JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Characters of the Poincaré Group

GÉRARD FUCHS AND PIERRE RENOUARD
Centre de Physique Théorique de l'Ecole Polytechnique, 17, rue Descartes, Paris V, France

(Received 16 May 1966; Revised Manuscript Received 31 January 1969)

We calculate the characters of the Poincaré group P as solutions of differential equations in a way which is valid for a large class of Lie groups. We discuss the solutions of linear differential equations on an analytic manifold in a space of distributions. Then, we calculate the characters of $SL(2, R)$ and the central distributions on P. Finally, we give the characters of all unitary irreducible representations of P, including mass 0, and some expressions which may be "characters" of nonunitary representations.

INTRODUCTION

The calculation of the characters of the irreducible representations of semisimple complex Lie groups as solutions of differential equations has already been carried out by Berezin.[1] His method, however, rests on specific properties of these groups.

We shall calculate the characters of the Poincaré group P using the following method, which is valid for a large class of Lie groups. We look for distributions (a) which are on P [$\in \mathfrak{D}'(P)$], (b) which are central (class) distributions, (c) which are eigendistributions of the Laplace operators $\mathbf{P}_2$ and $\mathbf{W}^2$ of P, and (d) finally, for unitary representations, which are extremal distributions of positive type.

In Sec. 1, we justify and discuss our method; we show how to handle distributions and differential operators on a group; we give some general results and a theorem on the solutions of a system of partial derivative equations of first order in the space of distributions.

In Sec. 2, as a first application, we calculate the characters of $SL(2, R)$ (for which Berezin's method already does not work). In Sec. 3, we give a local expression of the central distributions on P, some of which appear to have a nonzero transversal order.

Section 4 is devoted to the explicit calculation of the characters of P: (a) Outside $m^2 = w^2 = 0$, we get the characters of all unitary irreducible representations and, in addition, other solutions which can be looked at as the characters of some known nonunitary representations. (b) For $m^2 = w^2 = 0$, the initial program fails. We have not only to add to $\mathbf{P}^2$ and $\mathbf{W}^2$ the helicity operator but, moreover, to change our space of solutions, which we take to be the dual of a subspace of $\mathfrak{D}(\bar{P})$. We then get such pathologies as central "distributions," which are no longer constant on conjugation classes, and "distributions" of positive type which are not bounded.

We obtain the characters of the unitary helicity representations up to an unknown factor.

Appendices A and C contain some technical calculations of limits, which appear from the fact that we have to give a meaning to distributions whose supports are singular surfaces. Appendix B gives some formulas related to our parametrization of P.

1. GENERALITIES

A. Definition of a Character

Let G be any locally compact group and U a strongly continuous representation of G in a Banach space $\mathfrak{B}$.

For any positive measure μ on G such that

$$\int_G \|U(g)\|\, d\mu(g) < \infty \tag{1.1}$$

[$\|\ \|$ is the norm in $\mathfrak{L}(\mathfrak{B})$], we define the continuous linear operator

$$U(\mu) = \int_G U(g)\, d\mu(g).$$

By means of linearity, we can extend this definition to the complex vector space generated by the μ satisfying condition (1.1); this space is an algebra for convolution: the algebra $M_U^1(G)$ of the U-bounded measures on G.

For some of these μ, the operators $U(\mu)$ may have a trace, i.e., be such that there exist two sequences $x_i \in \mathfrak{B}$ and $x_j' \in \mathfrak{B}'$ (dual of $\mathfrak{B}$) with $|x_i|$, $|x_j'| < A$, $\forall\ i, j$, such that we have, for any x in $\mathfrak{B}$,

$$U(\mu)x = \sum_i \lambda_i(x_i', x)x_i,$$

with

$$\sum_i |\lambda_i| < \infty.$$

By definition, then,

$$\operatorname{Tr} U(\mu) = \sum_i \lambda_i(x_i', x_i).$$

The space of these μ, which we call $A_U(G)$, is a vector space and a 2-sided ideal of $M_U^1(G)$ (the product of a trace-class operator by a bounded operator is again a trace-class operator).

The mapping $\mu \to \text{Tr}\, U(\mu)$ is a linear form from $A_U(G)$ to C. This form is called the character of U, if U is a completely irreducible representation.

If U is an unitary representation (on Hilbert space $\mathcal{H}$), $M_U^1(G) = M^1(G)$ (bounded measures on G). In the case where G is a Lie group, we can consider the subspace $\mathfrak{D}(G) \subset M^1(G)$ of infinitely differentiable functions with compact support (the injection being made with help of the left-invariant Haar measure dg).

But we do not yet know all general conditions on G and U such that $\mathfrak{D}(G) \subset A_U(G)$. We know that the inclusion is true if G is semisimple and U a UIR (unitary irreducible representation).[2] In this case, we even know that the restriction of χ to $\mathfrak{D}(G)$ is a locally integrable function, which is analytic on the regular elements of G. We also know that for the Poincaré group there are representations for which the inclusion does not hold.[3]

When $\mathfrak{D}(G) \subset A_U(G)$, then the restriction χ of a character to $\mathfrak{D}(G)$ is continuous for the topology of $\mathfrak{D}(G)$, i.e., the restriction is a distribution.

B. Properties of Characters

When the character of a representation U is a distribution given by

$$\langle \chi, \varphi \rangle = \text{Tr}\, U(\varphi),$$
$$U(\varphi) = \int_G U(g)\varphi(g)\, dg,$$

with G any Lie group (we take it as unimodular, for simplicity), dg the (right- and left-)invariant Haar measure on G, and $\varphi(g) \in \mathfrak{D}(G)$, it has the following properties:

(*a*) *χ is a central distribution.* If we call δ_γ the Dirac measure on G with support γ,

$$\langle \delta_\gamma, \varphi \rangle = \varphi(\gamma),$$

we have

$$\langle \chi, \varphi \rangle = \langle \chi, \delta_\gamma * \varphi * \delta_{\gamma^{-1}} \rangle$$

or

$$\langle \delta_{\gamma^{-1}} * \chi * \delta_\gamma, \varphi \rangle = \langle \chi, \varphi \rangle, \quad \forall\, \varphi, \tag{1.2}$$

where $*$ means convolution. For any γ in a neighborhood of the identity, we have $\gamma = e^{tX}$ (X in the Lie algebra $\mathfrak{G}$ of G); we define

$$\langle \delta'_{X_i}, \varphi \rangle = \frac{d}{dt}\, \varphi(e^{-tX_i})\big|_{t=0},$$

where X_i is a basis of $\mathfrak{G}$. The infinitesimal form of (1.2) is equivalent to the set of n differential equations

$$\langle \delta'_{X_i} * \chi - \chi * \delta'_{X_i}, \varphi \rangle = 0, \quad \forall\, \varphi. \tag{1.3}$$

(*b*) *χ is an eigenvector of Laplace operators.* More generally, we have an isomorphism between the enveloping algebra $\mathfrak{U}(\mathfrak{G})$ of $\mathfrak{G}$ and the algebra (for convolution) of distributions on G with support identity. In particular, a Casimir operator Q_i of G [center of $\mathfrak{U}(\mathfrak{G})$] of degree n_i acts on $\mathfrak{D}'(G)$ by means of $\delta_{Q_i}^{(n_i)} * T$, where $T \in \mathfrak{D}'(G)$ and $\delta_{Q_i}^{(n_i)}$ is the distribution of order n with support e. The differential operators D_{Q_i} associated with the $\delta_{Q_i}^{(n_i)}$ are called the Laplace operators of G. We have for a character

$$\langle \delta_{Q_i}^{(n_i)} * \chi, \varphi \rangle = q_i \langle \chi, \varphi \rangle, \quad \forall\, \varphi, \tag{1.4}$$

where the $q_i \in \mathbf{C}$ are given by $U(Q_i) = q_i \mathbf{1}$.

(*c*) *Translation by the elements of the center of G:* If $g_0 \in C(G)$, the operators $U(g_0)$ commute with the representation and, thus, are multiples of the identity

$$U(g_0) = \alpha \mathbf{1}.$$

With $\varphi_{g_0}(g) = \varphi(g_0 g)$, we have

$$\langle \chi, \varphi_{g_0} \rangle = \alpha \langle \chi, \varphi \rangle. \tag{1.5}$$

Now if the representation U is unitary, the character has additional properties.

(*d*) *χ is a distribution of positive type*, $\chi \gg 0$. With $\tilde{\varphi}(g) = \overline{\varphi(g^{-1})}$, we have

$$\langle \chi, \varphi * \tilde{\varphi} \rangle \geq 0, \tag{1.6}$$

because $U(\varphi * \tilde{\varphi})$ is a positive operator. As a consequence, χ has Hermitian symmetry, i.e.,

$$\langle \chi, \varphi \rangle = \overline{\langle \chi, \tilde{\varphi} \rangle}. \tag{1.7}$$

(*e*) *χ is a bounded distribution.* That means that χ is a linear form on the space $\mathfrak{D}_{L^1}(G)$ of infinitely differentiable functions belonging as also their derivatives (defined by left invariant vector fields) to the space $L^1(G)$ (integrable functions). It is a consequence of (d) (see Ref. 4; the proof given for R^n can be carried over to a Lie group). Since we have not found a practical way to take the condition $\chi \gg 0$ into account, the condition that χ is bounded, weaker but easier to handle, will be very useful.

(*f*) *χ is an extremal distribution.* If $\chi = \chi_1 + \chi_2$ with $\chi_1, \chi_2 \gg 0$, then[5]

$$\chi_1 = a\chi, \quad \chi_2 = (1 - a)\chi, \quad 0 < a < 1. \tag{1.8}$$

(*g*) *The eigenvalues of Laplace operators [chosen to be symmetric, i.e., $Q_i \equiv P(X_i) = \bar{P}(-X_i)$, P a polynomial] are real.*

C. The Rules of the Game

In the case of the Poincaré group P and some subgroups, we look for the distribution solutions of (1.3) and (1.4), obeying (1.5); if some solutions are bounded, we shall investigate properties (1.6)–(1.8).

Now the only thing we can assert is that, given a solution χ of (1.3) and (1.4) obeying (1.5) of positive type and extremal, we know how to build a factorial unitary representation, quasi-equivalent to an irreducible unitary representation whose character is χ (G is of type I; see Ref. 6); this is Gel'fand construction.[5,6]

We cannot be sure, however, we have found all the characters of G, as we do not know when $\mathfrak{D}(G) \subset A_U(G)$ (even if G is semisimple, there are no results when U is not unitary) and, moreover, there can exist characters whose restriction to $\mathfrak{D}(G)$ is 0 ($\chi = 0$ has all good properties).

Conversely, given a solution which is not of positive type, it remains an open question to say whether or not it is a character of a nonunitary representation. We always find a one-to-one correspondence between the solutions of our problem and the known representations for the groups $SL(2, R)$ and $SL(2, C)$. For the Poincaré group, we find objects which are candidates to be the characters of some known nonunitary representations and others we cannot associate with any representation we know.

At this stage, we leave aside the specific problem of mass zero representations of P, for which we work in a space of "distributions" $\mathfrak{D}_0'$, which is a quotient of $\mathfrak{D}'$ and which is only suggested by the specific form of our equations.

D. Method

We have to solve, on the manifold G, the system of partial differential equations

$$\delta'_{X_i} * T - T * \delta'_{X_i} = 0, \tag{1.9}$$

$$\delta_{Q_k}^{(n_k)} * T = q_k T. \tag{1.10}$$

This means, given an open covering U_i (each U_i being the domain of a chart), solve (1.9) and (1.10) restricted to each U_i, which gives (for each set of values q_k) a vector space of solutions $V_i \in \mathfrak{D}'(U_i)$, and then find the distributions $T \in \mathfrak{D}'(G)$ such that $T|_{U_i} \in V_i$.

