\times\mathcal{D}$, which are continuous and periodic in their "angle" variable $\alpha\in R$ and, for fixed α, are restrictions to $\mathcal{D}$ of functions which are bounded and continuous on some open neighborhood $\mathcal{D}_\delta$ of $\mathcal{D}$. Now if the functions appearing in (2.6) belong to $F(\mathcal{D})$ so do the functions in (2.7) (or at least their coefficients in an ϵ expansion have this property), and it is obvious that for small ϵ the Kruskal transformation takes $\mathcal{D}_\delta$ to another open neighborhood of $\mathcal{D}$. With each transformation $(\nu,\mathbf{y})\to(\phi,\mathbf{z})$ we can now associate a linear operator $\mathbf{V}$: $F(\mathcal{D})\to F(\mathcal{D})$ by

$$[\mathbf{V}f](\phi,\mathbf{z})=[f](\nu,\mathbf{y}), \tag{2.11}$$

so that $\mathbf{V}f$ is a function defined on a neighborhood of $\mathcal{D}$, and hence on $\mathcal{D}$. Since the point transformation is invertible so is the operator $\mathbf{V}$ and we have

$$[\mathbf{V}^{-1}g](\nu,\mathbf{y})=[g](\phi,\mathbf{z}). \tag{2.12}$$

In order to link the Mandel–Turner transformation to $\mathbf{V}$, we are obliged to introduce several more linear operators. Following the lines of an argument used in the quantum case,[5] we may write for $\mathbf{L}$, the Liouville operator defined in $F(\mathcal{D})$ through (2.6),

$$\int_c \frac{1}{z^1-\mathbf{L}}f\,dz^1=\mathbf{V}^{-1}\int_c\frac{1}{z'-\mathbf{VLV}^{-1}}\mathbf{V}f\,dz^1$$
$$=\mathbf{V}^{-1}\int_c\sum_{m=0}^{\infty}\frac{1}{z^1-\mathbf{VL}_0\mathbf{V}^{-1}}\left(\epsilon\mathbf{VL}_1\mathbf{V}^{-1}\frac{1}{z^1-\mathbf{VL}_0\mathbf{V}^{-1}}\right)^m\mathbf{V}f\,dz^1, \tag{2.13}$$

where C is any contour avoiding the singularities and we have used the usual resolvent expansion. Now the operator

$$\mathbf{VLV}^{-1}=-i(1+\epsilon\omega)\frac{\partial}{\partial\phi}-i\epsilon\mathbf{h}\cdot\frac{\partial}{\partial\mathbf{z}}$$

plays the role of the Liouville operator in coordinates ϕ, $\mathbf{z}$ and governs the motion given by (2.8) in which the angle variable is well separated from the others. We therefore introduce the projection operator defined on $F(\mathcal{D})$ by

$$[\mathbf{P}g](\mathbf{z})=\frac{1}{\tau}\int_0^{\tau}[g](\phi,\mathbf{z})d\phi, \tag{2.14}$$

which is the nullspace projector for $\mathbf{VL}_0\mathbf{V}^{-1}$ and use this to decompose $\mathbf{V}f$ in (2.13) into $\mathbf{PV}f+(1-\mathbf{P})\mathbf{V}f$. We are now in a position to define the fundamental projection operator Π: $F(\mathcal{D})\to F(\mathcal{D})$ as usual[11] by taking the contour C on the right-hand side of (2.13) to be a small circle γ_0 in the neighborhood of $z^1=0$. Since $\mathbf{VL}_0\mathbf{V}^{-1}$ (which is in fact $-i\partial/\partial\phi$) has a discrete spectrum there is no difficulty in doing this. Thus

$$\Pi f=\mathbf{V}^{-1}\int_{\gamma_0}\sum_{n=0}^{\infty}\frac{1}{z^1-\mathbf{VL}_0\mathbf{V}^{-1}}\left(\epsilon\mathbf{VL}_0\mathbf{V}^{-1}\frac{1}{z^1-\mathbf{VL}_0\mathbf{V}^{-1}}\right)^n \times[\mathbf{PV}f+(1-\mathbf{P})\mathbf{V}f]dz^1. \tag{2.15}$$

We now observe from (2.8) that $\mathbf{VL}_1\mathbf{V}^{-1}$ has no ϕ-dependence other than a term containing $\partial/\partial\phi$ on the extreme right. It follows that the contribution of $(1-\mathbf{P})\mathbf{V}f$ to (2.15) is zero since all the propagators assume the form $1/(z^1 - 2\pi n/\tau)$, $n \neq 0$, and vanish on integration. It is also easy to see that the only nonzero contribution coming from $\mathbf{PV}f$ is from the term $n=0$. In this way we have proved

$$\Pi f = \mathbf{V}^{-1}\int_{\gamma_0}\frac{1}{z^1}\mathbf{PV}f dz^1 = \mathbf{V}^{-1}\mathbf{PV}f. \tag{2.16}$$

The transformation $\mathbf{V}$ is defined on the whole of $F(\mathcal{D})$ and is to be compared with the transformation Λ_p^{-1} of Eq. (2.6) of Ref. 1b, whereas the Mandel–Turner transformation is defined only on the $\mathbf{P}$-projected part of $F(\mathcal{D})$. Following Ref. 1b, we have that the transformation $\mathbf{PV} = \mathbf{V}\Pi$ (corresponding to $\Lambda^{-1}\,\Pi$ of Ref. 1) will give rise to the Mandel–Turner transformation if and only if it satisfies the equation

$$\frac{\partial \mathbf{PV}}{\partial \epsilon} = \mathbf{PV}\frac{\partial \Pi}{\partial \epsilon} \tag{2.17}$$

or, in view of (2.16),

$$\mathbf{PV}\frac{\partial \mathbf{V}^{-1}}{\partial \epsilon}\mathbf{P} = 0. \tag{2.18}$$

We thus have the following position. Kruskal's transformations (2.7) provide us through (2.11) with operators $\mathbf{V}$ acting on phase functions and a given $\mathbf{V}$ gives a Mandel–Turner transformation if and only if (2.18) is satisfied. In the next section we show that for the class of systems considered there is an essentially unique canonical transformation to "nice" variables which satisfies (2.18).

III. PROOF OF EQUIVALENCE

In the remainder of this paper we consider only the simplest systems to which Kruskal's method may be applied,[8] those described by a Hamiltonian of the form

$$H = I + \epsilon\Omega(\alpha, I, q_1 \cdots q_N, p_1 \cdots p_N), \tag{3.1}$$

where $\alpha, I; q, p$ are canonically conjugate variables, Ω is periodic in α with period τ, and ϵ is a (small) parameter. It is worth pointing out that by canonical transformation a large class of Hamiltonians may be put into the form (3.1): in particular those of the form

$$H = \sum_{i=1}^{n}\omega_i J_i + \epsilon V(\alpha_1 \cdots \alpha_N, J_1 \cdots J_N), \tag{3.2}$$

where α, J are action-angle variables, V is periodic in the α's with period 1, and the ratio ω_i/ω_j is rational for all i and j.

The equations of motion from (3.1) are of the form (2.6) so we commence by finding the canonical changes of variable which make the new Hamiltonian independent of the new angle variable and which are, therefore, changes to "nice" variables in Kruskal's sense. Suppose the transformation takes $\alpha, I, \mathbf{q}, \mathbf{p}$ to variables $\beta, J, \mathbf{Q}, \mathbf{P}$ by means of a generating function $S(\alpha, J, \mathbf{q}, \mathbf{P}, \epsilon)$. Then

$$\beta = \frac{\partial S}{\partial J}, \quad I = \frac{\partial S}{\partial \alpha}, \quad \mathbf{Q} = \frac{\partial S}{\partial \mathbf{P}}, \quad \mathbf{p} = \frac{\partial S}{\partial \mathbf{q}}. \tag{3.3}$$

The transformations we seek are of the form

$\beta = \alpha +$ a function periodic in α

and similarly for $J, \mathbf{Q}, \mathbf{P}$ which implies that S has the form

$$S = \alpha J + \mathbf{q}\cdot\mathbf{P} + \sum_{n=1}\epsilon^n S^{(n)}(\alpha, J, \mathbf{q}, \mathbf{P}) \tag{3.4}$$

with each $S^{(n)}$ periodic in α with period τ. Since S is independent of time, the old and new Hamiltonians H and K take the same value at a given phase point, i.e.,

$$K(J, \mathbf{Q}, \mathbf{P}, \epsilon) = H(\alpha, I, \mathbf{q}, \mathbf{p}.\epsilon).$$

In the independent variables $\alpha, J, \mathbf{q}, \mathbf{P}, \epsilon$ this gives

$$K\left(J, \frac{\partial S}{\partial \mathbf{P}}, \mathbf{P}, \epsilon\right) = \frac{\partial S}{\partial \alpha} + \epsilon\Omega\left(\alpha, \frac{\partial S}{\partial \alpha}, \mathbf{q}, \frac{\partial S}{\partial \mathbf{q}}\right) \tag{3.5}$$

which is the basic equation to be solved for K and S. This is done by expanding $K(J, \mathbf{q}, \mathbf{P}, \epsilon) = \Sigma_{n=0}\epsilon^n K^{(n)}(J, \mathbf{q}, \mathbf{P})$ and solving (3.5) order by order in ϵ:

Zeroth order

(3.5) gives

$$K^{(0)}(J, \mathbf{q}, \mathbf{P}) = \frac{\partial S^{(0)}}{\partial \alpha},$$

whence

$$K^{(0)} = J, \tag{3.6}$$

$$S^{(0)} = \alpha J + \mathbf{q}\cdot\mathbf{P}. \tag{3.7}$$

First order

(3.5) gives

$$K^{(1)}(J, \mathbf{q}, \mathbf{P}) = \frac{\partial S^{(1)}}{\partial \alpha} + \Omega.$$

$K^{(1)}$ is determined by integrating over a period in α and $S^{(1)}$ from the remaining oscillatory part as

$$K^{(1)}(J, \mathbf{q}, \mathbf{P}) = \frac{1}{\tau}\int_0^\tau \Omega(\alpha, J, \mathbf{q}, \mathbf{P})d\alpha \equiv \bar{\Omega}(J, \mathbf{q}, \mathbf{P}), \tag{3.8}$$

$$S^{(1)} = -\int^\alpha (\Omega - \bar{\Omega})d\alpha + F^{(1)}(J, \mathbf{q}, \mathbf{P}) \tag{3.9}$$

with $F^{(1)}$ arbitrary.

General term

$$K^{(n)} = \frac{\partial S^{(n)}}{\partial \alpha} - \frac{\partial K^{(1)}}{\partial \mathbf{q}}\cdot\frac{\partial S^{(n-1)}}{\partial \mathbf{P}} + \frac{\partial \Omega}{\partial J}\frac{\partial S^{(n-1)}}{\partial \alpha} + \frac{\partial \Omega}{\partial \mathbf{P}}\cdot\frac{\partial S^{(n-1)}}{\partial \mathbf{q}} \tag{3.10}$$

$$+ \text{terms involving } K^{(r)},\ S^{(t)} \text{ with } 1 < r < n,\quad t < n-1.$$

Thus we can solve for $K^{(n)}$ in terms of lower order expressions and for $S^{(n)}$, which takes the form

$$\int^\alpha d\alpha\,\{\text{expression involving lower orders}\} + F^{(n)}(J, \mathbf{q}, \mathbf{P})$$

with $F^{(n)}$ arbitrary.

In this way, we can determine, through S, all possible sets of "nice" canonical coordinates and we see that these differ only in the choice of the functions $F^{(n)}(J, \mathbf{q}, \mathbf{P})$.

We have constructed the variable β to be cyclic for the new Hamiltonian K so that the conjugate momentum J is a constant of the motion. It is not obvious from the above that each choice of "nice" canonical coordinates gives the same J but we shall now show that this is, in fact, the case.

Suppose that $(\beta,J,\mathbf{Q},\mathbf{P})$, $(\beta^*,J^*,\mathbf{Q}^*,\mathbf{P}^*)$ are two sets of coordinates determined as above with different choices of $F^{(n)}$. Following Kruskal's terminology,[7] a $(J,\mathbf{Q},\mathbf{P})$-ring is the set of points $(\beta,J,\mathbf{Q},\mathbf{P})$ with $(J,\mathbf{Q},\mathbf{P})$ fixed and β varying from 0 to τ. It can be shown (Sec. C.3 of Ref. 7) that a $(J,\mathbf{Q},\mathbf{P})$-ring is a $(J^*,\mathbf{Q}^*,\mathbf{P}^*)$-ring. For a definite ring then

$$\tau J=\oint J d\beta+\mathbf{P}\cdot d\mathbf{Q}=\oint J^* d\beta^*+\mathbf{P}^*\cdot d\mathbf{Q}^*=\tau J^*, \qquad (3.11)$$

where $\oint$ means integrate round the ring and the second equality follows since the two sets of variables are canonical.[12] This shows that J and J^* take the same values.

A lengthier, but perhaps more revealing, argument may be formulated in the following way, which indicates that the $F^{(n)}$ dependence in J just cancels out. We write

$$S=S(\alpha,J,\mathbf{q},\mathbf{P},\epsilon),\quad I=\frac{\partial S}{\partial\alpha},\quad \mathbf{p}=\frac{\partial S}{\partial\mathbf{q}}$$

and calculate $J=J(\alpha,I,\mathbf{q},\mathbf{p},\epsilon)$ directly to obtain

$$\begin{aligned}J=&I-\epsilon\frac{\partial S^{(1)}}{\partial\alpha}(\alpha,I,\mathbf{q},\mathbf{p})-\epsilon^2\frac{\partial S^{(2)}}{\partial\alpha}(\alpha,I,\mathbf{q},\mathbf{p})\\&+\epsilon^2\frac{\partial^2 S^{(1)}}{\partial\alpha\partial I}\cdot\frac{\partial S^{(1)}}{\partial\alpha}+\epsilon^2\frac{\partial^2 S^{(1)}}{\partial\alpha\partial\mathbf{p}}\cdot\frac{\partial S^{(1)}}{\partial\mathbf{q}}+O(\epsilon^3).\end{aligned}\qquad (3.12)$$

We now substitute the expressions for S obtained in (3.7) *et seq.* and for definiteness choose the arbitrary lower limit in $\int^\alpha(\Omega-\overline{\Omega})d\alpha\equiv\hat{\Omega}$ such that $\overline{\hat{\Omega}}=0$ (cf. Ref. 8). In this way (3.12) yields (with $\Omega_p\equiv\partial\Omega/\partial p$ etc.)

$$\begin{aligned}J=&I+\epsilon(\Omega-\overline{\Omega})+\epsilon^2(\overline{\Omega}_I-\Omega_I)(\overline{\Omega}-\Omega)+\epsilon^2(\Omega_p-\overline{\Omega}_p)\hat{\Omega}_q\\&+\epsilon^2(\overline{\Omega}_p-\Omega_p)F_q^{(1)}+\epsilon^2\overline{\Omega}_q\hat{\Omega}_p-\epsilon^2\overline{\Omega}_qF_p^{(1)}-\epsilon^2\overline{\Omega}_q\overline{\hat{\Omega}}_p+\epsilon^2\overline{\Omega}_q\overline{F_p^{(1)}}\\&+\epsilon^2\Omega_I(\overline{\Omega}-\Omega)-\epsilon^2\overline{\Omega_I(\overline{\Omega}-\Omega)}-\epsilon^2\Omega_p\hat{\Omega}_q+\epsilon^2\Omega_pF_q^{(1)}+\epsilon^2\overline{\Omega_p\hat{\Omega}_q}\\&-\epsilon^2\overline{\Omega}_pF_q^{(1)}+O(\epsilon^3)\\=&I+\epsilon(\Omega-\overline{\Omega})+\epsilon^2\{\overline{\Omega},\hat{\Omega}\}-\tfrac{1}{2}\epsilon^2\overline{\{\Omega,\hat{\Omega}\}}+O(\epsilon^3).\end{aligned}\qquad (3.13)$$

Thus J is independent of the choice of $F^{(n)}$ and is, in fact, the invariant of Ref. 8, Eq. (4.34).

Now that we have characterized the transformations which may give rise to the operator $\mathbf{V}$, we shall determine which among these satisfy the Mandel—Turner equation. The condition for this is (2.18), or

$$\lim_{\delta\to 0}\frac{1}{\delta}[\mathbf{P}\mathbf{V}(\epsilon)\mathbf{V}^{-1}(\epsilon+\delta)\mathbf{P}f-\mathbf{P}f](\beta,J,\mathbf{Q},\mathbf{P})=0. \qquad (3.14)$$

Since the transformations $\mathbf{V}(\epsilon)$ and $\mathbf{V}^{-1}(\epsilon+\delta)$ are canonical, so is $\mathbf{V}(\epsilon)\mathbf{V}^{-1}(\epsilon+\delta)$ and since only terms of order δ are required, the last may be considered as an infinitesimal transformation. More explicitly let $\mathbf{V}(\epsilon)\mathbf{V}^{-1}(\epsilon+\delta)$ induce the canonical transformation

$$\beta,J,\mathbf{Q},\mathbf{P}\longrightarrow\beta',J',\mathbf{Q}',\mathbf{P}'$$

with generating function

$$F(\beta,J',\mathbf{Q},\mathbf{P}')=\beta J'+\mathbf{Q},\mathbf{P}'+\delta G(\beta,J',\mathbf{Q},\mathbf{P}')+O(\delta^2).$$

Then

$$\begin{aligned}\beta=\frac{\partial F}{\partial J'}&=\beta+\delta\frac{\partial G}{\partial J'}(\beta,J',\mathbf{Q},\mathbf{P}')+\cdots\\&=\beta+\delta\frac{\partial G}{\partial J}(\beta,J,\mathbf{Q},\mathbf{P})+O(\delta^2)\end{aligned}\qquad (3.15)$$

and, similarly,

$$\mathbf{Q}'=\mathbf{Q}+\delta\frac{\partial G}{\partial\mathbf{P}}(\beta,J,\mathbf{Q},\mathbf{P}), \qquad (3.16)$$

$$J'=J-\delta\frac{\partial G}{\partial\beta}(\beta,J,\mathbf{Q},\mathbf{P}), \qquad (3.17)$$

$$\mathbf{P}'=\mathbf{P}-\delta\frac{\partial G}{\partial\mathbf{Q}}(\beta,J,\mathbf{Q},\mathbf{P}). \qquad (3.18)$$

In this way one obtains

$$\begin{aligned}[\mathbf{V}(\epsilon)\mathbf{V}^{-1}(\epsilon+\delta)f](\beta,J,\mathbf{Q},\mathbf{P})&=[f](\beta',J',\mathbf{Q}',\mathbf{P}')\\&=[f](\beta,J,\mathbf{Q},\mathbf{P})+\delta\{f,G\}(\beta,J,Q,P)+O(\delta^2).\end{aligned}$$

The condition (3.14) now becomes

$$\overline{\{\overline{f},G\}}=0 \qquad (3.19)$$

or

$$\{\overline{f},\overline{G}\}=0,$$

where the bar indicates integration over a period of β. Equation (3.19) can be satisfied for arbitrary f only if $\overline{G}$ is a function of J alone, so the condition under which $\mathbf{V}$ will satisfy the Mandel—Turner equation reduces to

$$\frac{\partial\overline{G}}{\partial\mathbf{Q}}(J,\mathbf{Q},\mathbf{P})=\frac{\partial\overline{G}}{\partial\mathbf{P}}(J,\mathbf{Q},\mathbf{P})=0. \qquad (3.20)$$

The remaining part of our program is to determine G in terms of the S of (3.4) and see which choices of $F^{(n)}$ will satisfy (3.20). We write

$$\mathbf{Q}=\mathbf{q}+\sum_{n=1}\epsilon^n\frac{\partial S^{(n)}}{\partial\mathbf{P}}(\alpha,J,\mathbf{q},\mathbf{P})\equiv\mathbf{q}+\frac{\partial\mathcal{S}}{\partial\mathbf{P}}(\alpha,J,\mathbf{q},\mathbf{P},\epsilon), \qquad (3.21)$$

so

$$\mathbf{Q}'=\mathbf{q}+\frac{\partial\mathcal{S}}{\partial\mathbf{P}'}(\alpha,J',\mathbf{q},\mathbf{P}',\epsilon+\delta), \qquad (3.22)$$

and introduce the notation ∂_i for differentiation with respect to the ith variable of a function, remembering that the argument of G is $(\beta,J,\mathbf{Q},\mathbf{P},\epsilon)$ and that of $\mathcal{S}$ is $(\alpha,J,\mathbf{q},\mathbf{P},\epsilon)$. Utilizing (3.16) we have

$$\begin{aligned}\partial_4 G&=\lim_{\delta\to 0}\frac{1}{\delta}(Q'-Q)\\&=\lim_{\delta\to 0}(\partial_4\partial_2\mathcal{S})\frac{(J'-J)}{\delta}+(\partial_4\partial_4\mathcal{S})\frac{P'-P}{\delta}+\partial_4\partial_5\mathcal{S}\\&=-(\partial_4\partial_2\mathcal{S})(\partial_1 G)-(\partial_4\partial_4\mathcal{S})(\partial_3 G)+\partial_4\partial_5\mathcal{S}\end{aligned}\qquad (3.23)$$

and similar expressions for the other derivatives of G. In terms of $\alpha,J,\mathbf{q},\mathbf{P}$ as independent variables these equations become $(i=1,2,3,4)$

$$\partial_i G\left(\alpha+\frac{\partial \mathcal{S}}{\partial J}, J, \mathbf{q}+\frac{\partial \mathcal{S}}{\partial \mathbf{P}}, \mathbf{P}, \epsilon)\right)$$

$$=\partial_i \partial_5 \mathcal{S}-(\partial_i \partial_2 \mathcal{S})\, \partial_1 G\left(\alpha+\frac{\partial \mathcal{S}}{\partial J}, J, \mathbf{q}+\frac{\partial \mathcal{S}}{\partial \mathbf{P}}, \mathbf{P}, \epsilon\right)$$

$$-(\partial_i \partial_4 \mathcal{S}) \partial_3 G\left(\alpha+\frac{\partial \mathcal{S}}{\partial J}, J, \mathbf{q}+\frac{\partial \mathcal{S}}{\partial \mathbf{P}}, \mathbf{P}, \epsilon\right). \tag{3.24}$$

By expanding $\mathcal{S}$ and G as power series in ϵ, equation (3.24) can be written order by order as

$$\partial_i G^{(0)}(\alpha, J, \mathbf{q}, \mathbf{P})=\partial_i S^{(1)}(\alpha, J, \mathbf{q}, \mathbf{P}),$$

$$\partial_i G^{(1)}=2 \partial_2 S^{(2)}-(\partial_i \partial_2 S^{(1)}) \partial_1 G^{(0)} -(\partial_i \partial_4 S^{(1)}) \partial_3 G^{(0)}-(\partial_i \partial_1 G^{(0)} \partial_2 S^{(1)} -(\partial_i \partial_3 G^{(0)}) \partial_4 S^{(1)},$$

and so on, each time obtaining

$$\partial_i G^{(r-1)}=\partial_i(r S^{(r)}+X^{(r)}), \tag{3.25}$$

where $X^{(r)}$ is determined from $S^{(1)}, \ldots, S^{(r-1)}$.

Returning now to the possibility of satisfying (3.20), we see that we have

$$\partial_3 G^{(r-1)}=\partial_3\{E^{(r)}(\alpha, J, \mathbf{q}, \mathbf{P})+r F^{(r)}(J, \mathbf{q}, \mathbf{P})\},$$

$$\partial_4 G^{(r-1)}=\partial_4\{E^{(r)}(\alpha, J, \mathbf{q}, \mathbf{P})+r F^{(r)}(J, \mathbf{q}, \mathbf{P})\},$$

where $E^{(r)}$ is the sum of $X^{(r)}$ and the (known) oscillatory part of S from (3.9) *et seq.*, and $F^{(r)}$ is the arbitrary function introduced in the equation for $S^{(r)}$. Thus we can arrange for (3.20) to hold simply by choosing

$$r F^{(r)}(J, \mathbf{q}, \mathbf{P})=-\bar{E}^{(r)}(J, \mathbf{q}, \mathbf{P})+D^{(r)}(J),$$

where $D^{(r)}$ is an arbitrary function of J alone. In this way, it is shown that among "nice" canonical transformations, there are some which satisfy the Mandel—Turner equation, and these transformations differ by an additive function of J in the generating function.