To easily solve (1.9) and (1.10) (restricted to some U_i), we are led to perform definite changes of variables which appear to choose conjugation classes of the group for coordinate surfaces. Unfortunately, the changes of variables are not regular everywhere, in general, and so we use the following step-by-step method: First, we define an open set $U_i^1 \subset U_i$ in which the suitable change of variables can be done; the resolution of (1.9) and (1.10) in U_i^1 gives a vector space of solutions $V_i^1 \subset \mathfrak{D}'(U_i^1)$. Next, we define U_i^2 ($U_i^1 \subset U_i^2 \subset U_i$), and we search (using, in particular, limiting processes) for the solutions in V_i^1, which are the restrictions to U_i^1 of solutions in U_i^2, and for the extended solutions. In the same way, we define (if necessary)

$$U_i^1 \subset U_i^2 \subset U_i^3 \subset \cdots \subset U_i^{k_i} = U_i$$

and, successively, we extend our solutions from U_i^k to U_i^{k+1}. The choice of the U_i^k is determined only by practical considerations: They correspond to the domains of validity of explicit manipulations. Note that $0 \in V_i^k$ [$V_i^k \subset \mathfrak{D}'(U_i^k)$ is the space of solutions in U_i^k] and that the space of solutions in U_i^{k+1} with support in $U_i^{k+1} \backslash U_i^k$ is not necessarily 1 dimensional; thus, if a given element in V_i^k has "extensions" in V_i^{k+1}, it may have several of them.

In practice, we first solve equations (1.9) "locally"; that is, we give in U_i^1 all solutions of (1.9) and, in each U_i^k, all solutions of (1.9) which are zero in U_i^{k-1}. We do not give the global solutions of (1.9).

Next we apply the differential operators (1.10) to these local solutions of (1.9) [the operators (1.10), restricted to such special distribution, have simple expressions] and so solve "locally" the system (1.9) and (1.10): We get (low-dimensional) vector spaces of solutions. The extension of solutions from U_i^k to U_i^{k+1} is performed by techniques adapted to each special case.

The complete description of all the open sets U_i^k (or, equivalently, of the domains of validity of expressions we write) is tedious and generally not useful. We omit or abbreviate it in different, obvious ways.

E. Local Resolution of Equation (1.9) (Regular Case)

The system (1.9) can be written

$${}^tD_i T = 0, \tag{1.9'}$$

where the ${}^tD_i$ are the transposed operators of the homogeneous partial differential operators D_i, defined by $D_i\varphi = -\delta'_{X_i} * \varphi + \varphi * \delta'_{X_i}$ [$\varphi \in \mathfrak{D}(G)$]; that is,

$$\langle {}^tD_i T, \varphi\rangle = -\langle T, D_i\varphi\rangle, \quad T \in \mathfrak{D}'(G).$$

In each open set U_i (with coordinates z^j, $j = 1 \cdots n$), the operators D_i are of the form $D_i = a_i^j(z)\partial/\partial z^j$, where the a_i^j are C^∞ functions.

To solve (1.9′) we first perform a change (depending on the coordinate system) of the unknown distribution which leads to homogeneous equations.

We note that a Haar measure on G is a distribution μ, belonging to Lebesgue class (that is, defined by a C^∞ nonvanishing, differential form of maximal order), such that

$$ {}^tD_i\mu = 0, \tag{1.11}$$

so that, in each U_i, we have

$$\mu = m(z)\, dz^1 \wedge \cdots \wedge dz^n, \tag{1.12}$$

where $m(z)$ is a C^∞ nonvanishing function, and we write (1.11) as

$$\frac{\partial}{\partial z^j}(a_i^j(z)m(z)) = 0. \tag{1.13}$$

Thus, if we define for each distribution T

$$\tilde{T}(z) = (m(z))^{-1}T, \tag{1.14}$$

Eq. (1.9′) becomes

$$a_i^j(z)\frac{\partial}{\partial z^j}\tilde{T}(z) = 0. \tag{1.15}$$

We must notice that these equations, from the Lie algebra commutation relations and the fact that the right- and left-hand infinitesimal translations commute, form a complete system.

Now suppose that, in some open U, the rank $r(z)$ of the $n \times n$ determinant $|a_i^j(z)|$ is equal to a given constant r. [At each point $g \in G$, the rank of the system (1.9) is equal to the dimension of the conjugation class of g.] Then there are $n - r$ C^∞ functions $v^k(z)$, such that $a_i^j(z)\partial v^k(z)/\partial z^j = 0$ and every function f solution of $a_i^j(z)\partial f/\partial 2^j = 0$ is a function of the v, i.e., $f(z) = f(v(z))$. Suppose, furthermore, that there exist r C^∞ functions $x^1(z)$ such that the v's and x's can be chosen as local coordinates in U_i; that is,

$$\frac{\partial(v^1\cdots v^{n-r}, x^1\cdots x^r)}{\partial(z^1\cdots z^n)} \neq 0, \quad \text{in } U_i.$$

Definition: We say that a distribution T depends only on the v, if there exists $S \in \mathfrak{D}'(R^{n-r})$ such that

$$\langle\tilde{T}(v, x), \varphi\rangle = \left\langle S, \int\varphi(\cdot, x)\, dx\right\rangle, \quad \varphi \in \mathfrak{D}(U).$$

If T is such a distribution, we denote it by $S(v)$:

$$T = S(v) = \tilde{T}(v, x)m(v, x),$$

$$\langle T, \varphi\rangle = \langle S(v), \varphi\rangle = \left\langle S, \int \varphi(\cdot, x)m(\cdot, x)\, dx\right\rangle.$$

Lemma 1: For given v and a given S, $S(v)$ is independent of the (admissible) choice of the functions x.

Proof: Let $y^1(v, x)$ be another possible choice, and denote by $S(v)_{\{y\}}$ the distribution defined from S using the y's:

$$\langle S(v)_{\{y\}}, \varphi\rangle = \left\langle S, \int\varphi(\cdot, y)m(\cdot, y)\, dy\right\rangle$$

$$= \left\langle S, \int\varphi(\cdot, y(\cdot, x))m(\cdot, y(\cdot, x))\left|\frac{\partial y}{\partial x}\right| dx\right\rangle$$

$$= \left\langle S, \int\varphi(\cdot, x)m(\cdot, x)\, dx\right\rangle = \langle S(v'), \varphi\rangle,$$

because

$$m(v, x) = m(v, y(v, x))\left|\frac{\partial(v, y)}{\partial(v, x)}\right| = m(v, y(v, x))\left|\frac{\partial y}{\partial x}\right|.$$

Lemma 2: The solutions of (1.9) in U_i are all distributions of the form $S(v)$ (if the manifolds $v^k = Cte$ are connected).

Proof: In the coordinate system (v, x), Eq. (1.15) becomes

$$a_1^j(v, x)\frac{\partial}{\partial x^j}\tilde{T}(v, x) = 0,$$
$$i = 1, \cdots, n, \quad j = 1, \cdots, r,$$

with some nonvanishing $r \times r$ determinant: $|a_{i_k}^j(v, x)|$ $(k = 1 \cdots r)$; thus, (1.15) is equivalent to

$$\frac{\partial}{\partial x^j}\tilde{T}(v, x) = 0, \quad \forall j = 1, \cdots, r.$$

Thus, $\tilde{T}(v, x) = S \otimes dx^1 \wedge \cdots \wedge dx^r$ for some S.

Remark 1: We have $S(v) = S \otimes \alpha^r$, where α^r is any differential r-form such that $\mu = \alpha^r \wedge dv^1 \wedge \cdots \wedge dv^{n-r}$.

Remark 2: If we use the formal notation

$$\langle S, \varphi\rangle = \int S(v)\varphi(v)\, dv, \quad [\varphi \in \mathfrak{D}(R^{n-r})],$$

we have, for $\varphi \in \mathfrak{D}(U)$,

$$\langle S(v), \varphi\rangle = \int S(v)\, dv \int\varphi(v, x)m(v, x)\, dx$$
$$= \int S(v)\varphi(v, x)\, d\mu(v, x),$$

and so the result of the paragraph is that, under the given regularity conditions, we can use the formal notation

$$\langle T, \varphi\rangle = \int T(z)\varphi(z)\, d\mu(z).$$

The operator ${}^tD_i$ is then obtained from D_i by a formal integration by parts (and is formally identical to D_i by invariance of the Haar measure) and the solutions (as in the case of functions) are distributions which depend only on characteristic manifolds. In particular, distributions defined by functions are of the form $f(v)\mu$.

Examples: Let

$$S_1 = \delta_{a^{i_1}}^{(k_{i_1})} \cdots \delta_{a^{i_p}}^{(k_{i_p})}, \quad p \le n - r,$$

and

$$S_2 = \delta_{a^{i_1}}^{(k_{i_1})} \cdots \delta_{a^{i_p}}^{(k_{i_p})} S_{i_{p+1}\cdots i_{p+q}}, \quad p + q \le n - r,$$

respectively, be the distributions

$$\varphi \to \frac{\partial^{k_{i_1}}}{(\partial v^{i_1})^{k_{i_1}}} \cdots \frac{\partial^{k_{i_p}}}{(\partial v^{i_p})^{k_{i_p}}} \left(\int \varphi(v) \prod_{l \ne i_1 \cdots i_p} dv^l \right) \Big|_{v^i = a^{i_j}}, \quad \varphi \in \mathfrak{D}(R^{n-r}),$$

and

$$\varphi \to \left\langle S, \frac{\partial^{k_{i_1}}}{(\partial v^{i_1})^{k_{i_1}}} \cdots \frac{\partial^{k_{i_p}}}{(\partial v^{i_p})^{k_{i_p}}} \times \left(\int \varphi(v) \prod_{l \ne i_1 \cdots i_{p+q}} dv^l \right) \Big|_{v^{i_j} = a^{i_j}} \right\rangle, \quad j = 1, \cdots, p.$$

The corresponding distributions in $\mathfrak{D}'(U)$ defined above are denoted as

$$S_1(v) = \delta^{(k_{i_1})}(v^{i_1} - a^{i_1}) \cdots \delta^{(k_{i_p})}(v^{i_p} - a^{i_p})$$

and

$$S_2(v) = \delta^{(k_{i_1})}(v^{i_1} - a^{i_1}) \cdots \delta^{(k_{i_p})}(v^{i_p} - a^{i_p}) \times S(v^{i_{p+1}} \cdots v^{i_{p+q}}). \tag{1.16}$$

Note that $S_1(v)$ depends, even for $k_{i_j} = 0$, on the functions v, and not only on the manifold $W = \{(v^{i_j} - a^{i_j}) = 0, j = 1, \cdots, p\}$.

F. Local Resolution of (1.9). Some Results in Irregular Cases

Suppose, now, that $r(z) = \operatorname{rank} |a_i^j(z)|$ is equal to a constant r_1 in some open set U, except on some closed submanifold $W \subset U$ where $r(z) = r_2 < r_1$. Again, assume regularity conditions: There are $n - r_1$ characteristic manifolds $v^k(z)$ of ${}^tD_i$ defined on U, and there exist r_1 functions $x^1(z)$ such that $\partial(v, x)/\partial(z) \ne 0$ in U. Assume also that the manifolds $(\{v^k = Cte\})$ are connected.

The considerations of Sec. 1E apply in $U' = U \backslash W$, and so (1.9) is already solved in U', as long as the manifolds $\{v^1 = Cte\}$ remain connected in U'. The lowering of the rank of the system on W leads to search for solutions $T \in \mathfrak{D}'(U)$ such that either supp $T \subset W$ or, if some manifolds $\{v^k = Cte\}$ do not remain connected in U', is localized on such manifolds and not "constant" on it. We study two cases which lead to these two possibilities. These two cases are characterized by the fact that the n-vector fields corresponding to the operators D_i are tangent to W at each point of W.

Suppose first that W is a characteristic manifold of the system ${}^tD_i$, that is, there exist p functions t^l such that

$$W = \{z; t^l(v(z)) = 0, 1 \le l \le p\},$$

and secondly that there are $n - p - r_1$ functions v', of the v only, such that (t, v', x) is a coordinate system in a neighborhood of W (again called U). Equation (1.15) is written

$$a_i^j(t, v', x) \frac{\partial}{\partial x^j} \tilde{T}(t, v', x), \quad 1 \le i \le n, \quad 1 \le j \le r_1,$$

and so every distribution of the form (1.16) with support in W [i.e., $\sum_k \delta^{(k)}(t)\hat{T}_k(v')$, finite sum in any bounded open set] is a solution of (1.9). But they are not related to the lowering of the rank of the system on W. Now the operators ${}^tD_i$ have restrictions to W which, in the coordinate system (v', x) on W, is written

$${}^tD_{i|W} = a_i^j(0, v', x) \frac{\partial}{\partial x^j}. \tag{1.17}$$

The system (1.17) is obviously complete, its rank is r_2, and so it admits [with the $n - p - r_1$ functions $v_{|W}$] $r_1 - r_2$ new C^∞ characteristic manifolds $u^i(v', x)$, $i = 1, \cdots, (r_1 - r_2)$.

Let the $\hat{u}$ be functions defined in a neighborhood of W such that $\hat{u}_{|W} = u$ and such that there exist functions x' so that $(t, v', \hat{u}, x')$ is a coordinate system.

For any $S \in \mathfrak{D}'(R^{n-p-r_2})$ define

$$T = S(u', \hat{u})\delta^{(k_1)}(t^1) \cdots \delta^{(k_p)}(t^p) \in \mathfrak{D}'(U)$$

by

$$\langle T, \varphi \rangle = \left\langle S(v', u), \left(\prod_{j=1}^{p} \frac{\partial^{k_j}}{(\partial t^j)^{k_j}} \right) \int \varphi m(t, v', u, x') \, dx' \big|_{t=0} \right\rangle, \quad \varphi \in \mathfrak{D}(U). \tag{1.18}$$

Equation (1.18) is obviously independent of the choice of the functions x' and, if $k_j = 0$, $\forall j$, Eq. (1.18) is independent of the functions $\hat{u}$ such that $\hat{u}_{|W} = u$. In this case, we write

$$T = S(v', u)\delta(t^1) \cdots \delta(t^p).$$

Theorem: (i) If W has codimension 1 ($p = 1$) and if the derivatives

$$\frac{\partial^{r_1 - r_2}}{(\partial t)^{r_1 - r_2}} \det \|a_{i'}^{j'}\| \,\big|_{t=0}, \quad 1 \le i', \quad j' \le r_1, \tag{1.19}$$

of the determinants of all $r_1 \times r_1$ submatrices of $\|a_i^j\|$ do not simultaneously vanish in W, then every nontrivial (i.e., not depending only on the v) solution T

of (1.9) with supp $T \subset W$ is of the form $S(v', u)\delta(t)$, i.e., has transversal order 0.

(ii) If the quantities (1.19) vanish simultaneously in W, there exist nontrivial solutions of (1.9) with support in W, which have transversal order 1.

(iii) If codim $W = p > 1$, there always exist solutions with nonzero transversal order.

Remark: Equation (1.19) is independent of the coordinate system. (All "tangential" or "mixed" derivatives of order $r_1 - r_2$ are zero, and all derivatives of order less than $r_1 - r_2$ are zero.)

Proof: (1) We have supp $\tilde{T}(t, v', \hat{u}, x') \subset W$ so there exist distributions $S_k \in \mathfrak{D}'(W)$ such that

$$\tilde{T}(t, v', \hat{u}, x') = \sum_k \delta^{(k)}(t) S_k,$$

$$\langle \tilde{T}, \varphi \rangle = \sum_k \left\langle S_k, \frac{\partial^k}{(\partial t)^k} \varphi_{|W} \right\rangle,$$

where the sum over k is finite in each bounded open set; for simplicity, assume U is bounded. The transversal order of $\tilde{T}$ in U is N if $S_N \neq 0$ and $k > N \Rightarrow S_k = 0$. Note that N is independent of the coordinate system.

We substitute this expression in Eq. (1.15) and we get

$$\left\{ b_i^j(t, v', \hat{u}, x') \frac{\partial}{\partial \hat{u}^j} + c_i^l(t, v', \hat{u}, x') \frac{\partial}{\partial x'^l} \right\} \tilde{T}(t, v', \hat{u}, x') = 0$$

with $1 \leq i \leq n$, $1 \leq j \leq r_1 - r_2$, $1 \leq l \leq r_2$, and $b_i^j(0, v', u, x') = 0$, and [the nonzero distribution of the form $S_k \delta^{(k)}(t)$ with different k being linearly independent] we obtain, from the coefficients of $\delta^{(N)}(t)$ and $\delta^{(N-1)}(t)$,

$$c_i^l(0, v', u, x') \frac{\partial}{\partial x'^l} S_N = 0, \tag{1.20a}$$

$$c_i^l(0, v', u, x') \frac{\partial}{\partial x'^l} S_{N-1} = N \frac{\partial}{\partial t} b_i^j(t, v', \hat{u}, x')\Big|_{t=0} \frac{\partial S_N}{\partial u^j}. \tag{1.20b}$$

(1.20a) is equivalent to $\partial S_N / \partial x^1 = 0$, so its solutions are of the form

$$\langle S_N, \varphi \rangle = \left\langle S_N(\psi', \psi), \int \varphi(\psi', \psi, x')\, dx' \right\rangle, \quad \varphi \in \mathfrak{D}(W),$$

with $N = 0$. Equation (1.20a) is the only condition, and so any distribution of the form $S(v', u)\delta(t)$ is a solution of (1.9).

If $N \geq 1$, Eq. (1.20b) must be solved. First, we replace it by an equivalent system of r_1 equations. The conditions on the rank of the system (1.20) can be expressed as follows: There exist functions $g_\alpha^i(t, v', \hat{u}, x')$, $1 \leq i \leq n$, $1 \leq \alpha \leq r_1$, such that, if

$$A_\alpha^\beta(t, v', \hat{u}, x') = g_\alpha^i(t, v', \hat{u}, x') c_i^\beta(t, v', \hat{u}, x'), \quad 1 \leq \beta \leq r_2,$$
$$= g_\alpha^i(t, v', \hat{u}, x') b_i^\beta(t, v, \hat{u}, x'), \quad r_2 + 1 \leq \beta \leq r_1,$$

then det $\|A_\alpha^\nu\| \neq 0$ if $t \neq 0$, and $\|A_\alpha^\beta\|$ is of rank r_2 if $t = 0$. The following equation is equivalent to the equation which occurs if we substitute $N - 1$ for N:

$$A_\alpha^\beta\big|_{t=0} \frac{\partial S_{N-1}}{\partial x^\beta} = N \frac{\partial}{\partial t} A_\alpha^\beta\big|_{t=0} \frac{\partial u_j}{\partial x^\beta} \frac{\partial S_N}{\partial u^j} \tag{1.21}$$

(where x is either u or x') for every suitable set of functions g_α^i.

In order to discuss (1.21), we use the following lemma:

Lemma: Let $\|A_\alpha^\beta\|$ a matrix of order r_1 and rank r_2, there exist $r_1 - r_2$ independent null linear combinations of its lines

$$C_{(k)}^\alpha A_\alpha^\beta = 0, \quad 1 \leq k \leq r_1 - r_2,$$

and $r_1 - r_2$ null linear combinations of its columns

$$A_\alpha^\beta D_\beta^{(j)} = 0, \quad 1 \leq j \leq r_1 - r_2.$$

Define

$$C^{\alpha_1 \cdots \alpha_{(r_1 - r_2)}} = [C_{(1)} \wedge \cdots \wedge C_{(r_1 - r_2)}]^{\alpha_1 \cdots \alpha_{(r_1 - r_2)}}$$
$$= \frac{1}{(r_1 - r_2)!} \sum_{\sigma \in S_{r_1 - r_2}} \chi(\sigma) C_{(\sigma(1))}^{\alpha_1} \cdots C_{(\sigma(r_1 - r_2))}^{\alpha_{r_1 - r_2}}$$

and

$$D_{\beta_1 \ldots \beta_{(r_1 - r_2)}} = [D^{(1)} \wedge \cdots \wedge D^{(r_1 - r_2)}]_{\beta_1 \ldots \beta_{(r_1 - r_2)}}.$$

On the other hand, denote by $\hat{A}_{\beta_1 \cdots \beta_l}^{\alpha_1 \cdots \alpha_l}$ the subdeterminant of order $(r_1 - 1)$, obtained by omission of the l lines $\alpha_1 \cdots \alpha_l$ and of the l columns $\beta_1 \cdots \beta_1$.

Lemma:

$$\hat{A}_{\beta_1 \ldots \beta_{(r_1 - r_2)}}^{\alpha_1 \ldots \alpha_{(r_1 - r_2)}} = M C^{\alpha_1 \ldots \alpha_{(r_1 - r_2)}} D_{\beta_1 \ldots \beta_{(r_1 - r_2)}},$$

where M is a constant independent of α and β. The proof is elementary.

Now, going back to Eq. (1.21), we note that the functions $\partial u^j / \partial x^\beta$ can be taken as the linear combinations $D_\beta^{(j)}(v'x)$ for the determinant $\|A_\alpha^\beta(0, v', x)\|$. If we consider the linear combinations of both sides of (1.21) defined by $C_{(k)}^\alpha(v', x)$, the left-hand side of (1.21) vanishes and we obtain

$$C_{(k)}^\alpha(v', x) \frac{\partial A_\alpha^\beta}{\partial t}\bigg|_{t=0} (v', x) D_\beta^{(j)}(v', x) \frac{\partial S_N}{\partial u^j} = 0.$$

This system has nontrivial solutions if and only if the determinant of the matrix

$$\left\| C^{\alpha}_{(k)}(v', x) \frac{\partial A^{\beta}_{\alpha}}{\partial t}\Big|_{t=0} (v', x) D^{(j)}_{\beta}(v', \alpha) \right\|$$

(free indices k, j; $1 \leq k, j \leq r_1 - r_2$) vanishes in W. But, from the Lemma, we can write this determinant as

$$\frac{\partial A^{\beta_1}_{\alpha_1}}{\partial t} \cdots \frac{\partial A^{\beta_{r_1-r_2}}_{\alpha_{r_1-r_2}}}{\partial t}\Big|_{t=0} C^{\alpha_1 \cdots \alpha_{r_1-r_2}} D_{\beta_1 \cdots \beta_{r_1-r_2}}$$
$$= \frac{\partial^{(r_1-r_2)}}{(\partial t)^{(r_1-r_2)}} (\det \|A^{\beta}_{\alpha}(t, v', x)\|)\big|_{t=0},$$

which proves (i) and (ii).

(2) If $p > 1$, we have

$$\tilde{T} = \sum_k S_k \delta^{(k_1)}(t^1) \cdots \delta^{(k_p)}(t^p), \quad k = \{k_1 \cdots k_p\}$$

(the transversal order of $\tilde{T}$ is N if there exists $S_k \neq 0$ with $|k| = N$ and $|k| > N$ implies $S_k = 0$; $|k| = k_1 + \cdots + k_p$). In this case, Eq. (1.20a) is replaced by

$$c^l_i(0, v', u, x') \frac{\partial S_k}{\partial x'^l}$$
$$= \sum_{s=1}^{p} (k_s + 1) \frac{\partial}{\partial t^s} b^j_i(t, v', \hat{u}, x')\big|_{t=0} \frac{\partial}{\partial u^j} S_{\{k_1 \cdots (k_s+1) \cdots k_p\}}, \quad |k| = N - 1,$$

and, even if the compatibility conditions imply

$$\sum_{s=1}^{p} (k_s + 1) \frac{\partial}{\partial u^j} S_{\{k_1 \cdots (k_s+1) \cdots k_p\}} = 0,$$

this is always possible with nonzero $\partial S_{\{k_1 \cdots (k_s+1) \cdots k_p\}}/\partial u^j$, so that there are at least solutions with $N = 1$.

Suppose that W is the manifold defined by the equations

$$\begin{cases} t^l(v) = 0, & 1 \leq l \leq p, \\ w(v, x) = 0 \end{cases}$$

(t^1 are characteristic manifolds as before), where w is a C^∞ function such that

$$a^j_i(z) \frac{\partial w}{\partial z^j}\Big|_W = 0.$$

Define again coordinates (t, v', w, x). Clearly, then, distributions of the form

$$\prod_{l=1}^{p} \delta(t^l) \epsilon(w) T(v') \tag{1.22}$$

are solutions of (1.9) [since $\epsilon(w)$ is the sign of w, it is clear that the form (1.22) is independent of the choice of a function w with the above properties]. In this case, we do not discuss the conditions necessary for some nonzero transversal order to appear.