In summary then, this section has shown that, starting from the original form of the Hamiltonian (3.1), there exist many canonical transformations to "nice" variables (differing by the choice of $F^{(n)}$), that each of these leads to the same invariant J, and that from the members of this set we can pick a smaller set [differing by the choice of $D(J)$] which satisfy $\mathbf{PV}(\partial \mathbf{V}^{-1}/\partial\epsilon)\mathbf{P}=0$. However, on looking back to (3.3) it is easy to see that adding a function of J to the generating function simply adds a function of J to β but leaves $J, \mathbf{Q}, \mathbf{P}$ unchanged. Since the projector $\mathbf{P}$ integrates over the angle variable, all the transformations differing by $D(J)$ give exactly the same operator $\mathbf{PV}$ and this unique operator satisfies Eq. (2.17) with initial condition that $\mathbf{V}(\epsilon=0)$ is the identity mapping.

IV. DEGREE OF APPROXIMATION

In Secs. 2 and 3 we have made use of series expansions in powers of ϵ, but these have all been in the nature of formal series with no discussion of convergence properties. One might begin such a discussion by asking if these series converge in the usual sense. The answer is that, in general, they do not. In fact, canonical transformations such as we are discussing can be viewed, especially if we regard (3.2) as the starting point, as an example of the general dynamical problem of finding normal forms for Hamiltonians.[13] From the work of Kolmogorov, Moser, and others in this field it is known that the series employed may well have a zero radius of convergence. Of course, it is possible to find many Hamiltonians for which the series *will* converge but this is to be regarded as a lucky accident and in no way the most common case.

If we ask whether the transformations make any sense as asymptotic series then the answers are more satisfactory. As mentioned in Sec. 2, Kruskal[7] has shown how, if we work in a restricted time range $0<t<0(1/\epsilon)$ the transformed equations of motion supply solutions as near as we please to the exact solutions of the original equations. Of course, if we truncate the transformation series at some order ϵ^n say, then the old and new variables are related in a perfectly definite way via a convergent (finite) series for all t. However, for long times of order $1/\epsilon$ the solution of (2.8) will deviate from the true solution by an amount which is, in general, not small.

In a similar way, the invariant J is an asymptotic invariant. If one works with series truncated at order ϵ^n the J so obtained does not remain exactly constant in time but varies according to $\partial J/\partial t=O(\epsilon^{n+1})$. As the series for J does not, in general, converge this is the best one can obtain.

Finally, it is clear that identical considerations apply to the phase-function formulation of the transformation theory, provided that we consider only "smooth" phase functions. Thus if f satisfies a Lipschitz condition

$$\|f(\mathbf{x})-f(\mathbf{x}')\| \leqslant C|\mathbf{x}-\mathbf{x}'|, \quad \mathbf{x}=(\nu, \mathbf{y}), \tag{4.1}$$

where $\|f\|$ is, e.g., $\sup_{\mathbf{x}\in\mathcal{D}} |f(\mathbf{x})|$, the natural norm in $F(\mathcal{D})$, we have in the notation of (2.10)

$$\|f(\nu^*, \mathbf{y}^*)-f(\nu^{(n)}, \mathbf{y}^{(n)}\| \leqslant C\,|(\nu^*, \mathbf{y}^*)-(\nu^{(n)}, \mathbf{y}^{(n)})| =O(\epsilon^{n+1}) \tag{4.2}$$

for times in the appropriate range of order $1/\epsilon$.

V. TIME EVOLUTION AND SUPEROPERATORS

The equations of motion (2.8) determine the evolution of the transformed phase function f_R. The motion clearly has the nature of an oscillation with a superimposed drift taking place on a time scale longer by a factor $1/\epsilon$ than that of the oscillation. From the point of view of the Brussels formalism this is an example of a motion decomposable into "separate subdynamics" for the coherent and oscillatory parts. This can be seen by the explicit consideration of the superoperator Π of (2.16) which effects this decomposition. Using the notation of (2.7) one obtains

$$[\Pi f](\nu, \mathbf{y})=\frac{1}{\tau}\int_0^\tau d\theta[f]\{N(\theta, \mathbf{Z}(\nu, \mathbf{y})), \mathbf{Y}(\theta, \mathbf{Z}(\nu, \mathbf{y}))\}, \tag{5.1}$$

which shows clearly that Πf evolves on the slow time scale only. This evolution of phase functions in the Π subspace is an exact projection of the motion described by (2.6) and it is meaningful under the same conditions as (2.8). That the motion in the Π subspace is in fact a separate subdynamics follows from the commutation of Π with the Liouville operator $\mathbf{L}$:

$$\begin{aligned}\mathbf{V}(\mathbf{L}\Pi-\Pi\mathbf{L})\mathbf{V}^{-1}f&=\mathbf{VLV}^{-1}\mathbf{P}f-\mathbf{PVLV}^{-1}f\\&=-i\left((1+\epsilon\omega)\frac{\partial}{\partial\phi}+\epsilon\mathbf{h}\cdot\frac{\partial}{\partial\mathbf{z}}\right)\mathbf{P}f\\&\quad+i\mathbf{P}\left((1+\epsilon\omega)\frac{\partial}{\partial\phi}+\epsilon\mathbf{h}\cdot\frac{\partial}{\partial\mathbf{z}}\right)f\\&=0,\end{aligned}\tag{5.2}$$

where we have used (2.16) and (2.8). If we denote by $\Sigma(t)$ the time-evolution operator in the Π subspace, it follows from (5.1) and (5.2) that

$$\Sigma(t)=\exp(-i\mathbf{L}t)\Pi=\Pi\exp(-i\mathbf{L}t).\tag{5.3}$$

Following the arguments of Ref. 5 the usual operators of the Brussels school may be obtained by splitting Π into four components as follows:

$$\begin{aligned}&\mathbf{A}=\mathbf{P}\Pi\mathbf{P},\quad \mathbf{CA}=\mathbf{Q}\Pi\mathbf{P},\\&\mathbf{AD}=\mathbf{P}\Pi Q,\quad \mathbf{CAD}=\mathbf{Q}\Pi\mathbf{P},\end{aligned}\tag{5.4}$$

where the operators $\mathbf{C}$ and $\mathbf{D}$ are well defined when $\mathbf{A}^{-1}$ exists. The operator $\mathbf{C}$, of great importance in the theory of Prigogine, can be defined in the standard development of that theory by the equation

$$-\mathbf{QLQC}+\mathbf{CPLQC}+\mathbf{CPLP}=\mathbf{QLP}\tag{5.5}$$

(see Ref. 2 where, however, $\mathbf{PLP}$ has been taken to be zero).

But starting from the $\mathbf{C}$ defined by (5.4), we have

$$\begin{aligned}\mathbf{CPLQCA}+\mathbf{CPLPA}&=\mathbf{CPLQ}\Pi\mathbf{P}+\mathbf{CPLP}\Pi\mathbf{P}\\&=\mathbf{CPL}\Pi\mathbf{P}=\mathbf{CP}\Pi\mathbf{LP}=\mathbf{C}(\mathbf{A}+\mathbf{AD})\mathbf{LP}\\&=\mathbf{Q}\Pi\mathbf{PLP}+\mathbf{Q}\Pi\mathbf{QLP}=\mathbf{Q}\Pi\mathbf{LP}=\mathbf{QL}\Pi\mathbf{P}\\&=\mathbf{QLQ}\Pi\mathbf{P}+\mathbf{QLP}\Pi\mathbf{P}\\&=\mathbf{QLQCA}+\mathbf{QLPA}.\end{aligned}$$

A comparison with (5.5) demonstrates that the $\mathbf{C}$ obtained from (5.4) is identical to that of the usual approach. The operator $\mathbf{D}$ can be treated in a similar way, and we already know from its definition (2.15) that Π and hence $\mathbf{A}$ are the same operators as occur in the usual approach.

At this point, it is usual to introduce the operators χ and $\chi^\dagger$ acting on the $\mathbf{P}$-subspace and defined by the two relations

$$\chi\chi^\dagger=\mathbf{A},\quad \frac{\partial\chi^\dagger}{\partial\epsilon}=\chi^\dagger\left(\frac{\partial\mathbf{A}}{\partial\epsilon}+\mathbf{D}\frac{\partial(\mathbf{CA})}{\partial\epsilon}\right).\tag{5.6}$$

It is possible to obtain these superoperators simply in terms of $\mathbf{V}$ and in fact we now prove that the choice

$$\chi=\mathbf{PV}^{-1}\mathbf{P},\quad \chi^\dagger=\mathbf{PVP}\tag{5.7}$$

satisfies the defining equations (5.6). The first equation is trivial

$$\chi\chi^\dagger=\mathbf{PV}^{-1}\mathbf{PVP}=\mathbf{P}\Pi\mathbf{P}=\mathbf{A}.$$

For the second we note that

$$\begin{aligned}\mathbf{PV}&=\mathbf{PVP}+\mathbf{PVQ}=\chi^\dagger+\chi^{-1}\chi\mathbf{PVQ}\\&=\chi^\dagger+\chi^{-1}\mathbf{PV}^{-1}\mathbf{PVQ}=\chi^\dagger+\chi^{-1}\mathbf{P}\Pi\mathbf{Q}\\&=\chi^\dagger+\chi^{-1}\mathbf{AD}=\chi^\dagger(1+\mathbf{D}).\end{aligned}\tag{5.8}$$

Similarly,

$$\mathbf{V}^{-1}\mathbf{P}=(1+\mathbf{C})\chi.\tag{5.9}$$

But the Mandel–Turner equation (2.17) is

$$\frac{\partial(\mathbf{PV})}{\partial\epsilon}=\mathbf{PV}\frac{\partial\Pi}{\partial\epsilon}$$

and post-multiplication of this by $\mathbf{P}$ gives at once

$$\begin{aligned}\frac{\partial\chi^\dagger}{\partial\epsilon}&=\chi^\dagger(1+\mathbf{D})\frac{\partial}{\partial\epsilon}(\Pi\mathbf{P})=\chi^\dagger(1+\mathbf{D})\frac{\partial}{\partial\epsilon}(\mathbf{A}+\mathbf{CA})\\&=\chi^\dagger\frac{\partial\mathbf{A}}{\partial\epsilon}+\chi^\dagger\mathbf{D}\frac{\partial}{\partial\epsilon}(\mathbf{CA})\end{aligned}$$

which is the second equation of (5.6).

Next, we decompose the operator $\Sigma(t)$ in a way analogous to (5.4) as[1]

$$\begin{aligned}&\mathbf{P}\Sigma(t)\mathbf{P}=\exp(-i\Omega\psi t)\mathbf{A},\quad \mathbf{Q}\Sigma(t)\mathbf{P}=\mathbf{C}\exp(-i\Omega\psi t)\mathbf{A},\\&\mathbf{P}\Sigma(t)\mathbf{Q}=\exp(-i\Omega\psi t)\mathbf{AD},\quad \mathbf{Q}\Sigma(t)\mathbf{Q}=\mathbf{C}\exp(-i\Omega\psi t)\mathbf{AD},\end{aligned}$$

in which for simplicity of comparison we have restricted ourselves to the case $\mathbf{PLP}=0$. It follows at once that

$$\Omega\psi\mathbf{A}=\mathbf{PL}\Pi\mathbf{P}=\mathbf{PLQ}\Pi\mathbf{P}.\tag{5.10}$$

In the usual form of the theory, this operator can be defined from the relation[2]

$$\Omega\psi=\mathbf{PLQC},$$

which is an immediate consequence of (5.10).