More singular cases appear in the practical situations we meet (for example, the rank of the system may lower on a manifold W defined by the intersection of characteristic manifolds which are tangent together, or the manifold W may have a vertex), but, in those cases, either we solve (1.9) by convenient methods or we do not give a complete explicit solution, which is not necessary to solve the complete system (1.9) and (1.10).

As an example, we give here the solutions of (1.9) with support $\{e\}$, the identity on the group G (rank $|a^j_i(z)| = 0$ on $\{e\}$): From the isomorphism we mentioned in property (b) (Sec. 1B), T (supp $T = \{e\}$) is a central distribution; that is, $\delta_x * T = T * \delta_x \ \forall\, x \in \mathfrak{G}$ if and only if T belongs to the center of $\mathfrak{U}(\mathfrak{G})$ so that there exists some Q in the center of $\mathfrak{U}(\mathfrak{G})$, such that $T = \delta_Q = \delta_Q * \delta = D^t_Q \delta$, where D_Q is the Laplace operator defined by $D_Q f = \delta_Q * f, f \in \mathfrak{D}$.

2. CHARACTERS OF $SL(2, \mathbb{R})$

We choose $SL(2, \mathbb{R})$ for an introduction to our methods of calculation because it is a simple group with only three parameters and its structure and representations are well known; however, the fact that it is a real group already obliges us to solve the problem of "class functions" and "eigenfunctions" of Laplace operators in a space of distribution; lastly, $SL(2, \mathbb{R})$ being a little group of the Poincaré group, the calculation gives us some insight into the difficulties to come.

A. Parametrization of $SL(2, \mathbb{R})$

The group $SL(2, \mathbb{R})$ is the group of 2×2 real matrices of determinant 1:

$$g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \text{with} \quad \alpha\delta - \beta\gamma = 1.$$

It has a center $C(G)$, the two matrices $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, and, in addition, singular elements, the other matrices of trace ± 2, which cannot be diagonalized.

We introduce the new variables

$$\alpha \pm \delta = u, v,$$
$$\beta \pm \gamma = r, s.$$

In the space $R^4(u, v, r, s)$, the group manifold is then the quadric

$$u^2 - v^2 - r^2 + s^2 = 4. \tag{2.1}$$

The singular matrices are represented by the submanifolds $u = \pm 2$, which are cones in R_3. The elements of $C(G)$, corresponding to the values $u = \pm 2$ and $v = r = s = 0$ of the parameters, are the vertices of these cones.

The Haar measure can be written

$dv\,dr\,ds/|u|$, for $u \neq 0$, $du\,dv\,dr/|s|$, for $s \neq 0$.

We write $M^{01}M^{02}M^{12}$, the generators of the Lie algebra of $SL(2, R)$, with the commutation relations

$$[M^{01}, M^{02}] = M^{12}, \quad [M^{02}, M^{12}] = -M^{01}, \quad [M^{12}, M^{01}] = -M^{02}.$$

M^{12} is then the generator of the compact subgroup U_1. The differential expression of these generators acting on functions of the u, v, r, and s variables bounded by (2.1),by right- or left-infinitesimal translation in G, is:

$$\begin{aligned} 2M^{01}_{R,L} &= \pm u\frac{\partial}{\partial r} \pm r\frac{\partial}{\partial u} + v\frac{\partial}{\partial s} + s\frac{\partial}{\partial v},\\ 2M^{02}_{R,L} &= \pm u\frac{\partial}{\partial v} \pm v\frac{\partial}{\partial u} - s\frac{\partial}{\partial r} - r\frac{\partial}{\partial s},\\ 2M^{12}_{R,L} &= \mp u\frac{\partial}{\partial s} \pm s\frac{\partial}{\partial u} + r\frac{\partial}{\partial v} - v\frac{\partial}{\partial r}. \end{aligned} \tag{2.2}$$

In fact, we always consider the system (2.2) "restricted" to some open sets where the relation (2.1) will then just be expressed by the elimination of the "bad" variable.

The Casimir operator of $SL(2, R)$, which is given in the enveloping algebra by

$$Q = (M^{01})^2 + (M^{02})^2 - (M^{12})^2, \tag{2.3}$$

defines in the same way a Laplace operator. As we said, we are only interested in its restriction ${}^tD_Q$ to the central distributions we seek first.

B. Central Distributions

From (2.2) our equations (1.9) for central distribution are in $SL(2, R)$:

$$\begin{aligned} &\left(v\frac{\partial}{\partial s} + s\frac{\partial}{\partial v}\right)\tilde{T}_c = 0,\\ &\left(s\frac{\partial}{\partial r} + r\frac{\partial}{\partial s}\right)\tilde{T}_c = 0,\\ &\left(r\frac{\partial}{\partial v} - v\frac{\partial}{\partial r}\right)\tilde{T}_c = 0. \end{aligned} \tag{2.4}$$

The rank of the system is constant at 2 outside $C(G)$. The characteristic manifolds are

$$u^2 \equiv 4 + v^2 + r^2 - s^2 = a^2.$$

They are 1-sheeted hyperboloids in R^3 for $a^2 > 4$, 2-sheeted hyperboloids for $a^2 < 4$; we saw already that $u^2 = 4$ was a cone.

From our previous discussions we thus see that the solutions of (2.1) are

$$T_c = T_1(u) + \epsilon(s)\theta(4 - u^2)T_2(u), \quad \text{outside } C(G),$$

and

$$T_c = \sum_N (a_N\theta(u) + b_N\theta(-u))\Theta_N, \quad \text{(finite sum) if } \operatorname{supp} T_c \subset C(G), \tag{2.5}$$

where $\Theta_N = (\delta_Q)^{*N} * \delta$; explicitly, we have

$$\Theta_N = \sum_{p,q\geq 0}^{p+q\leq N} (-)^{p+q} C_N^{p+q} C_{p+q}^{p} \delta^{(2p)}(v)\delta^{2q}(r)\delta^{(2N-2p-2q)}(s).$$

[$T_1(u)$ and $T_2(u)$ are defined for $u^2 - 4 \neq 0$.]

C. Eigendistributions

We are now going to look, among distributions T_c, for the solutions of (1.10):

$${}^tD_Q T_c = -qT_c. \tag{2.6}$$

We have seen that T_c has a very simple form when $u^2 - 4 \neq 0$. In any such open set, the explicit form ${}^tD_Q$ also becomes simple; with (2.2) and (2.3), Eq. (2.6) becomes

$$\left((u^2 - 4)\frac{\partial^2}{\partial u^2} + 3u\frac{\partial}{\partial u} + 4q\right)\tilde{T}_c = 0. \tag{2.7}$$

We shall first solve (2.7) in a open set U complementary of e and such that $u > -2$.

The functional solutions of such a differential equation in u are well known in the analytic field (Re $u > -2$): (2.7) is a Fuchsian equation of first type with a singular-regular point at $u = 2$; it admits one regular holomorphic solution $R(u)$ and one singular solution $S(u) = A(u)/(u - 2)^{\frac{1}{2}}$, with $A(u)$ holomorphic.

When the equation is restricted to the half real axis $u > -2$, there is no longer a connection between the two sides of the point $u = 2$, and we have, as four independent solutions of (2.7), the functions

$$\begin{aligned} f_1(u) &= 0, \quad u < 2, & f_3(u) &= 0, \quad u < 2,\\ &= R(u) \quad u > 2, & &= S(u), \quad u > 2,\\ f_2(u) &= R(u), \quad u < 2, & f_4(u) &= S(u), \quad u < 2,\\ &= 0, \quad u > 2, & &= 0, \quad u > 2. \end{aligned}$$

Noticing that

$${}^tD_Q\theta(\pm s)\theta(4 - u^2)T(u) = \theta(\pm s)\,{}^tD_Q\theta(4 - u^2)T(u),$$

we can conclude that the central distributions solutions of (2.7) in U with a support different from the only manifold $u = 2$ are combinations of

$$R_e(u) = \theta(u - 2)R(u), \quad R_i^{\pm}(u) = \theta(\pm s)\theta(2 - u)R(u),$$
$$S_e(u) = \theta(u - 2)S(u), \quad S_i^{\pm}(u) = \theta(\pm s)\theta(2 - u)S(u).$$

$S(u)$ and, of course, $R(u)$ are locally summable functions on G, and the products θR and θS actually define distributions.

To calculate more easily the action of ${}^tD_Q$ on these distributions (particularly with the usual formulas of derivation of a product), it is interesting to avoid coincidence of singularities by replacing $\theta(u-2)$ and $\theta(2-u)$ by

$$\lim_{\eta\to+0}\theta(u-2-\eta),\quad \lim_{\eta\to+0}\theta(2-\eta-u).$$

(We know that we can permute limit and derivation.) It appears clearly that the quantity η so introduced has the effect of restricting our distributions to a complement of a neighborhood of identity. The limit $\eta\to 0$ will thus give us the action of ${}^tD_Q$ in all $u > -2$.

Using the form given for $S(u)$, from which we get

$$S(2\pm\eta)=\frac{A(2)}{\sqrt{\eta}}\quad S'(2\pm\eta)=\frac{1}{2}\frac{A'(2)}{\sqrt{\eta}}\mp\frac{1}{2}\frac{A(2)}{\eta\sqrt{\eta}}$$

[at $O(\sqrt{\eta})$ and apart a multiplicative factor of no importance], and the fact that $R(u)$ and $S(u)$ are solutions of (2.7), we obtain

$$\begin{aligned}4\,{}^tD_QR_e(u)&=\lim_{\eta\to+0}\{[2\delta(u-2-\eta)+4\eta\delta'(u-2-\eta)]R(2)\\&\quad+4\eta\delta(u-2-\eta)R'(2)\},\end{aligned}$$

$$\begin{aligned}4\,{}^tD_QR_i^\pm(u)&=\lim_{\eta\to+0}\theta(\pm s)\{[-2\delta(u-2+\eta)\\&\quad+4\eta\delta'(u-2+\eta)]R(2)\\&\quad+4\eta\delta(u-2+\eta)R'(2)\},\end{aligned}$$

$$\begin{aligned}4\,{}^tD_QS_e(u)&=\lim_{\eta\to+0}[4(\sqrt{\eta})\delta'(u-2-\eta)A(2)\\&\quad+2(\sqrt{\eta})\delta(u-2-\eta)A'(2)],\end{aligned}$$

$$\begin{aligned}4\,{}^tD_QS_i^\pm(u)&=\lim_{\eta\to+0}\theta(\pm s)[4(\sqrt{\eta})\delta'(u-2+\eta)A(2)\\&\quad+2(\sqrt{\eta})\delta(u-2+\eta)A'(2)].\end{aligned}$$

We have calculated in Appendix A the limits of the δ and δ' distributions with the result

$$\lim_{\eta\to+0}\delta(u-2-\eta)=\delta(u-2),$$
$$\lim_{\eta\to+0}\eta^\alpha\delta'(u-2-\eta)=0,\quad\forall\,\alpha>0,$$
$$\lim_{\eta\to+0}\theta(\pm s)\delta(u-2+\eta)=\theta(\pm s)\delta(u-2),$$
$$\lim_{\eta\to+0}(\sqrt{\eta})\theta(\pm s)\delta'(u-2+\eta)=-2\pi\delta(v)\delta(r)\delta(s),$$

and so

$$\begin{aligned}&4\,{}^tD_QR_e(u)=2\delta(u-2)R(2),\\&4\,{}^tD_QR_i^\pm(u)=-2\theta(\pm s)\delta(u-2)R(2),\\&4\,{}^tD_QS_e(u)=0,\\&4\,{}^tD_QS_i^\pm(u)=-8\pi A(2)\delta(v)\delta(r)\delta(s).\end{aligned}$$

On the other hand, we can easily see that

$${}^tD_Q\delta^{(k)}(u-2)=c_1\delta^{(k+1)}(u-2)+c_2\delta^{(k)}(u-2),\quad c_1\neq 0$$

[the distributions $\delta^{(K)}(u-2)$ have a meaning outside $C(G)$], from which we can see that there does not exist any central distribution T of support $u=2$ such that, separately or combined with the results above, we get ${}^tD_Q(T)=0$.

Lastly, the action of ${}^tD_Q$ on Θ_N giving Θ_{N+1} and the sum over N being finite, we can conclude that the eigendistributions of (2.6) in $u>-2$ are

$$\begin{gathered}R_e(u)+R_i^+(u)+R_i^-(u)=R(u),\\S_e(u),\\S_i^+(u)-S_i^-(u)=S_i(u,s/|s|).\end{gathered}\tag{2.8}$$

A similar argument leads us, in $u<2$, to the eigendistributions

$$\begin{gathered}\hat{R}(u),\\\hat{S}_e(u),\\\hat{S}_i^+(u)-\hat{S}_i^-(u)=\hat{S}_i(u,s/|s|)\end{gathered}\tag{2.9}$$

[$\hat{R}(u)$ and $\hat{S}(u)$ are the regular and singular solutions of (2.7) in U'].

We postpone to the next paragraph the patching of the restrictions we have found in $u>-2$ and $u<2$. This problem will lead us to the discussion of the spectrum of Q.

D. Spectrum of Laplace Operator; Characters of $SL(2,R)$

1. *Explicit Expressions of R, $\hat{R}$, S, and $\hat{S}$*

The differential equation

$$(u^2-4)f''+3uf'+4qf=0$$

can be solved explicitly with a change of variable suited to each region and next with a change of function; namely,

$$|u|>2,\quad |u|=2\cosh\eta,\quad 0<\eta<+\infty,\quad g=f/\sinh\eta,$$
$$|u|<2,\quad u=2\cos\Phi,\quad 0<\Phi<\pi,\quad g=f/\sin\Phi.$$

The equation then becomes

$$g''\pm(4q-1)g=0,$$

and the solutions are

$$R(u) = \begin{cases} \sinh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta, \\ \sin (1 - 4q)^{\frac{1}{2}}\Phi/\sin \Phi, \end{cases}$$

$$\hat{R}(u) = \begin{cases} \sinh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta, \\ \sin (1 - 4q)^{\frac{1}{2}}(\pi - \Phi)/\sin (\pi - \Phi), \end{cases}$$

$$S(u) = \begin{cases} \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta, \\ \cos (1 - 4q)^{\frac{1}{2}}\Phi/\sin \Phi, \end{cases}$$

$$\hat{S}(u) = \begin{cases} \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta, \\ \cos (1 - 4q)^{\frac{1}{2}}(\pi - \Phi)/\sin (\pi - \Phi). \end{cases}$$

(These functions being defined up to a factor, the determination chosen for the root $(1 - 4q)^{\frac{1}{2}}$ has no importance: R and $\hat{R}$ are well continuous on the points $u = \pm 2$.)

2. Patching

We remember that we are only interested in distributions on G satisfying the "parity" requirement

$$\tilde{T}(-u, -s) = \pm \tilde{T}(u, s).$$

On the other hand, the distributions of formulas (2.8) and (2.9) which we want to patch together have to be equal in the intersection J: $-2 < u < 2$. We then see that only the homologous lines of (2.8) and (2.9) can fulfill these conditions.

For the first and third lines, we have the supplementary condition that $(1 - 4q)^{\frac{1}{2}}$ must be a real number, integer, or zero. This can be written

$$q = K(1 - K), \quad K = \tfrac{1}{2}, 1, \tfrac{3}{2} \cdots,$$

for then

$$(1 - 4q)^{\frac{1}{2}} = 2K - 1$$

and we have the two global eigendistributions of ${}^tD_Q$:

$$R^K(u) = \begin{cases} \sinh (2K - 1)\eta/\sinh \eta, & u > 2, \\ \sin (2K - 1)\Phi/\sin \Phi, & |u| < 2, \\ (-1)^{2K} \sinh (2K - 1)\eta/\sinh \eta, & u < -2, \end{cases}$$

$$S_i^K(u, s/|s|) = \theta(4 - u^2)[\cos (2K - 1)\Phi/\sin \Phi][\theta(s) - \theta(-s)].$$

[The parity of these two solutions is $(-1)^{2K}$.] We can notice here that the variable $\cos \Phi = \frac{1}{2}u$ we have introduced is nothing but the cosine of the angle of rotation corresponding to the element of U_1 group to which $g(v, r, s)$ is conjugate. This leads us to let Φ vary between 0 and 2π with the relation $\sin \Phi/|\sin \Phi| = s/|s|$. Φ is then actually the angle of the rotation, and we can write more simply

$$S_i^K(u, s/|s|) = S_i^K(\Phi) = \cos (2K - 1)\Phi/\sin \Phi, \quad |u| < 2.$$

The second lines of (2.8) and (2.9) do not give us much trouble. $S_e(u)$ and $\hat{S}_e(u)$ both being zero in J, we have for any q the two global eigendistributions of Q of opposite parity:

$$S_e^{q+}(u) = \theta(-u - 2) \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta + \theta(u - 2) \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta,$$

$$S_e^{q-}(u) = -\theta(-u - 2) \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta + \theta(u - 2) \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta.$$

3. The Characters of SL(2, R)

With the central eigendistributions we have obtained for each value of q, we are now going to form the linear combinations that can be characters, and compare them to the known representations associated with the same q.[7,8]

1. $q \neq K(1 - K)$. We have only two eigendistributions S_e^{q+} and S_e^{q-} which we say are the characters of two nonequivalent irreducible representations. We write

$$\chi_q^{0,\frac{1}{2}}(u) = \theta(u - 2) \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta \pm \theta(-u - 2) \cosh (1 - 4q)^{\frac{1}{2}}\eta/\sinh \eta.$$

For complex or real negative q, the corresponding representations are surely not unitary, the distributions being unbounded. They are well known.

For real $q > \frac{1}{4}$, χ_q^0 and $\chi_q^{\frac{1}{2}}$ decrease [as $\cos (a\eta)e^{-\eta}$] and can be the characters of unitary representations. There are effectively two known representations: the so-called principal series C^0 and $C^{\frac{1}{2}}$.

For $0 < q < \frac{1}{4}$, $\chi_q^{\frac{1}{2}}$ still decreases, but only as $\exp [-1 + (1 - 4q)^{\frac{1}{2}}]$. Only χ_q^0 corresponds to unitary representations: those of the "complementary series" (which do not appear in the regular representation of the group). We are not able to show that $\chi_q^{\frac{1}{2}}$ is not of positive type.

2. $q = K(1 - K)$. The space of central eigendistributions is 4-dimensional, generated by S_e^{K+}, S_e^{K-}, S_i^K, and R^K. Only one of these distributions has parity $(-1)^{2K+1}$, and we can isolate it as a character

$$\chi_K^{(-)^{2K+1}}(u) = \theta(u - 2) \cosh (2K - 1)\eta/\sinh \eta + (-1)^{2K+1}\theta(-u - 2) \cosh (2K - 1)\eta/\sinh \eta.$$

For $K > 1$, this distribution, unbounded, is the character of nonunitary representations. For $K = \frac{1}{2}$, it is the character of the unitary representation $C_{\frac{1}{4}}^0$ of the complementary series. For $K = 1$, the corresponding representation is not unitary, although the character is bounded.

Among the three left distributions, $R(u)$ is continuous and is the character of the finite-dimensional representations

$$\chi_K^f(u) = \begin{cases} \sinh(2K-1)\eta/\sinh\eta, & u > 2, \\ \sin(2K-1)\Phi/\sin\Phi, & |u| < 2, \\ (-1)^{2K}\sinh(2K-1)\eta/\sinh\eta, & u < -2. \end{cases}$$

The dimension of these representations is $2K-1$ (and χ is well normalized by $\chi_f^K(e) = 2K-1$). They are, of course, nonunitary, except for $K=1$, which gives the trivial representation. $K = \frac{3}{2}$ corresponds to the representation of the group by itself: We verify that $\chi^f_{\frac{3}{2}}(u) = u!$.

We are left with a 2-dimensional space of solutions. We can obtain two combinations giving bounded distributions

$$\tfrac{1}{2}[(S_e^{K(-)^{2K}} - R^K) + \alpha_{1,2}S_i^K].$$

We would like to find the values of $\alpha_{1,2}$ so that we have two extremal distributions of positive type, which would be the characters of the unitary representations of the two discrete series D_K^+ and D_K^-.

The condition $\langle\chi,\varphi\rangle = \langle\chi,\tilde{\varphi}\rangle$ gives us $\alpha_{1,2}$ pure imaginary. The distributions $S^{K(-)^{2K}} - R^K \pm \alpha S_i$ being conjugated, the condition $\chi \gg 0$ is of the type $|\alpha| \le a(K)$, and the extremality gives

$$\chi_K^{\pm}(u,s) = \tfrac{1}{2}[S_e^{K(-)^{2K}}(u) - R^K(u) \pm ia(K)S_i^K(u, s/|s|)].$$

Up to now, we have not been able to calculate $a(K)$, and we can only refer to the result obtained (by very different methods) by Gel'fand,[9] $a(K) = 1$, $\forall\, K$.

We then have

$$\chi_K^{\pm}(u, s/|s|) = \begin{cases} e^{-(2K-1)\eta}/\sinh\eta, & u > 2, \\ \pm ie^{\pm i(2K-1)\Phi}/\sin\Phi, & |u| < 2, \\ (-1)^{2K}e^{-(2K-1)\eta}/\sinh\eta, & u < -2. \end{cases}$$

4. *Frobenius' Formula*

We conclude this first part by noticing a fact which appears as a generalization of the Frobenius formula for finite groups[10]; this formula relates the character of a representation of a finite group G, induced by a representation of a subgroup H, to the character of this representation: We have

$$\chi_G(g) = \sum_r \chi_H(g_{r_0}^{-1} g g_r)\delta_{r,r_0},$$

where the g_r are some representatives of the cosets G/H and where g_r defined by $g_r^{-1}g$, $g_r \in H$. In particular, $\chi_G(g)$ is equal to zero if the element g is not the conjugate of an element of H.

Now all irreducible representations of $SL(2, R)$ are either induced representations or extracted from induced representations; the inducing subgroup is the solvable group H of elements

$$h = \begin{pmatrix} \lambda & x \\ 0 & \lambda^{-1} \end{pmatrix},$$

λ, x real, and the inducing representations the 1-dimensional irreducible representations of H defined by $(\operatorname{sgn}\lambda)^\epsilon\,|\lambda|^{i(s+1)}$, $\epsilon = 0, 1$, $s \in C$. If we now write $s = \pm(1-4q)^{\frac{1}{2}}$, we have: For $q \neq K(1-K)$, $2K$ positive integer, the representations induced by $(\epsilon, \pm s)$ are irreducible and equivalent [according to ϵ we get the two representations corresponding to the eigenvalue q of Laplace operator of $SL(2, R)$]; for $q = K(1-K)$ $[s = \pm(2K-1)]$ if, $\epsilon = (-1)^{2K-1}$, the two induced representations are again irreducible and equivalent (we have found their characters); for $s = +(2K-1)$, $\epsilon = (-1)^{2K}$, the induced representation is reducible with, as subrepresentation, the finite-dimensional spinor representation; in the quotient, space acts the reducible representation $D_K^+ \oplus D_K^-$; for $s = -(2K-1)$, $\epsilon = (-1)^{2K}$, the induced representation is again reducible, but this time, the subrepresentation is $D_K^+ \oplus D_K^-$ and the quotient representation is the spinor representation.

Now on the group manifold, the subgroup H is defined by $r = s$ and $|u| \ge 2$ and by the matrices of $SL(2, R)$ for which $|u| < 2$ are surely not the conjugates of elements of H. If we look at the distributions

$$\chi_q^0, \quad \chi_q^{\frac{1}{2}}, \quad \chi_K^{(-)2K+1}$$

and

$$\chi_K^f + \chi_K^+ + \chi_K^-$$

(trace of the reducible induced representations), we can see that they have for support

$$|u| \ge 2.$$

Such a phenomenon will be still more striking for the characters of the representations of the Poincaré group, which are also induced representations when the translations are not trivially represented.

3. CENTRAL DISTRIBUTIONS ON THE POINCARÉ GROUP

The calculation of the central distributions on $SL(2, R)$ has been done very easily, because the structure of the set of conjugation classes of the group was simple.

For the Poincaré group, the differential method we introduce proves all its interest: One can, of course, get the invariants under conjugation in $\bar{P}$ directly by algebraic calculation or geometrical considerations and thus know the support of the distributions we are looking for.[11,12] In addition, we obtain the transversal order of these distributions.