Finally, for use in the next section, we make two more remarks. The basic collision operator ψ of the theory is usually defined by the perturbation series[1]

$$\psi=\lim_{z\to+i0}\left\langle 0\left|\epsilon\mathbf{L}_1\sum_{r=1}\left(\frac{1}{z-\mathbf{L}_0}\mathbf{Q}\epsilon\mathbf{L}_1\right)^r\right|0\right\rangle=\sum_{s=2}\epsilon^s\psi^{(s)},\tag{5.11}$$

where the "matrix-element" notation means, for example,

$$\langle n|\mathbf{L}|m\rangle=\frac{1}{\tau}\int_0^\tau\exp(-2\pi in\alpha/\tau)\mathbf{L}\exp(\pi im\alpha/\tau)\,d\alpha.$$

Likewise one defines the "creation" and "destruction" operators $\mathcal{C}$ and $\mathcal{D}$ as follows:

$$\mathcal{C}_n=\lim_{z\to+i0}\left\langle n\left|\sum_{r=1}\left(\frac{1}{z-\mathbf{L}_0}\mathbf{Q}\epsilon\mathbf{L}_1\right)^r\right|0\right\rangle=\sum_{s=1}\epsilon^s\mathcal{C}_n^{(s)},\tag{5.12}$$

$$\mathcal{D}_n=\lim_{z\to+i0}\left\langle 0\left|\sum_{r=1}\left(\epsilon\mathbf{L}_1\mathbf{Q}\frac{1}{z-\mathbf{L}_0}\right)^r\right|n\right\rangle=\sum_{s=1}\epsilon^s\mathcal{D}_n^{(s)}.\tag{5.13}$$

In the further perturbational development of the theory, the operators $\mathbf{C}$ and $\mathbf{D}$ that we have used earlier, are defined,[1b,11] not by the difficult nonlinear equation (5.5), but by a series involving $\mathcal{C}$, $\mathcal{D}$, ψ, and the derivatives, evaluated at $z=+i0$, of the z-dependent expressions in the defining equations (5.11)–(5.13). These are the definitions of **C** and **D** employed in Turner's solution[4] of the Mandel–Turner equation in the form (5.6), which is

$$\chi^{-1} = \mathbf{P} + \sum_{p,q=1} C_{pq}\mathbf{D}^{(p)}\mathbf{C}^{(q)} + \sum_{p,q,r,s} C_{pqrs}\mathbf{D}^{(p)}\mathbf{C}^{(q)}\mathbf{D}^{(r)}\mathbf{C}^{(s)} \tag{5.14}$$

with

$$C_{pq} = q/(p+q), \quad C_{pqrs} = -sp/(p+q)(p+q+r+s), \text{ etc.}$$

The derivative operators mentioned above appear for the first time at fourth order in ϵ in the development (5.14), and in the particular calculation in Sec. 6, they in fact give a zero contribution at this order, so that only the simple operators $\mathcal{C}$ and $\mathcal{D}$ appear explicitly. At higher orders, more complicated terms may well contribute.

VI. CALCULATION OF THE INVARIANT: A COMPARISON OF METHODS

From the foregoing results we know that Hamiltonian systems of the type (3.1) possess an (asymptotic) invariant J. It is often useful to calculate an explicit expression for this in terms of the original variables α, I, q, p and there exists a variety of methods for doing this. One may follow Kruskal's method[7] for determining "nice" coordinates in terms of which J is an action integral or, equivalently, solve the generating function equation for "nice" canonical variables as is done in Sec. 3. These two methods are systematic but involve tedious calculation since the expression for J at any order involves information about the other "nice" variables at lower orders. A more direct method (see below) was proposed by McNamara and Whiteman[8] who showed that their method was equivalent to that of Kruskal. Finally, a method for calculating invariants has been developed by the Brussels school[14,15] using the operators of their formalism of statistical mechanics. Since we know, from the above, the correspondance between the methods of Kruskal and Prigogine we can now link all these methods. In particular, we show below that the methods of Ref. 8 and those of the Brussels group are equivalent, and that the latter removes some difficulties which rendered the former unsystematic.

We begin by considering the method of McNamara and Whiteman[8] as applied to Hamiltonians of the form (3.1). This proceeds by solving, in a perturbation scheme, the equation for the invariant

$$J = \sum \epsilon^n J^{(n)}:$$

$$[J,H] \equiv \sum_i \left(\frac{\partial J}{\partial p_i}\frac{\partial H}{\partial q_i} - \frac{\partial J}{\partial q_i}\frac{\partial H}{\partial p_i}\right) + \frac{\partial J}{\partial I}\frac{\partial H}{\partial \alpha} - \frac{\partial J}{\partial \alpha}\frac{\partial H}{\partial I} = 0. \tag{6.1}$$

One introduces two operations defined on functions f which are periodic in α (period τ) by

$$\bar{f} = \frac{1}{\tau}\int_0^\tau f\,d\alpha, \quad \hat{f} = \int^\alpha (f - \bar{f})\,d\alpha, \tag{6.2}$$

in terms of which the equations to be solved become

$$J^{(n)} = [\widehat{J^{(n-1)}}, \Omega] + G^{(n)}, \tag{6.3}$$

$$\overline{[J^{(n-1)}, \Omega]} = 0. \tag{6.4}$$

The authors develop a calculus for manipulating the "bar" and "hat" operations with the bracket [,] of (6.1) and use of this enables them to produce a solution for J which, in the special case $\overline{\Omega} = 0$, can be made explicit to fourth order in ϵ. At higher orders the method becomes prohibitively complicated.

The methods of the Brussels school have already been outlined in the preceding sections and here we require in particular the Liouville operator **L**, the projectors **P** and **Q** of Sec. 2, and the operators ψ, $\mathcal{C}$, $\mathcal{D}$, and χ of (5.11) through (5.14). We also require the result, shown by various authors,[1,14,15] that starting from an invariant $J^{(0)}$ of H_0 which satisfies

$$(\mathbf{PLP} + \psi)J^{(0)} = 0, \tag{6.5}$$

one can construct an invariant of H by

$$J = J^{(0)} + \sum_{n\neq 0} \exp(2\pi i n\alpha/\tau)\mathcal{C}_n J^{(0)} \tag{6.6}$$

and conversely an invariant J satisfies (6.5) and (6.6).

We turn now to the connection between these two formalisms and construct a dictionary for translating from one to the other. First we notice that the "bar" operation of (6.2) is exactly the projector **P** of Sec. 2, and that the bracket operation with Ω found in (6.3), (6.4) is essentially the action of $\mathbf{L}_1$, the perturbation in the Liouville operator. Next, if we decompose a phase function f

$$f = \bar{f} + \sum_{n\neq 0} \exp(2\pi i n\alpha/\tau) f_n(I, \mathbf{q}, \mathbf{p}) = \mathbf{P}f + \mathbf{Q}f,$$

we obtain

$$\frac{1}{z - \mathbf{L}_0}\mathbf{Q}f = \sum_{n\neq 0} \frac{1}{z - 2\pi n/\tau}\exp(2\pi i n\alpha/\tau) f_n.$$

There is no difficulty in taking $z \to +i0$ in this to yield

$$\frac{1}{-\mathbf{L}_0}\mathbf{Q}f = \sum_{n\neq 0} \frac{-\tau}{2\pi n}\exp(2\pi i n\alpha/\tau) f_n = -i\int^\alpha d\alpha \sum_{n\neq 0} \exp(2\pi i n\alpha/\tau) f_n$$
$$= -i\int^\alpha d\alpha (f - \bar{f}). \tag{6.7}$$

We also have $\mathbf{P}(1/z - \mathbf{L}_0)\mathbf{Q}f = 0$, so that the constant of integration in (6.7) must be chosen so that $\mathbf{P}\int^\alpha d\alpha (f - \bar{f}) = 0$. Comparison with the definitions of Ref. 8 now give

$$\mathbf{P}f = \bar{f}, \quad \mathbf{L}_1 f = i[f, \Omega], \quad (1/-\mathbf{L}_0)\mathbf{Q}f = -i\hat{f} \tag{6.8}$$

which is the required dictionary. As an example of its use one may verify from (5.11) that

$$\psi^{(2)} f = \mathbf{PL}_1 \frac{1}{-\mathbf{L}_0}\mathbf{QL}_1\mathbf{P}f = \overline{i[[\bar{f}, \Omega], \Omega]}$$
$$= -\frac{i}{2}[\overline{[\hat{\Omega}, \Omega]}, \bar{f}] \tag{6.9}$$

which shows that, for the systems considered here, ψ is not identically zero, contrary to the cases looked at in Refs. 3, 5, and 6.

We now proceed to the explicit construction of J and first determine the particular choice of $J^{(0)}$ to be used in (6.5) and (6.6). Let $\mathcal{J}$ be the function in $F(\mathcal{D})$ (suitably truncated if need be) defined by

$$[\mathcal{J}](\beta,J,\mathbf{Q},\mathbf{P})=J. \tag{6.10}$$

Then $\mathcal{J}=\mathbf{P}\mathcal{J}$ and $J=[\mathbf{V}^{-1}\mathcal{J}](\alpha,I,\mathbf{q},\mathbf{p})$ so that J as a function of $\alpha,I,\mathbf{q},\mathbf{p}$ is just $\mathbf{V}^{-1}\mathbf{P}\mathcal{J}$. But from (5.9) we see that

$$J(\alpha,I,\mathbf{q},\mathbf{p})=\Big[\chi\mathcal{J}+\sum_{n\neq 0}\exp(2\pi in\alpha/\tau)C_n\chi\mathcal{J}\Big]\,(\alpha,I,\mathbf{q},\mathbf{p})$$
$$=\chi I+\sum_{n\neq 0}\exp(2\pi in\alpha/\tau)\mathbf{C}_n\chi I, \tag{6.11}$$

where the operators χ and $\mathbf{C}$ are given their realization in the variables $\alpha,I,\mathbf{q},\mathbf{p}$. A comparison of (6.3)–(6.6) and 6.11 entitles us to make the following two assertions.

(i) The function χI is essentially the function G of (6.3) and it satisfies Eq. (6.5)

$$(\mathbf{PLP}+\psi)\chi I=0 \tag{6.12}$$

which is exactly condition (6.4) of McNamara and Whiteman.

(ii) In view of 6.12 the action of $\mathbf{C}_n$ in 6.11 reduces to $\mathcal{C}_n$ and the angle-dependent part of the invariant can be constructed with the operator $\mathcal{C}$ to obtain the complete invariant in the form (6.6), which is exactly condition (6.3).

More explicitly, we now calculate the first few orders and begin by showing

$$G^{(n)}=-\chi^{(n+1)}I. \tag{6.13}$$

From the expression (5.14) for χ^{-1} we have, using our "dictionary"

$$\chi^{(0)}I=I,\quad \chi^{(1)}I=0, \tag{6.14}$$

$$\chi^{(2)}I=-\tfrac{1}{2}\mathcal{D}^{(1)}\mathcal{C}^{(1)}I=-\tfrac{1}{2}\mathbf{PL}_1\frac{1}{-\mathbf{L}_0}\mathbf{Q}\frac{1}{-\mathbf{L}_0}\mathbf{L}_1\mathbf{P}I$$
$$=-\tfrac{1}{2}\mathbf{PL}_1\frac{1}{-\mathbf{L}_0}\mathbf{Q}(\Omega-\overline{\Omega})=\frac{i}{2}\mathbf{PL}_1\hat{\Omega}$$
$$=-\tfrac{1}{2}\mathbf{P}[\hat{\Omega},\Omega]=-\tfrac{1}{2}\overline{[\hat{\Omega},\Omega]}. \tag{6.15}$$

In a similar way, after a certain amount of manipulation

$$\chi^{(3)}I=-\tfrac{1}{3}\mathcal{D}^{(2)}\mathcal{C}^{(1)}I-\tfrac{2}{3}\mathcal{D}^{(1)}\mathcal{C}^{(2)}I$$
$$=-\tfrac{1}{3}\overline{[\hat{\Omega},[\hat{\Omega},\Omega]]}-\tfrac{2}{3}\overline{[\hat{\Omega},[\hat{\Omega},\overline{\Omega}]]}. \tag{6.16}$$

Results (6.14)–(6.16) are identical to those of Ref. 8.

When $\overline{\Omega}=0$ we have the result

$$\mathcal{C}_n^{(1)}I=n\Big\langle\frac{1}{-\mathbf{L}_0}\mathbf{L}_1\Big\rangle 0\ I=\Omega_n,\quad \mathcal{C}_n^{(s)}I=0,\quad s>1, \tag{6.17}$$

which enables us to calculate the fourth order term (in the case $\overline{\Omega}=0$) as

$$\chi^{(4)}I=-\tfrac{1}{4}\mathcal{D}^{(3)}\mathcal{C}^{(1)}I+\tfrac{3}{8}\mathcal{D}^{(1)}\mathcal{C}^{(1)}\mathcal{D}^{(1)}\mathcal{C}^{(1)}I.$$

After use of (6.8) and rather extensive rearrangement this is seen to be identical to the $-G^{(3)}$ of Ref. 8.

Knowing the angle-independent parts $G^{(n)}$, one may construct the complete invariant using (6.3), that is

$$J=\sum_{n=0}\epsilon^n J^{(n)}=\sum_{n=0}[\widehat{J^{(n-1)}},\Omega]+G^{(n)}$$
$$=\sum_{n=0}\epsilon^n\Big(\frac{1}{-\mathbf{L}_0}\mathbf{QL}_1J^{(n-1)}+G^{(n)}\Big)$$
$$=\sum_{n=0}\epsilon^n(G^{(n)}+\mathcal{C}^{(1)}G^{(n-1)}+\cdots+\mathcal{C}^{(n)}I)$$
$$=(1+\mathcal{C})\sum_{n=0}\epsilon^nG^{(n)}$$

which shows that (6.3) is exactly (6.6).