As we said in Sec. 1D, we only solve Eqs. (1.9) "locally."

A. Parametrization of the Poincaré Group

The Poincaré group P is the semidirect product of the proper Lorentz group $\mathfrak{L}_0$ [connected component of the invariance group of the quadratic form $\mathbf{x}^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2$ on R^4] by the 4-dimensional translation group R^4. In fact, we are interested in the covering $\bar{P}$ of P:

$$\bar{P} = R^4 \times \bar{\mathfrak{L}}_0, \quad \text{with} \quad \bar{\mathfrak{L}}_0 = SL(2, \mathbb{C}).$$

An element of this group can be represented by a pair (x, Λ), where x is the Hermitian matrix

$$\mathsf{x} = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix},$$

associated with the real 4-vector x^μ ($\det \mathsf{x} = \mathbf{x}^2$):

$$\Lambda \in SL(2, \mathbb{C}), \text{ i.e., } \Lambda = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad \alpha, \beta, \gamma, \delta \in \mathbb{C}, \quad \alpha\delta - \beta\gamma = 1.$$

The group law is then

$$(\mathsf{x}_1, \Lambda_1)(\mathsf{x}_2, \Lambda_2) = (\mathsf{x}_1 + \Lambda_1 \mathsf{x}_2 \Lambda_1^*, \Lambda_1\Lambda_2),$$

Λ^* = Hermitian conjugate of Λ.

We define the six matrices $\sigma_{\alpha\beta}$, $\alpha, \beta = 0, 1, 2, 3$, $\sigma_{\alpha\beta} = -\sigma_{\beta\alpha}$, by

$$\sigma_{0i} = \tfrac{1}{2}\sigma_i, \quad \sigma_{ij} = -\tfrac{1}{2}\epsilon_{ijk}\sigma_k, \quad i, j, k = 1, 2, 3,$$

where ϵ_{ijk} is the completely antisymmetric tensor of order 3, $\epsilon_{123} = +1$, and the σ_i are the Pauli matrices

$$\sigma_1 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \quad \sigma_2 = \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix}, \quad \sigma_3 = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}.$$

The commutation relations of the $\sigma_{\alpha\beta}$ are those of the generators of the Lie algebra of $SL(2, \mathbb{C})$:

$$[\sigma_{\alpha\beta}\, \sigma_{\gamma\delta}] = -g_{\alpha\gamma}\sigma_{\beta\delta} + g_{\beta\gamma}\sigma_{\alpha\delta} - g_{\beta\delta}\sigma_{\alpha\gamma} + g_{\alpha\delta}\sigma_{\beta\gamma},$$
$$g_{\alpha\beta} = 0, \quad \text{if } \alpha \neq \beta, \quad g_{00} = 1, \quad g_{ii} = -1.$$

We then write (sum over repeated indices)

$$\Lambda = \tfrac{1}{2}[u\mathbf{1} + z^{\alpha\beta}\sigma_{\alpha\beta}], \quad \Lambda \in SL(2, \mathbb{C}),$$
$$z^{\alpha\beta} = -z^{\beta\alpha} \in \mathbb{R}.$$

(The $z^{\alpha\beta}$ transform under Lorentz group as the components of a tensor of order 2.) By identification, we get

$$u = u_1 + iu_2 = \operatorname{Re}(\alpha + \delta) + i \operatorname{Im}(\alpha + \delta),$$
$$z^{01} = -\operatorname{Re}(\beta + \gamma), \quad z^{23} = -\operatorname{Im}(\beta + \gamma),$$
$$z^{02} = \operatorname{Im}(\beta - \gamma), \quad z^{31} = -\operatorname{Re}(\beta - \gamma),$$
$$z^{03} = -\operatorname{Re}(\alpha - \delta), \quad z^{12} = -\operatorname{Im}(\alpha - \delta).$$

Let $\hat{z}_{\alpha\beta} = \frac{1}{2}\epsilon_{\alpha\beta\gamma\delta}z^{\gamma\delta}$ be the dual tensor ($\epsilon_{\alpha\beta\gamma\delta}$ is the completely antisymmetric tensor of order 4, $\epsilon_{0123} = -1$). With these variables, the group $\bar{\mathfrak{L}}_0$ is a manifold in R^8 defined by the equations (which come from $\det \Lambda = 1$)

$$4u_1u_2 + z^{\alpha\beta}\hat{z}_{\alpha\beta} = 0,$$
$$2(u_1^2 - u_2^2 - 4) + z^{\alpha\beta}z_{\alpha\beta} = 0. \tag{3.1}$$

The center $C(\bar{\mathfrak{L}}_0)$ of $\bar{\mathfrak{L}}_0$ has two elements

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix}$$

which are the vertices of cones containing the nondiagonalizable matrices with trace ± 2 ($u_1 = \pm 2$, $u_2 = 0$). The center of the Poincaré group is

$$C(\bar{P}) = \left\{\left(\mathbf{0}, \pm \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}\right)\right\}.$$

We note that the $z^{\alpha\beta}$ can be chosen as coordinates only in the open sets $U_1(u_1 > 0)$ and $U_2(u_1 < 0)$. (We do not exhibit explicitly a complete system of charts on $\bar{\mathfrak{L}}_0$, but it can be done in a straightforward manner.) Coordinates on $\bar{P}$ are given (outside $u_1 = 0$) by (x^μ, $z^{\alpha\beta}$).

B. Equations (1.9)

$\bar{P}$ is a 10-parameter group, so that the system (1.9) contains ten equations: four of them (3.2) express invariance under conjugation by R^4, the other six (3.3) invariance under conjugation by $\bar{\mathfrak{L}}_0$. An explicit calculation gives [in the form (1.15)]

$$\mathfrak{T}_\alpha \tilde{T} \equiv (u_2^2 g_{\alpha\beta} - z_{\alpha\gamma}z^\gamma_\beta - u_1 z_{\alpha\beta} - u_2\hat{z}_{\alpha\beta})\frac{\partial}{\partial x_\beta}\tilde{T} = 0, \tag{3.2}$$

$$\mathfrak{M}_{\lambda\mu}\tilde{T} \equiv \left(z^\beta{}_\lambda \frac{\partial}{\partial z^{\beta\mu}} - z^{\alpha\mu}\frac{\partial}{\partial z^{\alpha\lambda}} + x_\lambda \frac{\partial}{\partial x^\mu} - x_\mu \frac{\partial}{\partial x^\lambda}\right)\tilde{T} = 0 \tag{3.3}$$

[where the "bad" variables—for instance, u_1 and u_2 outside $\{u_1 = 0\}$—are to be expressed in terms of the other six by means of (3.1) and $\tilde{T}$ is defined in the corresponding chart, as in (1.15)].

The peculiar form of (3.2), in the coefficients of which the variables x^α do not appear, and of (3.3), where the x^α and $z^{\mu\nu}$ variables are separated, leads us to use first a fractional approach which is of greater interest for geometrical interpretation.

C. Invariance under Conjugation by Translation

We first study Eqs. (3.2).

Outside $\{u_1 = 0\}$, we write (with obvious notations) (3.2) in the form

$$\Omega_{\alpha\beta}(z)\frac{\partial}{\partial x_\beta}\tilde{T}(x, z) = 0,$$

where we have $\det \|\Omega_{\alpha\beta}(z)\| = 4u_2^2$, and so, outside $\{u_2 = 0\}$, solutions are distributions which do not depend on x:

$$T = T(z), \quad u_2 \neq 0.$$

Then one can see that, on the manifold $\{u_2 = 0\}$, the rank r of $\|\Omega_{\alpha\beta}\|$ is 2, except on $R^4 \times C(\bar{\mathfrak{L}}_0)$ $(z_{\alpha\beta} = 0)$, where $r = 0$. The differential operators of (3.2) restricted to $\{u_2 = 0\}$ are

$$(z_{\alpha\gamma}z^\gamma_\beta + u_1 z_{\alpha\beta})\frac{\partial}{\partial x_\beta}, \tag{3.4}$$

where the variables are now related by

$$\hat{z}^{\alpha\beta}z_{\alpha\beta} = 0, \tag{3.5}$$

$$2(u_1^2 - 4) + z^{\alpha\beta}z_{\alpha\beta} = 0. \tag{3.6}$$

The characteristic manifolds of (3.1) are [outside $R^4 \times C(\bar{\mathfrak{L}}_0)$] the four functions $\xi^\sigma = \hat{z}^{\sigma\tau}x_\tau$, among which only two, of course, are linearly independent [$\det \|\hat{z}^{\sigma\tau}\| \simeq u_1^2u_2^2$ is of rank two on $\{u_2 = 0\}$ outside $R^4 \times C(\bar{\mathfrak{L}}_0)$].

Now, as $\partial^2(\det \|\Omega_{\alpha\beta}\|)/(\partial u)^2 = 0$, we know from the theorem in Sec. 1F that the solutions with support in $\{u_2 = 0\}$ are of the form

$$\delta(u_2)T(z, \xi), \quad \text{outside} \quad R^4 \times C(\bar{\mathfrak{L}}_0).$$

Distributions of the form $\prod_{\alpha\beta}\delta(z_{\alpha\beta})T(x)$ are obviously solutions with support in $R^4 \times C(\bar{\mathfrak{L}}_0)$. We do not discuss here the form of solutions with nonzero transversal order and with such a support.

Geometrically[11,12] the condition $u_2 = 0$ means that the matrix Λ of the corresponding element in $\bar{P}$ is a conjugate either of a rotation ($|u_1| < 2$), or of a pure Lorentz transformation ($|u_1| > 2$), or of a transformation of the Euclidean group $E(2)$ ($|u_1| = 2$). Indeed, the matrix $\Omega_{\alpha\beta}$ is proportional to the matrix $(1 - \tilde{\Lambda}^t)_{\alpha\beta}$, where $\tilde{\Lambda}^t$ is the transpose of the 4×4 matrix $\tilde{\Lambda} \in \mathfrak{L}_0$ defined by $\Lambda \in \bar{\mathfrak{L}}_0$.

If $u_2 = 0$, there exists a 2-plane, pointwise invariant by $\tilde{\Lambda}$, generated by the vectors $\hat{z}^{\mu\nu}y_\nu$ $(y \in R^4)$ and of equation $z^{\alpha\beta}x_\beta = 0$. This 2-plane is spacelike, or has a timelike direction, or is tangent to the light cone, according as $|u_1| < 2$, or $|u_1| > 2$, or $|u_1| = 2$, respectively.

D. Invariance under Conjugation by Lorentz Group

We solve now Eqs. (3.3) separately:

$$\left(\mathfrak{M}^{\rho\sigma}_{\lambda\mu}(z)\frac{\partial}{\partial z^{\rho\sigma}} + \mathfrak{M}^\nu_{\lambda\mu}(x)\frac{\partial}{\partial x^\nu}\right)\tilde{T}(x, z) = 0. \tag{3.7}$$

We first look at the system restricted to distributions which depend only on z. It reduces then to the system of class-equations of $SL(2, \mathbb{C})$:

$$\mathfrak{M}^{\rho\sigma}_{\lambda\mu}(z)\frac{\partial}{\partial z^{\rho\sigma}}\tilde{T}(z) = 0. \tag{3.8}$$

The rank of the system (3.8) is 4 outside $C(\bar{\mathfrak{L}}_0)$; one can see that there exist two nontrivial combinations of the equations

$$z^{\lambda\mu}\mathfrak{M}^{\rho\sigma}_{\lambda\mu} = 0, \quad \hat{z}^{\lambda\mu}\mathfrak{M}^{\rho\sigma}_{\lambda\mu} = 0,$$

which are independent unless $z_{\lambda\mu} = 0, \forall\, \lambda, \mu$. (We give in Appendix B a list of relations between $z, \hat{z}, \cdots$, which are useful in all calculations.) The two corresponding characteristic manifolds are $z^{\alpha\beta}z_{\alpha\beta}$ and $z^{\alpha\beta}\hat{z}_{\alpha\beta}$, i.e., $u_1^2 - u_2^2$ and $u_1 u_2$; the patching between the different open sets of a covering of $SL(2, \mathbb{C})$ shows that, in fact, u_1 and u_2 are themselves characteristic manifolds. So the solutions of (3.8) on $\bar{\mathfrak{L}}_0$ are distributions of the form $T(u_1, u_2)$ outside $C(\bar{\mathfrak{L}}_0)$. On the other hand, solutions of (3.8) with support in $C(\bar{\mathfrak{L}}_0)$ are of the form (see end Sec. I)

$$T_{\{a\},\{b\}} = \sum_{N_1N_2} [a_{N_1N_2}\theta(u_1) + b_{N_1N_2}\theta(-u_1)]A^{N_1}B^{N_2} \times \left(\prod_{\alpha<\beta}\delta(z^{\alpha\beta})\right),$$

(finite sum) where A and B are the Laplace operators of $SL(2, \mathbb{C})$. The solutions of (3.7) which depend only on z are thus of the form $T(u_1u_2)$ [outside $R^4 \times C(\bar{\mathfrak{L}}_0)$] and $T_{\{a\},\{b\}}$.

Acting on distributions depending only on $\mathbf{x}$ we reduce (3.7) to the invariance equations of the Minkowsky space under $SL(2, \mathbb{C})$,

$$\mathfrak{M}^\nu_{\lambda\mu}(x)\frac{\partial}{\partial x^\lambda}\tilde{T}(x) = 0. \tag{3.9}$$

Outside $\{x^\mu = 0\}$, there are only three independent equations with the characteristic manifold $\mathbf{x}^2 = x_\mu x^\mu$. So the solutions of (3.3) which depend only on x are, outside $\{x_\mu = 0\}$, of the form

$$T(\mathbf{x}^2), \ \theta(\mathbf{x}^2)\epsilon(x^0)T(\mathbf{x}^2), \ \text{and} \ \delta^{(k)}(\mathbf{x}^2)\epsilon(x^0).$$

[The formal notation $\theta(\mathbf{x}^2)\epsilon(x^0)T(\mathbf{x}^2)$ does not mean that the product makes sense: It means "any odd, invariant distribution with support in the light cone." The use of such formal notation does not spoil the validity of the subsequent calculations if one handles it carefully. Analogous remarks are to be understood when we use similar notations in the following.]

On $\{x^\mu = 0\}$, the rank of (3.9) is zero, and the solutions with support in $\{x^\mu = 0\}$ are

$$\Box^N \prod_\lambda \delta(x^\lambda) = N! \sum_{\substack{\Sigma n_\lambda = N \\ n_\lambda \geq 0}} \frac{(-1)^{n_0}}{\prod_\lambda (n_\lambda !)} \prod_\lambda \delta^{(2n_\lambda)}(x^\lambda), \quad \forall N.$$

Finally, we look at the complete system (3.3). From $z^{\lambda\mu}\mathfrak{M}^{\rho\sigma}_{\lambda\mu} = 0$ and $\hat{z}^{\mu\lambda}\mathfrak{M}^{\rho\sigma}_{\lambda\mu} = 0$, we have

$$z^{\lambda\mu}\mathfrak{M}_{\lambda\mu} = \xi^\sigma \frac{\partial}{\partial x^\sigma} \tag{3.10}$$

and

$$\hat{z}^{\lambda\mu}\mathfrak{M}_{\lambda\mu} = \hat{\xi}^\sigma \frac{\partial}{\partial x^\sigma}, \tag{3.11}$$

with $\xi^\mu = z^{\mu\nu}x_\nu$ and $\hat{\xi}^\mu = \hat{z}^{\mu\nu}x_\nu$. Now, in a neighborhood of any point which does not belong to $R^4 \times C(\mathfrak{L}_0)$, Eq. (3.3) can be replaced by an equivalent system of six equations which contains (3.10), (3.11), and four linear combinations of $\mathfrak{M}_{\lambda\mu}$ (the system of these four last equations being of rank 4). So the rank of (3.3) is six [outside $R^4 \times C(\bar{\mathfrak{L}}_0)$], except when:

$\xi^\mu = 0$ or $\hat{\xi}^\mu = 0$ or ξ^μ proportional to $\hat{\xi}^\mu$,

then

$$r = 5;$$

$\xi^\mu = \hat{\xi}^\mu = 0$, then $r = 4$.

The system containing ten derivatives admits (when $r = 6$) four characteristic manifolds: We have already found three of them, $u_1, u_2, \mathbf{x}^2$; the fourth can be taken as

$$Q = \xi^2 = \xi^\mu \xi_\mu = (u_1^2 - 2)\mathbf{x}^2 - 2(\mathbf{x}, \tilde{\Lambda}, \mathbf{x}).$$

We do not discuss further the system (3.3). We now go back to the study of the complete system (1.9).

E. Central Distributions

The system (1.9) is equivalent to

$$\begin{cases} \left(C^{\lambda\mu}_{(i)}(z)\mathfrak{M}^{\rho\sigma}_{\lambda\mu}(z)\dfrac{\partial}{\partial z^{\rho\sigma}} + C^{\lambda\mu}_{(i)}(z)\mathfrak{M}_{\lambda\mu}(x)\dfrac{\partial}{\partial x^\nu}\right) \times \tilde{T}(z, x) = 0, \\ \hat{\xi}^\nu \dfrac{\partial}{\partial x^\nu}\tilde{T}(z, x) = 0, \\ \xi^\nu \dfrac{\partial}{\partial x^\nu}\tilde{T}(z, x) = 0, \\ \Omega^{\mu\nu}(z)\dfrac{\partial}{\partial x^\nu}\tilde{T}(z, x) = 0, \end{cases}$$

where the $C^{\lambda\mu}_{(i)}$, $i = 1, 2, 3, 4$, define four linear combinations of the six $\mathfrak{M}_{\lambda\mu}$ such that the rank of

$$\| C^{\lambda\mu}_{(i)}(z)\mathfrak{M}^{\rho\sigma}_{\lambda\mu}(z) \|$$

is 4 outside $R^4 \times C(\bar{\mathfrak{L}}_0)$. Now only four among the six last equations can be independent, so it is obvious that the rank r of the system (1.9) is at most 8; in fact:

(1) $r = 8$ if $u_2 \neq 0$, with the corresponding solutions

$$T = T(u_1, u_2). \tag{3.12}$$

We have seen that, on $\{u_2 = 0\}$, $\|\Omega^{\mu\nu}\|$ is of rank 2 outside $R^4 \times C(\bar{\mathfrak{L}}_0)$. On the other hand, the term $\xi^\nu\partial/\partial x^\nu$ is a linear combination of $\Omega^{\mu\nu}\partial/\partial x^\nu$,

$$4\xi^\nu = (u_1 x_\mu - \xi_\mu)\Omega^{\mu\nu}, \quad \text{if} \quad u_2 = 0$$

[the coefficients $(u_1 x_\mu - \xi_\mu)$ of the linear combination vanish simultaneously if and only if $x_\mu = 0, \forall \mu$, but this implies $\xi_\mu = 0$], and so there remain seven equations which are independent unless

$$\hat{\xi}^\mu = 0 \tag{3.13}$$

or

$$\hat{\xi}^\nu = C_\mu(z, x)\Omega^{\mu\nu}(z) \neq 0. \tag{3.14}$$

If we introduce $\hat{\eta}_\mu = \hat{z}_{\mu\nu}\hat{\xi}^\nu$, we can see that (3.13) is equivalent to

$$\hat{\xi}^0 = 0, \quad \hat{\eta}^0 = 0, \tag{3.13'}$$

and (3.14) is equivalent to

$$(u_1^2 - 4) = 0, \quad \hat{\eta}^0 = 0, \tag{3.14'}$$

and so

(2) $r = 7$ if $u_2 = 0$ outside $\{\hat{\xi}^0 = 0, \hat{\eta}^0 = 0\} \cup [(u_1^2 - 4) = 0$ and $\hat{\eta}^0 = 0]$, with the solutions

$$\delta(u_2)T(u_1, Q), \tag{3.15}$$

$$\delta(u_2)\theta(4 - u_1^2)\theta(Q)\epsilon(\hat{\xi}^0)T(u_1, Q), \tag{3.16}$$

$$\delta(u_2)\theta(4 - u_1^2)\theta(-Q)\epsilon(\hat{\eta}^0)T(u_1, Q), \tag{3.17}$$

$$\delta(u_2)\theta(4 - u_1^2)\delta^{(k)}(Q)T(u_1, \epsilon(\hat{\xi}^0), \epsilon(\hat{\eta}^0)), \tag{3.18}$$

$$\delta(u_2)\delta^{(k)}(4 - u_1^2)\epsilon(\hat{\eta}^0)T(Q, \epsilon(u_1)). \tag{3.19}$$

(3a) $r = 6$ if $u_2 = 0$, $\hat{\xi}^0 = 0$, and $\hat{\eta}^0 = 0$ outside $(R^4 \times C(\bar{\mathfrak{L}}_0))$.

One can see that any solution with support in $\{u_2 = 0, \hat{\xi}^0 = 0$, and $\hat{\eta}^0 = 0\}$ is of the form

$$\sum_{p,q} \delta(u_2)\delta^{(p)}\left(\frac{\hat{\xi}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)\delta^{(q)}\left(\frac{\hat{\eta}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)T_{p,q}(u_1), \tag{3.20}$$

but any distribution of that form is not a solution. We do not need the precise description of solutions, and so we do not discuss it; however, we note that

$$(\mathbf{P}^2)^m(\mathbf{W}^2)^n\delta(u_2)\delta\left(\frac{\hat{\xi}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)\delta\left(\frac{\hat{\eta}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)T(u_1) \quad (3.21)$$

are solutions [$(\mathbf{P}^2)$ and $(\mathbf{W}^2)$ are Laplace operators].

(3b) $r = 6$ if $u_2 = 0$, $u_1^2 = 4$, and $\hat{\eta}^0 = 0$ outside $(R^4 \times C(\bar{\mathfrak{L}}_0))$.

That manifold is the intersection of the characteristic manifolds $u_2 = 0$, $u_1^2 = 4$, and $Q = 0$, but $\{u_1^2 = 4\}$ and $\{Q = 0\}$ are tangent together and the regularity conditions assumed in the study of (1.15) are not fulfilled. The system (1.9) restricted to that manifold has a new characteristic manifold

$$\lambda = \hat{\xi}^0/(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}. \quad (3.22)$$

(We note that $\hat{\xi}^\mu\hat{\xi}^\nu/\hat{z}^{\mu\alpha}\hat{z}^\nu{}_\alpha$ is independent of μ and ν.) The solutions of (1.9), with the given support, are of the form

$$\sum_n \sum_{2p+q=2n} \delta(u_2)\delta^{(p)}(u_1^2 - 4)\delta^{(q)}\left(\frac{\hat{\eta}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)T_{p,q}(\lambda) \quad (3.23)$$

and

$$\delta(u_2)\delta(u_1^2 - 4)\delta^{(q)}\left(\frac{\hat{\eta}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)\frac{1}{\lambda^q}. \quad (3.24)$$

[Any distribution of the form (3.23) is not a solution, but

$$(\mathbf{W}^2)^n\delta(u_2)\delta(u_1^2 - 4)\delta\left(\frac{\hat{\eta}^0}{(\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)T(\lambda) \quad (3.25)$$

are solutions.]

(4) $r = 3$ on $R^4 \times C(\bar{\mathfrak{L}}_0)$ outside $C(\bar{P})$. The corresponding solutions are

$$(\mathbf{W}^2)^n \prod_{\mu<\nu} \delta(\hat{z}^{\mu\nu})T(\mathbf{x}^2, \epsilon(u_1)), \quad (3.26)$$

$$(\mathbf{W}^2)^n \prod_{\mu<\nu} \delta(\hat{z}^{\mu\nu})\theta(\mathbf{x}^2)\epsilon(x^0)T(\mathbf{x}^2, \epsilon(u_1)), \quad (3.27)$$

$$(\mathbf{W}^2)^n \prod_{\mu<\nu} \delta(\hat{z}^{\mu\nu})\delta^{(k)}(\mathbf{x}^2)[a + b\epsilon(u_1)]\epsilon(x^0), \quad (3.28)$$

$$(A)^p(B)^q \prod_{\mu<\nu} \delta(\hat{z}^{\mu\nu})[a + b\epsilon(u_1)], \quad (3.29)$$

where A and B are the Laplace operators of $SL(2, \mathbb{C})$.