Thus we have shown that use of 6.6 with $J^{(0)}=\chi I$ allows a systematic order-by-order calculation of the invariant. This calculation is equivalent to the method of McNamara and Whiteman but has the great advantage that, since χ is known to all orders in ϵ, one may calculate by a completely determined procedure to any desired order.

VII. CONCLUSIONS

We have seen that for the simplest systems to which the averaging method is applicable there exists one and only one of Kruskal's transformation which satisfies the Mandel–Turner equation. This furnishes us with a simple mechanical example exhibiting some features of the general scheme of the Brussels group, in particular the notion of subdynamics. Of course, owing to the extreme simplicity of the case there is here no question of thermodynamic behavior but merely a clear separation of mechanical motions. One is tempted to generalize by complicating the model, for example, by starting from (3.2) without such strong assumptions on the frequencies. It seems to the authors that, on the formal level, most of the content of this paper could be extended in a straightforward way but that great difficulties will arise in any attempt to justify the formal series as asymptotic series (cf. the well-known difficulties of the averaging method for many degrees of freedom[12]). Finally, we comment on the fact that in the usual development of the Brussels scheme[1] the transformations considered are seen to include but be more general than canonical transformations. This feature does not show up in our very simple case. We could of course examine systems which evolve according to a Liouville equation not derivable from a Hamiltonian and then there would be no question of canonical transformations. However, it is interesting that even for Hamiltonian systems, if they have more than one rapidly oscillating variable, the "nice" transformations cannot always be made canonical and, in fact, the conditions for this depend on delicate considerations of degeneracy which have not yet been completely resolved.[16,17,18] In these cases it may still prove possible to establish a connection with the superoperator transformation theory by a method not dependent on canonical transformations at the intermediate stage.

ACKNOWLEDGMENTS

The authors wish to thank Professor I. Prigogine and Drs. J. Wm. Turner, P. Mandel, and A.P. Grecos for their assistance during the course of this work.

*Present address: The University of British Columbia, Vancouver, Canada.

[1]I. Prigogine, *et al.* (a) Proc. Natl. Acad. Sci. USA **65**, 789 (1970), (b) Proc. Natl. Acad. Sci. USA **66**, 709 (1970).
[2]I. Prigogine, *et al.*, Physica (Utr.) **45**, 418 (1969).
[3]P. Mandel, Physica (Utr.) **50**, 77 (1970).
[4]J. W. Turner, Physica (Utr.) **51**, 351 (1971).
[5]J. Rae and R. Davidson, J. Stat. Phys. **3**, 135 (1971).
[6]I. Prigogine, Cl. George, and J. Rae, Physica **56**, 25 (1971).
[7]M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[8]B. McNamara, and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).
[9]I. Prigogine, *Topics in Non-Linear Physics, Proceedings of International School of Non-Linear Math. and Phys. Munich 1966* (Springer, New York, 1968).
[10]P. Mandel, Physica (Utr.) **48**, 397 (1970).
[11]Cl. George, Bull. Acad. Roy. Belg. Cl. Sci. **53**, 623 (1967).
[12]V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).
[13]J. Moser, Mem. Amer. Math. Soc. **81**, 1 (1968).
[14]R. Balescu *et al.*, Bull. Acad. Roy. Belg. Cl. Sci. **55**, 1055 (1969).
[15]A. P. Grecos, Physica (Utr.) **57**, 50 (1971).
[16]T. P. Coffey, J. Math. Phys. **10**, 426 (1969).
[17]M. Kummer, J. Math. Phys. **12**, 4 (1971).
[18]T. P. Coffey and G. W. Ford, J. Math. Phys. **10**, 998 (1969).

Simplex transformations of the spin-1/2 Ising model. I

D. W. Wood and H. P. Griffiths

Mathematics Department, University of Nottingham, Nottingham, England

(Received 18 January 1972)

The problems of developing power series expansions for the thermodynamic functions of the spin-1/2 Ising model are discussed. The basis of the discussion is a generalized Ising-type model which incorporates collective potentials between special groups of k particles on the lattice sites. The addition of these k-body potential functions throws a new light onto the problems of developing series expansions for the conventional spin-1/2 Ising model. A system of transformation equations connecting the high and low temperature series developments of the model is obtained which contains the well-known high–low transformation of the Ising model. It is likely that this generalized transformation will provide a useful additional technique in the theory of the conventional Ising model, as well as providing a basis for the discussion of other lattice models which exhibit critical phenomena.

I. INTRODUCTION

This paper is about the problems which are encountered in forming exact power series expansions for the thermodynamic functions of an Ising model system. In recent years these expansions have been studied in great detail. Many of the principle references and a general lead into the subject can be found in reviews by Domb[1] and by Fisher.[2] Our discussion throughout will be restricted to the special case of the spin-$\frac{1}{2}$ Ising model, however, throughout the literature formal generalizations of much of the theory can be found, and applied to the contents of this paper.

There are two principle forms of a series expansion development of the Ising model thermodynamic functions; these are, (a) high temperature expansions valid in regions $T > T_c$, where T_c is the critical temperature, and (b) low temperature expansions in regions $T < T_c$.

An immense amount of data is available in the literature for both types (a) and (b). This data is the principle source of our present knowledge concerning the critical point behavior of a variety of transition phenomena, which can be represented in terms of some type of Ising model defined on a regular lattice L_q, where q denotes the degree of each vertex in the lattice. The evaluation of successive terms in (a) and (b) can be related to the enumeration and counting of sets of subgraphs in L_q; in this much ingenuity has been devised to reduce to an acceptable minimum the amount of direct enumeration necessary in determining a given set of subgraphs which contribute to the terms in (a) and (b). The most successful of these techniques has been centered on the series in (b).

The first such special technique was devised by Domb,[3] and is a transformation equation which relates the two forms of expansion (a) and (b). This transformation is known as the high temperature—low temperature ($h-l$) transformation. This transformation assumes the absence of any singularities in the canonical partition function in unlikely regions of the phase diagram corresponding to the model. The $h-l$ transformation exploits some simple symmetry properties of the partition function, which results in a reduction of the direct enumeration needed to develop the series (b). This technique has been successfully developed by a number of authors.[4–9]

A further reduction in effort was achieved by Sykes, Essam, and Gaunt,[7] who devised a technique known as the method of partial generating functions, which was originally applied to lattices which are divisible into two interpenetrating sublattices. Basically this method exploits a symmetry relation between the (b) series developed for the Ising model of a ferromagnet, and the corresponding series for the antiferromagnet. Using this method these workers extended the series (b) for a variety of lattices.

An alternative transformation relating the two forms (a) and (b) was proposed by Nagle,[8] who gave a more generalized treatment which included a variety of network models. This transformation known as the 'closed-weak' transformation does not so far seem to have been used in developing series expansions for the Ising model.

The most recent work in the development of the (b) series is contained in a major series of calculations by Sykes and coworkers,[10,11] where the original method of partial generating functions has been developed and extended to include all the common lattice structures.

In this paper we introduce what we believe to be a new transformation between series of type (a) and (b) which contains much symmetry, and which we call the simplex transformation. This transformation is defined on a generalized set of Ising-like models which have an interesting mathematical and physical relation with the conventional spin-$\frac{1}{2}$ Ising model. We envisage an extended Hamiltonian H_{L_q} for lattice L_q of the form

$$H_{L_q} = H_1 + H_2 + \cdots + H_n, \tag{1}$$

when H_k in (1) is defined as a contribution to H_{L_q} arising from groups of k sites in L_q. Thus the conventional Ising model Hamiltonian H_I is

$$H_I = H_1 + H_2, \tag{2}$$

where H_1 is a single particle term (the magnetic field interaction with L_q for the magnetic Ising Model), and H_2 is the contribution arising from the pair interactions. The clusters of k sites over which H_k is defined are the k-point simplex subgraphs s_k of L_q. Thus s_n is the largest simplex subgraph in L_q, that is the $(n+1)$-point simplex cannot be embedded in L_q. Thus for the two-dimensional triangular lattice ($q=6$) $n=3$, and for the three-dimensional face centered cubic lattice ($q=12$) $n=4$, where in both cases L_q is formed as a graph where only nearest neighbor points are connected.

Copyright © 1973 American Institute of Physics

The $h-l$ transformation of Domb arises as a natural consequence of the groupings of graphs in the terms of (a) and (b). In the (b) series groups of graphs are formed which have the same number of points and lines, that is, the same number of s_1 and s_2 simplex subgraphs. This is a reproducible pattern in relation to the inclusion of other simplex clusters in Eq. (1), and in this way a new type of transformation is obtained which we call the simplex transformation, which throws an interesting light on the expansions (b).

The plan of the paper is as follows. In Sec. II the prototype expansions (a) and (b) for the spin-$\frac{1}{2}$ Ising model, and the $h-l$ transformation are briefly reviewed. In Secs. III and IV we describe the generalization of (a) and (b) which result from interaction Hamiltonians in the form of Eq. (1). In Sec. V we derive the new simplex transformation, which is illustrated with the specific example of the triangular lattice in Sec. VI. Finally, in Sec. VII we illustrate a natural physical model based on Eq. (1), which is a model of a lattice fluid in which triplet potential functions are included.

II. THE HIGH TEMPERATURE-LOW TEMPERATURE TRANSFORMATION

The form of the series expansions of the spin-$\frac{1}{2}$ Ising model partition function Z_N^I for a lattice L_q of N sites in which only the nearest neighbor pairs of sites in L_q are connected is well known. Following the notation of Domb[1] the expansions are summarized below.

A. The high temperature hyperbolic tangent expansion

We can write Z_N^I in the form

$$Z_N^I = \mu^{-N/2} z^{-Nq/4} \Lambda_N(\mu, z), \tag{3}$$

where $\mu = \exp(-2mH/kT)$ and $z = \exp(-2J/kT)$. The function Λ_N can be expanded in the form

$$\Lambda_N(\mu, z) = \left(\frac{1+z}{2}\right)^{Nq/2} (1+\mu)^N \left(1 + \sum_{r=1}^{Nq/2} \sum_j (r:j) v^r \tau^{n_0}\right), \tag{4}$$

where $v = (1-z)/(1+z)$ and $\tau = (1-\mu)/(1+\mu)$. In (4) the symbol $r:j$ refers to a graph of r lines, and the index j serves to distinguish topological types. $(r:j)$ denotes the number of weak embeddings[12] of $r:j$ in L_q, and n_0 is the number of vertices of the graph which are of odd degree, that is vertices at which an odd number of lines meet. Following Domb[1] (see also Nagle[9]) the expansion (4) can be rearranged as a power series in $(1-z)$, and we readily obtain the form

$$\Lambda(\mu, z) = 1 + \mu + \sum_{r=1}^{\infty} \frac{\phi_{(r)}(\mu)}{(1+\mu)^{2r-1}} (1-z^2)^r, \tag{5}$$

where $\Lambda(\mu, z) = \Lambda_{N=1}(\mu, z)$, and $\phi_{(r)}(\mu)$ are symmetric polynomials of degree $2r$. The symmetry of the polynomials is apparent from Eq. (4) since

$$\mu^{-1/2}\Lambda(\mu, z) = \mu^{1/2}\Lambda(\mu^{-1}, z), \tag{6}$$

which follows directly from the observation that n_0 is necessarily even.

B. Low temperature expansions

By successively introducing perturbations on the ground state of the spin-$\frac{1}{2}$ Ising model in taking groups of s lattice sites for which the site variables $\sigma = -1$, we can readily derive the following expansion

$$\Lambda_N(\mu, z) = 1 + \sum_{s=1}^{N} \sum_k [k:s:r] u^{sq/2-r} \mu^s, \quad (u = z^2), \tag{7}$$

where $k:s:r$ denotes a subgraph of L_q containing s points, r lines $[0 \le r \le \frac{1}{2}s(s+1)]$ of topological type k. $[k:r:s]$ is the number of strong embeddings[12] of the graph in L_q. The terms in this expansion are commonly grouped so as to form the so-called high field expansion in which the high field polynomials are defined by

$$\Lambda(\mu, z) = 1 + \sum_{s=1}^{\infty} f_s(z)\mu^s. \tag{8}$$

The $h-l$ transformation is based upon the Eqs. (5) and (8) from which we readily derive the relations

$$f_s(1) = 0, \quad s > 1 \tag{9}$$

and

$$f_s^{(p)}(1) = \text{coefficient of } \mu^s \text{ in } [(-1)^p p!\,\phi_{(p)}(\mu)/(1+\mu)^{2p-1}], \tag{10}$$

where $f_s^{(p)}(1)$ is the pth derivative of $f_s(u)$ evaluated at $u=1$. The connecting equations (9) and (10) have been exploited by numerous workers[4-9] in evaluating $f_s(u)$.

III. GENERALIZED LOW TEMPERATURE SIMPLEX EXPANSIONS

Returning now to the expansion (7), we can regard Λ as a generating function, the expansion of which generates groups of strong embeddings of subgraphs in L_q; we see that in the coefficient of $\mu^s u^{(1/2)qs-r}$ we have

$$\text{coefficient of } \mu^s u^{(1/2)qs-r} = \sum_k [k:r:s], \tag{11}$$

where the sum on the right of (11) is the total number of strong embeddings (evaluated at $N=1$) for subgraphs in L_q with s points and r lines. The sums in (11) are precisely what we need in the derivation of the high field polynomials $f_s(z)$. Thus for $s=10$ on the triangular lattice r can take the values $0, 1, 2, \ldots, 19$ giving 20 such sums. Of course in the evaluation of $f_s(z)$ we need only determine the sums in (11) which contribute to $f_s(z)$, and it is this which is exploited using the connecting equations (9) and (10).

Given that the Ising model problem leads naturally to a partitioning of the strong embeddings of all the subgraphs in L_q, whereby the graphs in each element of the partition contain the same number of points and lines, one can seek alternative partitions of the same set of strong embeddings. For any graph $k:s:r$ the point and the line are both subgraphs of $k:s:r$, consequently the grouping represented by Eq. (11) groups these graphs in L_q with the same number of point and line subgraphs. It is natural to ask if we can extend this type of grouping whereby additional subgraphs of $k:s:r$ are identified thus yielding a different partition of the total set of graphs in L_q in which a larger number of groups of graphs in L_q are partitioned. We require to know what type of subgraphs of $k:s:r$ can be incorporated into expansions of the type in Eq. (7).

Suppose we select a subgraph g, then we seek a meaningful expansion of the form

$$\Lambda(\mu,u,w)=\sum_{s=0}^{\infty}[k:s:r:n_g]u^{(1/2)sq-r}w^{\gamma(n_g)}\mu^s, \tag{12}$$

where the variable w is some additional variable in the generating function, related to the graphs g in L_q, and $k:s:r:n_g$ is the strong embedding lattice constant ($N=1$) of a graph of s points, r lines, and n_g subgraphs g. Now we shall require to impose a constraint on g in that

$$[k:s:r:n_g]=[k:s:r], \tag{13}$$

which is a statement to the effect that the exponent $\gamma(n_g)$ of w in Eq. (12) must be independent of the various space types[13] of the strong embeddings in L_q of the graph $k:s:r:n_g$. Equation (13) represents a very strong condition on the possible graphs for g, and in fact restricts g to be a simplex graph. We will denote the j-point simplex graph by s_j. Clearly, the conventional Ising model expansions in Eq. (7) are graph generating functions in which information is carried relating to the s_1 and s_2 subgraphs of the graphs contributing to the expansion.

Consider now a lattice L_q such that $s_1, s_2, \ldots, s_n \in L_q$, and let the variables associated with these simplex groups of sites be $\mu, u_1, u_2, \ldots,$ and u_{n-1}, respectively; then the generalized form of Eq. (12) for this lattice will be

$$\Lambda(\mu,u_1,u_2,\ldots,u_{n-1})=\sum_{n_{s_1}}^{\infty}[k:n_{s_1}:n_{s_2}:\cdots:n_{s_n}] \times u_1^{(1/2)n_{s_1}q-n_{s_2}}u_2^{\gamma_2(n_{s_1},n_{s_2},n_{s_3})}\cdots u_{n-1}^{\gamma_{n-1}(n_{s_1},n_{s_2},\ldots,n_{s_n})}\mu^{n_{s_1}}, \tag{14}$$

where $k:n_{s_1}:n_{s_2}:\cdots:n_{s_n}$ is a subgraph of L_q containing n_{s_1}-points, n_{s_2}-lines, $\cdots$, and n_{s_n} n-point simplex subgraphs. The nature of the original Ising model, and the lattice L_q determines the maximum n for which $s_n \in L_q$. Thus the triangular lattice with nearest neighbor interactions contains s_1, s_2, and s_3 simplexes; however, if next nearest neighbor interactions are included the s_4 simplex can be added to this list, and the face centered cubic lattice with only nearest neighbor interactions present contains simplexes up to s_4.

We next have to obtain the exponent functions $\gamma_{j-1}(n_{s_1}, n_{s_2}, \ldots, n_{s_j})$ in Eq. (14). These can be readily obtained in the usual way (see Domb[1]); thus, if t_{ji} ($i \leq j$) is the number of simplexes s_j sharing a common simplex s_i in L_q, we find

$$\gamma_{j-1}=\tfrac{1}{2}t_{j1}n_{s_1}-t_{j2}n_{s_2}+2t_{j3}n_{s_3}-4t_{j4}n_{s_4}+\cdots +(-1)^{j+1}2^{j-2}t_{jj}n_{s_j}, \tag{15}$$

where following the usual practice the exponents are reduced to the smallest integer form in that if $-J_{i-1}$ represents some collective i-site potential in the ground state of (1), then $u_{i-1}=\exp(-4J_{i-1}/kT)$. The high field groupings of Eq. (14) corresponding to Eq. (8) now become high field polynomials $f_{n_{s_1}}(u_1,\ldots,u_{n-1})$ defined by

$$\Lambda(\mu,u_1,\ldots,u_{n-1})=\sum_{n_{s_1}=0}^{\infty}f_{n_{s_1}}\mu^{n_{s_1}}, \tag{16}$$

and the relation corresponding to Eq. (11) becomes coefficient of

$$\{\mu^{n_{s_1}}u_1^{\gamma_1(n_{s_1},n_{s_2})}u_2^{\gamma_2(n_{s_1},n_{s_2},n_{s_3})}\cdots, u_{n-1}^{\gamma_{n-1}(n_{s_1},n_{s_2},\ldots,n_{s_n})}\} \sum_k[k:n_{s_1}:n_{s_2}:\cdots:n_{s_n}]. \tag{17}$$

The sum in (17) is the total number of strong embeddings ($N=1$) of all the subgraphs of L_q which possess n_{s_1}-points, n_{s_2}-lines, $\cdots$, and n_{s_n} simplexes s_n. There will of course be more sums of the type (17) in $f_{n_{s_1}}$ than there are sums of the type in Eq. (11) in the corresponding f_s. Essentially, each term in f_s is split up into a partition of graphs each containing the same number of $s_3, s_4, \cdots$, and s_n simplex subgraphs.

IV. GENERALIZED HIGH TEMPERATURE SIMPLEX EXPANSIONS

We now seek the generalization of Eq. (4) which corresponds to the generalized low temperature simplex expansion of Eq. (14). The form of the expansion (14) was derived on the basis of including a term $J_{j-1}\sum_{s_j\in L_q}\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_j}$ in Eq. (1), where each product of j-site variables is defined over the j-point simplexes $s_j \in L_q$. We can readily incorporate this feature in a general development analogous to Eq. (4) for a lattice L_q such that $s_1, s_2, \ldots, s_n \in L_q$, and the resulting expansion is (see also Sec. VII)

$$\Lambda_N(\mu,z_1,z_2,\ldots,z_{n-1}) =\left(\prod_{i=2}^{n}\left(\tfrac{1}{2}(1+z_{i-1})\right)^{NP(s_i)}\right)(1+\mu)^N \times\left(1+\sum_{n_{s_2}=1}\sum_{n_{s_3}=1}\cdots\sum_{n_{s_n}=1}\sum_j(j:n_{s_2}:n_{s_3}:\cdots:n_{s_n}) \times v_1^{n_{s_2}}v_2^{n_{s_3}}\cdots v_{n-1}^{n_{s_n}}\tau^{n_o}\right), \quad (z_i^2=u_i). \tag{18}$$

This expansion requires some clarification in that the graphical interpretation is considerably different from that in Eq. (4). In Eq. (18) $v_i=\tanh(J_i/kT)=(1-z_i)/(1+z_i)$, $i=1,2,\ldots,n-1$, and $P(s_i)$ is the lattice constant for weak embeddings of $s_i \in L_q$. The graph $j:n_{s_2}:n_{s_3}:\cdots:n_{s_n}$ is a graph made up of $n_{s_2}-s_2$ simplexes, $n_{s_3}-s_3$ simplexes, $\ldots$, and $n_{s_n}-s_n$ simplexes but any simplex s_{i-1} in this graph is not itself a subgraph of any of the simplexes $s_i, s_{i+1}, \ldots, s_n$ which may also be contained in the graph. Again the index j in (18) serves to distinguish topological types. The way in which the graphs in (18) are defined means that in a diagrammatic representation of a graph each of the simplicial complexes must be clearly distinguished. Finally $(j:n_{s_2}:n_{s_3}:\cdots n_{s_n})$ is again the number of weak embeddings of the graph in L_q.

To illustrate the graphs in Eq. (18) consider any lattice such that $s_3 \in L_q$, then the graphs contributing to the term v_1v_2 are shown in Fig. 1 together with their respective contributions to the coefficient, where the shaded triangles correspond to the s_3 simplexes and the lines to the s_2 simplexes. In Eq. (18) n_o is again the number of odd degree vertices in the graphs; however, we must redefine our notion of the degree of a vertex. Any single $s_j \in L_q$ covering the lattice sites $i_1, \ldots, i_j$ introduces the product of site variables $\sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_j}$ only once in the expansion, and consequently we can define

FIG. 1. The contributions of the graphs contributing to the coefficient of v_1v_2 in Eq. (18).

the degree of a vertex in any graph as the total number of simplexes sharing this particular vertex. Thus for the graph (a) in Fig. 2, which is made up of s_2 and s_3 simplexes the degrees of the vertices a, b, c, d, e, f, and g are 1, 3, 3, 2, 1, 1, and 2, respectively. This graph would contribute a term in τ^5 to the coefficient of $v_1^2v_2^3$ in Eq. (18). In the case of the three-dimensional face centered cubic lattice s_4 simplexes can occur, and this necessitates a clear distinction between the triangles and the tetrahedra. For this lattice the graph (b) in Fig. 2 is an example of a graph made up of one s_2 simplex, one s_3 simplex, and one s_4 simplex which contributes a term in τ^3 to the coefficient of $v_1v_2v_3$. A more detailed discussion of the expansion in Eq. 18 is included in Sec. VII where the expansion is developed with reference to a model of a lattice fluid.

V. THE HIGH TEMPERATURE-LOW TEMPERATURE SIMPLEX TRANSFORMATION

Starting from the simplex expansion of Eqs. (14) and (18) it should be possible to set up a transformation corresponding to Eqs. (9) and (10) relative to the same assumptions as made by Domb.[3] We consider a lattice L_q such that $s_1, s_2, \dots, s_n \in L_q$ and define two index sets J and K to be the set of odd and even integers, respectively, from the integers $2, 3, 4, \dots, n$. We readily observe that for any graph $j: n_{s1}: n_{s2}: \cdots : n_{s_n}$ in Eq. (18) (i) n_o is odd if Σn_{s_j}, $j\in J$ is odd, and (ii) n_o is even if Σn_{s_j}, $j \in J$ is even.

From (i) and (ii) the following form of Eq. (18) immediately commends itself:

$$\Lambda(\mu, z_1, \dots, z_{n-1}) = \left(\prod_{i=2}^{n}[\tfrac{1}{2}(1+z_{i-1})]^{P(s_i)}\right)(1+\mu) \times\left[1+\sum_{n_{s3}=1}^{\infty} v_2^{n_{s3}} \sum_{n_{s5}=1}^{\infty} v_4^{n_{s5}} \cdots \left(\sum_{n_{s_j}:\, j\in K} (j : n_{s2} : \cdots : n_{s_n}) \times \prod_{i\in K} v_{i-1}^{n_{s_i}} \tau^{n_o}\right)\right], \qquad (19)$$

where the final sum in Eq. (19) extends over all possible sets $\{n_{s_j}: j\in K\}$ for a fixed set $\{n_{s_j}: j \in J\}$. Clearly only odd powers of τ occur in this sum if the sum of the members of the fixed set $\{n_{s_j}: j\in J\}$ is odd, otherwise the sum contains only even powers of τ.

It will be convenient to define a sequence of partial partition functions $\Lambda_{\{n_{s_j}:\, j\in J\}}(\mu, z_1, z_3, \cdots)$ by

$$\Lambda_{\{n_{s_j}:\, j\in J\}} = \sum_{\{n_{s_j}:\, j\in K\}} (j : n_{s2} : n_{s3} : \cdots : n_{s_n}) \prod_{i\in K} v_{i-1}^{n_{s_i}} \tau^{n_o}, \qquad (20)$$

which defines one such partial partition function for each set $\{n_{s_j}: j\in J\}$. It follows from (i) and (ii) that the partial partition functions satisfy the following symmetry relations

$$\Lambda_{\{n_{s_j}:\, j\in J\}}(\mu) = -\Lambda_{\{n_{s_j}:\, j\in J\}}(\mu^{-1}), \quad \text{if } \textstyle\sum n_{s_j}: j\in J \text{ is odd}, \qquad (21)$$

and

$$\Lambda_{\{n_{s_j}:\, j\in J\}}(\mu) = \Lambda_{\{n_{s_j}:\, j\in J\}}(\mu^{-1}), \quad \text{if } \textstyle\sum n_{s_j}: j\in J \text{ is even}.$$

We can now write Eq. (19) in the form

$$\Lambda = \left(\prod_{i\in J}(1+v_{i-1})^{-P(s_i)}\right) \sum_{n_{s_j}:\, j\in J}\left((1+\mu)\prod_{i\in K}[\tfrac{1}{2}(1+z_{i-1})]^{P(s_i)} \times \Lambda_{\{n_{s_j}:\, j\in J\}}\right)\prod_{i\in J} v_{i-1}^{n_{s_i}}. \qquad (22)$$

We can transform the expression in parenthesis in the sum of Eq. (22) in the usual way to obtain a form corresponding to Eq. (5), which is

$$\Lambda = \left(\prod_{i\in J}(1+v_{i-1})^{-P(s_i)}\right) \sum_{n_{s_j}:\, j\in J}\left(\sum_{n_{s_j}:\, j\in K} \frac{\phi_{\{n_{s_j}:\, j\in J+K\}}(\mu)}{(1+\mu)^{\Sigma_j j n_{s_j}-1}} \times \prod_{i\in K}(1-u_{i-1})^{n_{s_i}}\right)\prod_{i\in J} v_{i-1}^{n_{s_i}}, \qquad (23)$$

where a polynomial $\phi_{\{n_{s_j}:\, j\in J+K\}}(\mu)$ is defined for each set of integers n_{s_j}, $j = 2, 3, \dots, n$. These polynomials are of maximum degree $\Sigma_j j n_{s_j}$ and have the following symmetry properties resulting from Eqs. (21):

$$\phi_{\{n_{s_j}:\, j\in J+K\}}(\mu^{-1}) = -\mu^{-\Sigma_j j n_{s_j}}\phi_{\{n_{s_j}:\, j\in J+K\}}(\mu), \quad \text{if } \textstyle\sum n_{s_j}: j\in J \text{ is odd}, \qquad (24)$$

and

$$\phi_{\{n_{s_j}:\, j\in J+K\}}(\mu^{-1}) = \mu^{-\Sigma_j j n_{s_j}}\phi_{\{n_{s_j}:\, j\in J+K\}}(\mu), \quad \text{if } \textstyle\sum n_{s_j}: j\in J \text{ is even}.$$

To obtain the expansion in Eq. (23) we have split the original expansion of Eq. (18) into two parts each of which satisfies a symmetric or antisymmetric relation in the context of Eqs. (21). It is quite feasible to deduce other forms of Λ, but generally much is to be gained from a form which can be defined on a symmetric set of polynomials. The symmetry of the polynomials $\phi_{(r)}(\mu)$ in Eq. (5) plays a vital role in applying the relations (9) and (10) to the derivation of the high field polynomials.

We now proceed by expressing Eq. (16) in the form

$$\Lambda(\mu, z_1, z_2, \dots, z_{n-1}) = \sum_{n_{s1}=0}^{\infty} \Phi_{n_{s1}}(z_1, v_2, z_3, v_4, \cdots)\mu^{n_{s1}}, \qquad (25)$$

which is simply a matter of making the substitution $z_j = (1-v_j)/(1+v_j)$ for $(j+1)\in J$, whence we can formally equate Eqs. (25) and (23). The expansion of Eq. (23)

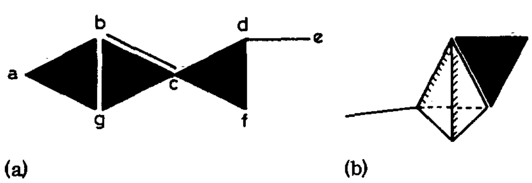

FIG. 2. The graphs (a) and (b) contribute to the coefficients of $v_1^2v_2^3$ and $v_1v_2v_3$, respectively, in the expansion of Eq. (18).

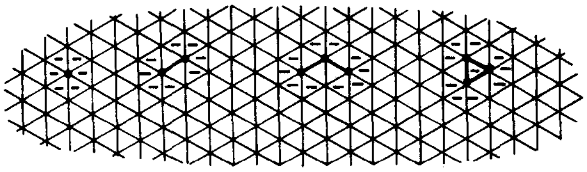

FIG. 3. The exponents of u_2 are equal to one half of the number of negative triangles which determines the variable u_2 for which the exponents will be integers.

yields the relation

$$\Lambda(\mu, 1, 0, 1, 0, \cdots) = 1 + \mu \tag{26}$$

and hence

$$\Phi_{n_{s_1}}(1, 0, 1, 0, \cdots) = 0, \quad n_{s_1} \geqslant 2. \tag{27}$$

On differentiating Eq. (26) l_1 times with respect to u_1, l_2 times with respect to $v_2, \cdots$, and l_{n-1} times with respect to v_{n-1} or u_{n-1} depending on whether n is odd or even, respectively, we obtain

$$\sum_{n_{s_1}=0}^{\infty} \Phi_{n_{s_1}}^{(l_1, l_2, \ldots, l_{n-1})}(1, 0, 1, 0, \ldots)\mu^{n_{s_1}} = \sum_{k_j:\, j \in K}^{\{l_j\}} A_{\{k_j\}} \phi_{(l_1, k_2, l_3, k_4, \ldots)}(\mu) / (1+\mu)^{2l_1 + \sum_{j\in J}(j+1)l_j + \sum_{j \in K}(j+1)k_j - 1} \tag{28}$$

where the constants $A_{\{k_j\}}$ are easily determined, and which yields the relations

$$\Phi_{n_{s_1}}^{(l_1, l_2, \cdots l_{n-1})}(1, 0, 1, 0, \ldots) = \text{coefficient of } \mu^{n_{s_1}} \text{ in } \sum_{k_j:\, j \in K}^{\{l_j\}} A_{\{k_j\}} \phi_{(l_1, k_2, l_3, k_4, \ldots)}(\mu) / (1+\mu)^{2l_1 + \sum_{j\in J}(j+1)l_j + \sum_{j \in K}(j+1)k_j - 1} \tag{29}$$

which are the generalizations of the relations in Eq. (10), which we call the simplex transformation.

We notice that both $f_s(u)$ in Eq. (8) and $\Phi_s(z_1, v_2, \cdots)$ in Eq. (25) are defined on identical sets of strong embeddings in L_q, that is, the same total set of graphs contribute to both functions; however, in Φ_s we have a different partition of these graphs whereby we identify a larger number of groups in the total set of graphs. For a lattice L_q which contains simplexes beyond s_2 Eq. (29) yields many more connecting equations on the total set of graphs which contribute to f_s; however, the larger number of groupings in Φ_s necessitates more connecting equations if the contributions from the individual groupings are to be determined. It seems to us that the simplex transformation of Eq. (29) may represent an additional and very useful technique in determining low temperature expansions for a number of Ising model problems.

VI. THE APPLICATION OF THE SIMPLEX TRANSFORMATION TO THE TRIANGULAR LATTICE

We now illustrate the simplex transformation of Eq. (29) with a specific example in which we take L_q to be the two-dimensional triangular lattice in which only nearest neighbor sites are connected. The lattice is shown in Fig. (3), and contains the s_3 simplex. For the graphs contributing to the expansion of Eq. (14) we will use the notation $k:s:r:t$ to denote a graph of type k containing s points, r lines, and t triangles. From Eq. (15) we find

$$\gamma_2 = 3s - 2r + 2t \tag{30}$$

and we can write

$$\Lambda_{(\mu, u_1, u_2)} = \sum_{s=0}^{\infty} [k:s:r:t] u_1^{3s-r} u_2^{3s-2r+2t} \mu^s. \tag{31}$$

The expansion variable u_2 carries the information relating to the number of triangles contained as subgraphs of $k:s:r:t$. The value of the exponent given by Eq. (30) is illustrated in Fig. 3 where the number of triangles containing a negative sign determines the index γ_2; the four graphs shown in Fig. 3 are sufficient to illustrate the general case. In Eq. (15) t_{31}, t_{32}, and t_{33} are, respectively, the number of triangles sharing a common point, a common line, and a common triangle, and are 6, 2, and 1, respectively.

The first four polynomials $f_s(u_1, u_2)$ are shown in diagrammatic form in Fig. 4.

The grouping of the contributions to f_4 shown in Fig. 4 corresponds to the general grouping of Eq. (17); graphs forming the coefficient of $u_1^m u_2^n$ are collected together, thus f_4 contains seven such groups, whereas in the grouping of the scheme in Eq. (11) the polynomial contains six groups. In general the number of groupings in the form of Eq. (17) must be determined by an examination of the lattice. In the case of the triangular lattice we illustrate the groupings that will occur in f_6 or f_7 by constructing the grids shown in Fig. (5). The numbers m in the top row of the grids represent possible terms in u_1^m, and the columns denote terms in u_2^n. The entries (t, r) in the grid represent graphs that can occur containing t triangles and r lines; consequently, in f_6 and f_7 there are 17, and 25 groupings, respectively, in the scheme of Eq. (17).

In this case the polynomials $\phi_{(r,t)}(\mu)$ for the triangular lattice, which appear in Eq. (23) are symmetric if t is even and antisymmetric if t is odd. Furthermore, the polynomials contain no term independent of μ, and therefore the maximum degree of $\phi_{(r,t)}$ is $2r + 3t - 1$, giving a maximum number $r + [\frac{3}{2}t]$ of independent coefficients in the polynomial, where [ ] denotes the integer

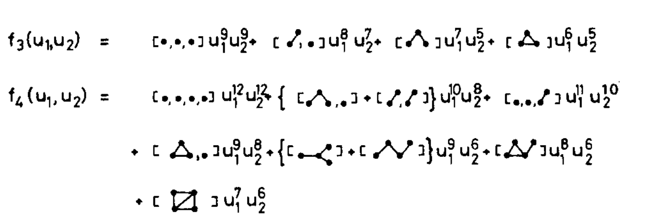

FIG. 4. The first four polynomials of Eq. (16) for the triangle lattice.

n \ m	18	17	16	15	14	13	12	11	10	9	
18	(0,0)										
16		(0,1)									
14			(0,2)	(1,3)							
12				(0,3)	(1,4)	(2,5)					
10					(0,4)	(1,5)	(2,6)	(3,7)			
8						(0,5)	(1,6)	(2,7)	(3,8)	(4,9)	
6							(0,6)				

n \ m	21	20	19	18	17	16	15	14	13	12	11	10	9
21	(0,0)												
19		(0,1)											
17			(0,2)	(1,3)									
15				(0,3)	(1,4)	(2,5)							
13					(0,4)	(1,5)	(2,6)	(3,7)					
11						(0,5)	(1,6)	(2,7)	(3,8)	(4,9)			
9							(0,6)	(1,7)	(2,8)	(3,9)	(4,10)	(5,11)	(6,12)
7								(0,7)	(1,8)				

FIG. 5. The grouping of graphs contributing to $f_6(u_1, u_2)$ and $f_7(u_1, u_2)$ for the triangular lattice in the scheme of Eq. (17).

part of. Given the polynomials $f_1(u_1, u_2), \ldots, f_5(u_1, u_2)$ (see Ref. 7), we can use Eq. (29) to determine the polynomials $\phi_{(r,t)}(\mu)$ for which $r + [\frac{3}{2}t] \leq 5$, which gives rise to the following set for (r, t):

$$\begin{aligned}(r, t) = (5, 0)(4, 0)(3, 0)(2, 0)(1, 0)(0, 0)&\\ (4, 1)(3, 1)(2, 1)(1, 1)(0, 1)&\\ (2, 2)(1, 2)(0, 2)&\\ (1, 3)(0, 3).&\end{aligned} \tag{32}$$

Thus to evaluate $f_6(u_1, u_2)$ we have 16 connecting equations for the 17 groupings shown in Fig. 5; consequently, we need only obtain the data for one of the entries in the grid to establish f_6. The simplest entry is (0, 6) which contains only one graph which is a hexagon of sites, and which has a strong embedding lattice constant of 1. Following the determination of f_6, we can extend the set in Eq. (32) by including

$$(r, t) = (6, 0)(5, 1)(3, 2)(2, 3) \text{ and } (0, 4), \tag{33}$$

thereby yielding 21 connecting equations for f_7 leaving the data corresponding to 4 elements in the grid of Fig. 5 to be determined.

Once we have obtained $f_s(u_1, u_2)$ we have obtained more configurational information about strong embeddings in L_q than we have in the single variable case $f_s(u_1)$. Alternatively, if a given set of high field polynomials $f_s(u_1)$ in Eq. (8) is known, we can use these to form additional connecting equations for the polynomials $f_s(u_1, u_2)$, which can be used to supplement the connecting equations obtained from Eq. (29). The number of additional connecting equations formed in this way is just the number of columns in the grid such as those shown in Fig. 5; since the entries in each grid must satisfy

$$\sum_{k,t}[k:s:r:t] = \sum_{k}[k:s:r], \quad r = 0, 1, 2, \cdots. \tag{34}$$

Recently Sykes and coworkers[11] have extended the series development of Eq. (8) for a variety of lattices. For the triangular lattice $f_{10}(u_1)$ has been obtained. The number of groupings of the type in Eq. (11) in this polynomial is 20, whereas the number of groupings in the scheme of Eq. (17) is 58. From the connecting equations (29) we can obtain 40 constraints on $f_{10}(u_1, u_2)$. Thus we can supplement these constraints with any 18 of the set of Eqs. (34) providing a full complement of 58 equations to determine $f_{10}(u_1, u_2)$. This procedure, which can be extended to other lattices, could also be used as a consistency check on the results of Sykes *et al.*[11] by altering the choice of constraints obtained from Eq. (34).

VII. A LATTICE GAS WITH TRIPLET POTENTIALS

Within the theory of classical fluids it has long been recognized that the N-particle potential function $U_N(\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_N)$ may not be expressed entirely in terms of additive pair potential functions. It has been suggested by several authors[14-18] that in order to obtain an accurate theory for the thermodynamics of simple fluids the potentials arising from three-body forces may have to be taken into account. The variety of network models that have been so successful in developing the theory of cooperative phenomena do not so far seem to have been extended with a view to including such triplet potentials. In our view such models would be well worth examining, particularly in the critical region.

To incorporate triplet potentials in a model of a lattice gas is simply to extend the original model of Yang and Lee.[19] Following Fisher's[20] notation we can define a potential function U_N of the fluid over a regular lattice of M sites, which includes triplet functions in the form

$$U_N = \sum_{1 \leq i < j \leq M} \phi(i, j) t_i t_j + \sum_{1 \leq i < j < k \leq M} \psi(i, j, k) t_i t_j t_k, \tag{35}$$

where $\phi(i, j)$ and $\psi(i, j, k)$ are pair and triplet potential functions for atoms located at lattice sites i, j and i, j, k, respectively. The variables t_i, $i = 1, 2, \ldots, M$ are the usual occupation variables taking values of 1 or 0.

For the purposes of a lattice model we can define the potential functions as follows:

$$\begin{aligned}\phi(i, j) &= \phi_r \quad \text{if } i \text{ and } j \text{ are } r\text{th neighbor sites}, \quad 1 \leq r \leq R\\ &= 0 \quad \text{otherwise},\end{aligned} \tag{36}$$

and

$$\begin{aligned}\psi(i, j, k) &= \psi_{pqs} \quad \text{if } i \text{ and } j \text{ are } p\text{th neighbors},\\ &\qquad j \text{ and } k \text{ are } q\text{th neighbors, and}\\ &\qquad k \text{ and } i \text{ are } s\text{th neighbors},\\ &\qquad \max(p, q, s) \leq S\\ &= 0 \quad \text{otherwise}.\end{aligned} \tag{37}$$

On transforming the expression (35) to the spin-$\frac{1}{2}$ Ising model spin up, spin down representation by using the relation

$$t_i = \tfrac{1}{2}(1+\sigma_i), \quad \sigma_i = \pm 1, \tag{38}$$

we obtain

$$U_N = \tfrac{1}{4}N\lambda_0 + \tfrac{1}{4}\lambda_0 \sum_{i=1}^{M}\sigma_i + \sum_{r=1}^{\max(R,S)} \tfrac{1}{4}(\phi_r + D_r)\sum_{(ij)_r}\sigma_i\sigma_j + \sum_{p\leq q\leq s}^{S}\ \sum_{(ijk)_{pqs}} \tfrac{1}{8}\psi_{pqs}\sigma_i\sigma_j\sigma_k, \tag{39}$$

where γ_0 and D_r, $r=1,2,\ldots,S$ are constants formed from finite sums of the interaction parameters ϕ_r and ψ_{pqs}, and where $(ij)_r$ denotes the summation over rth neighbor pairs, and $(ijk)_{pqs}$ denotes the sum over the triplet sites defined in Eq. (37). Following the usual procedure, we can consider Eq. (39) to define the Hamiltonian of a spin equivalent Ising model for the lattice fluid which incorporates triplet potential functions. In this representation of the problem the canonical partition function of the spin equivalent Ising model is simply related to the grand canonical partition function of the original fluid model (see Fisher[20]); in this way we can determine the thermodynamic functions of interest for the model of the fluid.

The two-dimensional triangular lattice and the three-dimensional face centered cubic lattice form simple examples of Eq. (39). For these two lattices we can form a triplet potential model which will only involve nearest neighbor links of the lattice. This is the case when $R=S=1$, and consequently it is sufficient to examine the properties of a spin equivalent Ising model defined by a Hamiltonian $\mathscr{H}_N$ in the form

$$\mathscr{H}_N = -J_1 \sum_{(ij)_1}\sigma_i\sigma_j - J_2 \sum_{(ijk)_{111}}\sigma_i\sigma_j\sigma_k - f\sum\sigma_i, \tag{40}$$

where J_1, J_2, and f are constants which can be readily determined through the transformation Eqs. (38) and (39). The problem is to evaluate the partition function

$$Z_N = \mathrm{Tr}\exp(-\beta\mathscr{H}_N). \tag{41}$$

The Hamiltonian of Eq. (40) is clearly an example of the class of Hamiltonians defined in Eq. (1) in which $n=3$. The simplex transformation discussed in Sec. V could conveniently be applied to the problem of developing the low temperature expansion of the partition function. On combining the simplex transformation method with the present data available on the conventional Ising model problem it would be possible to obtain the low temperature expansions of the thermodynamic functions of this triplet model up to the same order of perturbation as recently achieved by Sykes and coworkers.[11] Some details of these calculations have recently been given by Griffiths and Wood.[21]

The high temperature expansion of Eq. (41) is an example of the expansion of Eq. (18) in which $n=3$. The form of this expansion can be obtained in the usual way by using the Van der Waerden identities

$$\exp(J_1\beta\sigma_i\sigma_j) = \cosh(\beta J_1) + \sigma_i\sigma_j\sinh(\beta J_1), \tag{42}$$

and

$$\exp(\beta J_2\sigma_i\sigma_j\sigma_k) = \cosh(\beta J_2) + \sigma_i\sigma_j\sigma_k\sinh(\beta J_2). \tag{43}$$

The graphs contributing to this expansion are the simplicial complexes of the type shown in Fig. 1. We have derived examples of these expansions by obtaining some of the initial terms in the series expansions for the free energy F per site in zero field ($f=0$), and for the susceptibility per site in zero field χ_0 of the Ising model of Eq. (40). The expansions are listed below.

Zero field free energy

(a) triangular lattice

$$-F/kT - \log 2 - 3\log\cosh(\beta J_1) - 2\log\cosh(\beta J_2)$$
$$= 2v_1^3 + 3v_1^4 + 15v_1^2v_2^2 + 6v_1^5 + 102v_1^3v_2^2 + 11v_1^6 + 543v_1^4v_2^2 + 90v_1^2v_2^4 + v_2^6 + 24v_1^7 + 2520v_1^5v_2^2 + 1120v_1^3v_2^4 + \cdots \tag{44}$$

(b) face centered cubic lattice

$$-F/kT - \log 2 - 6\log\cosh(\beta J_1) - 8\log\cosh(\beta J_2)$$
$$= 8v_1^3 + 4v_1v_2^2 + 33v_1^4 + 336v_1^2v_2^2 + 24v_2^4 + \cdots \tag{45}$$

Zero field susceptibility

(a) triangular lattice

$$kT\chi_0 = 1 + 6v_1 + 30v_1^2 + 6v_2^2 + 138v_1^3 + 120v_1v_2^2 + 606v_1^4 + 1134v_1^2v_2^2 + 30v_2^4 + 2586v_1^5 + 7890v_1^3v_2^2 + 1200v_1v_2^4 + 10818v_1^6 + 44046v_1^4v_2^2 + 18654v_1^2v_2^4 + 180v_2^6 + \cdots \tag{46}$$

(b) face centered cubic lattice

$$kT\chi_0 = 1 + 12v_1 + 132v_1^2 + 72v_2^2 + 1404v_1^3 + 2880v_1v_2^2 + 14652v_1^4 + 66112v_1^2v_2^2 + 6280v_2^4 + \cdots \tag{47}$$

VIII. SUMMARY

We have considered in some detail the methodology which surrounds the techniques of developing power series expansions for the classical network models which are used frequently in the theory of critical phenomena. It has been shown that the powerful and commonly adopted technique known as the high temperature-low temperature transformation is a special case of a generalized transformation which we have called the simplex transformation. This transformation can form the basis of an approach to many problems which are defined in terms of lattice models.

The conventional Ising model problem which has been well studied using a variety of techniques can be considered afresh in the light of this new transformation. For example the present work on the inclusion of longer range pair interactions[9] could be extended using this technique. We have also shown that the simplex transformation is ideally suited to the treatment of lattice gas models in which three-body forces are considered; in a lattice such as the face centered cubic lattice four-body potential functions could also be included without many additional complications.

The present authors are presently engaged in applying the simplex transformation to problems in which three-body potential functions are included.

ACKNOWLEDGMENTS

One of us (H. P. G.) would like to thank the S. R. C. for the award of a maintenance grant.

[1]C. Domb, Adv. Phys. **9** (34, 35) (1960).
[2]M. E. Fisher, Rep. Prog. Phys. **30**, 615 (1967).
[3]C. Domb, Proc. R. Soc. A **199**, 199 (1949).
[4]M. F. Sykes, thesis, Oxford University, 1956.
[5]C. Domb, and B. J. Hiley, Proc. R. Soc. A **268**, 506 (1962).
[6]A. J. Wakefield, Proc. Camb. Philos. Soc. **47**, 419 (1952).
[7]M. F. Sykes, J. W. Essam, and D. S. Gaunt, J. Math. Phys. **6**, 283 (1965).
[8]J. F. Nagle, J. Math. Phys. **9**, 1007 (1968).
[9]N. W. Dalton and D. W. Wood, J. Math. Phys. **10**, 1271 (1969).
[10]M. F. Sykes, D. S. Gaunt, J. W. Essam, and D. L. Hunter, J. Math. Phys. 14, 1060 (1973).
[11]M. F. Sykes, D. S. Gaunt, S. R. Mattingly, J. W. Essam, and C. J. Elliot, J. Math. Phys. 14, 1066 (1973).
[12]The concepts of weak and strong embeddings are defined and discussed in detail by M. F. Sykes *et al.* J. Math. Phys. **7**, 1557 (1966).
[13]The concept of a space type relating to embeddings of graphs in L_q is discussed in detail by Domb[1], Chap. 5.
[14]G. S. Rushbrooke, *Physics of Simple Liquids* (North-Holland, Amsterdam, 1968).
[15]J. S. Rowlinson, Discuss. Faraday Soc. **40**, (1965).
[16]G. S. Rushbrooke, Discuss. Faraday Soc. **43**, (1967).
[17]G. S. Rushbrooke and M. Silbert, Mol. Phys. **12**, 505 (1967).
[18]J. S. Rowlinson, Mol. Phys. **12**, 513 (1967).
[19]C. N. Yang and T. D. Lee, Phys. Rev. **87**, 404 (1952); Phys. Rev. **87**, 410 (1952).
[20]M. E. Fisher, *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, 1964), Vol. 7.
[21]H. P. Griffiths and D. W. Wood, J. Phys. C , to appear Aug. 1973.