(5) $r = 0$ on $C(\bar{P})$ with solutions

$$(\mathbf{P}^2)^m(\mathbf{W}^2)^n \prod_{\mu<\nu} \delta(\hat{z}^{\mu\nu}) \prod_\lambda \delta(x^\lambda)[a + b\epsilon(u_1)]. \quad (3.30)$$

We give briefly the geometrical interpretation of all "signs" which appear in these formulas. The vector $\hat{\xi}^\mu(u_2 = 0)$ lies in the invariant 2-plane we talked about when discussing the conjugation under R^4; when $|u_1| < 2$ (2-plane with timelike directions), the sign of the component of $\hat{\xi}^\mu$ along a time direction (if $Q > 0$) or a space direction ($Q < 0$) is Lorentz invariant. So is $\hat{\xi} = 0$. A similar situation occurs when $|u_1| = 2$ (2-plane tangent to the light cone): The sign of the space component of $\hat{\xi}^\mu$ is invariant for $Q < 0$, and the sign of the "light" component is invariant for $Q = 0$. It only remains to notice that the vector $\hat{\eta}^\mu$ also lies in the invariant 2-plane, in which it is orthogonal to $\hat{\xi}^\mu$ so that its time component up to a factor is just the space component of $\hat{\xi}$.

As a conclusion to this section, we emphasize the fact that we have found central distributions with nonzero transversal order. This means that the linear forms they define also depend on the values of the derivatives of the test-functions transversal to the corresponding class of conjugation, i.e., in some sense on the geometry of the conjugation classes in the neighborhood of this class.

4. CALCULATION OF THE CHARACTERS

We now calculate those among the central distributions of $\bar{P}$ which are eigendistributions of the Laplace operators of the group. As we have said, the program fails on the point $m^2 = 0$, $w^2 = 0$, where we have to change our space of solutions and solve Eqs. (1.9) and (1.10) again.

As in the previous section, we do not make explicit the calculation of limits coming from the singularities of the supports of central distributions.

The Laplace operators associated with $\mathbf{P}^2 = P_\mu P^\mu$ and $\mathbf{W}^2 = W^\mu W_\mu$, with $W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}M_{\nu\rho}P_\sigma$, are obtained from the differential expressions

$$P_L^\mu = \frac{\partial}{\partial x_\mu}, \quad (4.1)$$

$$2M_L^{\mu\nu} \equiv u_1\frac{\partial}{\partial z_{\mu\nu}} + u_2\frac{\partial}{\partial \hat{z}_{\mu\nu}} + \hat{z}^{\alpha\mu}\frac{\partial}{\partial \hat{z}_\nu^\alpha} - \hat{z}^{\alpha\nu}\frac{\partial}{\partial \hat{z}_\mu^\alpha} + 2x^\mu\frac{\partial}{\partial x_\nu} 2x^\nu\frac{\partial}{\partial x_\mu}.$$

A. Central Eigendistributions of $\mathbf{P}^2$, $m^2 \neq 0$

$\mathbf{P}^2$ does not act on the $\hat{z}_{\rho\sigma}$ variables, so we can look separately at its action on distributions with support in $u_1^2 - 4 \gtrless 0$.

Without giving the details of the open sets we use for our step-by-step calculation, we, however, for the sake of geometrical interpretation, draw, for a given Lorentz transformation $\{\hat{z}_{\mu\nu}\}$ with $u_2 = 0$, the intersection of the quadric Q with the 2-plane the transformation leaves invariant in M. See Fig. 1. This is

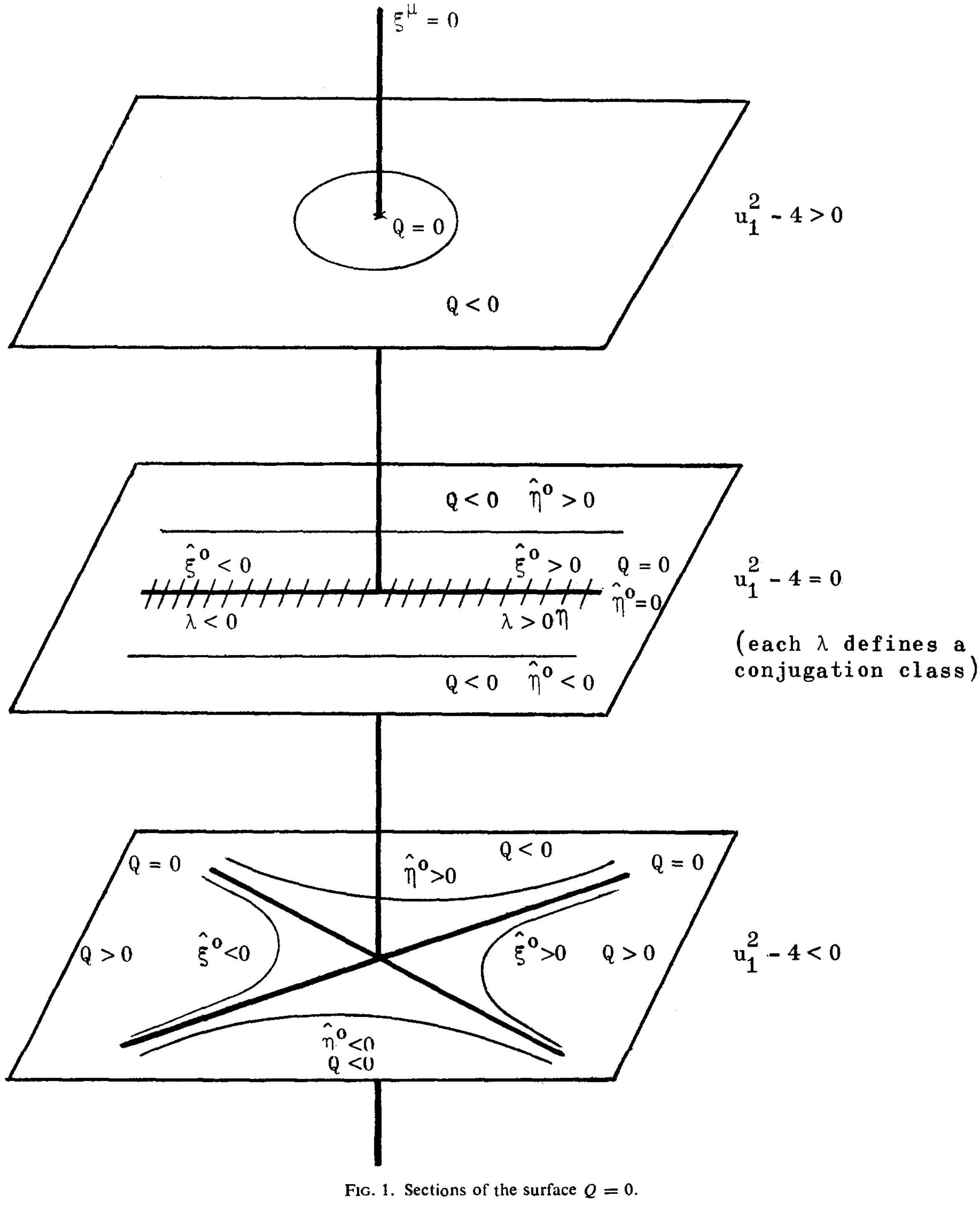


FIG. 1. Sections of the surface $Q = 0$.

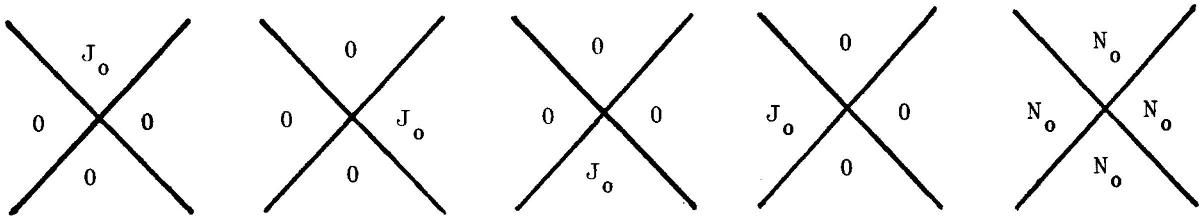


FIG. 2. Solutions of Eq. (4.3) outside $\hat{\xi}^0 = \hat{\eta}^0 = 0$.

easily understood from the algebraic identity

$$Q = \hat{\xi}^2 = \hat{\eta}_\mu \hat{\eta}_\nu + (u_1^2 - 4)\hat{\xi}_\mu \hat{\xi}_\nu / \hat{z}_{\mu\alpha}\hat{z}_\nu^\alpha, \quad \forall\, \mu, \nu, u_2 = 0. \quad (4.2)$$

We first look at the central eigendistributions of $\mathbf{P}^2$ outside $Q = 0$, $u_1^2 - 4 = 0$. For distributions T of the form (3.12), $\mathbf{P}^2 T = 0$ and we have no contribution (we have supposed $m^2 \neq 0$). For distributions of the forms (3.15) and (3.16), we are led to solve

$$\delta(u_2)\left(4(u_1^2 - 4)Q\frac{\partial^2 T}{\partial Q} + 4(u_1^2 - 4)\frac{\partial T}{\partial Q} + m^2 T\right) = 0. \quad (4.3)$$

The solutions of (4.3) are easily found to be

$$\delta(u_2)[J_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))F(u_1) + N_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))G(u_1)], \quad (4.4)$$

where J_0 and N_0 are the regular and singular Bessel functions of order 0 (we choose the determination of the square root with argument in $[0, \pi]$; F and G are arbitrary distributions depending only on u_1).

If we now include $Q = 0$, but stay outside $\hat{\xi}^0 = \hat{\eta}^0 = 0$, we have to look at the action of $\mathbf{P}^2$ on (3.18) and take into account the derivatives of θ functions in (3.16) and (3.17).

We see easily that $\mathbf{P}^2$ acting on (3.18) with $\delta^{(k)}(Q)$ gives a distribution of the same type with $\delta^{(k+1)}(Q)$, so that it cannot contribute to an eigendistribution (k is only allowed a finite number of values).

Being careful with the derivatives of $\theta(Q)$ [i.e., using limiting processes and writing $Q = 0$ with help of (4.2)], we get, for (3.16) and (3.17), the following eigendistributions of $\mathbf{P}^2 - m^2$:

$$\delta(u_2)\theta(4 - u_1^2)J_0((m^2 Q)^{\frac{1}{2}}/(4 - u_1^2))\theta(Q)\theta(\pm\hat{\xi}^0)F_\pm^+(u_1),$$
$$\delta(u_2)\theta(4 - u_1^2)J_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))\theta(-Q)\theta(\pm\hat{\eta}^0)F_\pm^-(u_1),$$
$$\delta(u_2)\theta(4 - u_1^2)N_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))G(u_1). \quad (4.5)$$

See Fig. 2.

Adding $\hat{\xi}^0 = \hat{\eta}^0 = 0$, but outside $\{\hat{z}^{\mu\nu} = 0\}$, we have to look at the action of $\mathbf{P}^2$ on (3.20). Again, we see easily that the transversal order is raised. Now we have to see what happens to (4.4) and (4.5) on $\hat{\xi}^0 = \hat{\eta}^0 = 0$.

Moreover, we also take into account the values of m^2, noticing that, if we want our Bessel functions to define distributions, they must not have an exponential behavior near $u_1^2 - 4 = 0$ (we see further that the distributions depending on u_1 which we use are not 0 on $u_1^2 - 4 = 0$).

With the notations

$$I_0(x) = J_0(ix) \quad \text{and} \quad K_0(x) = \tfrac{1}{2}i\pi[J_0(ix) + iN_0(ix)],$$

we have the solutions for $m^2 > 0$,

$$\delta(u_2)\theta(4 - u_1^2)I_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))\theta(Q)\epsilon(\hat{\xi}^0)F_1(u_1), \quad (4.6)$$

$$\delta(u_2)\theta(4 - u_1^2)J_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))\theta(-Q)\epsilon(\hat{\eta}^0)F_2(u_1), \quad (4.7)$$

$$\delta(u_2)\theta(4 - u_1^2)[I_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))\theta(Q) + J_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))\theta(-Q)]F_3(u_1), \quad (4.8)$$

$$\delta(u_2)\theta(4 - u_1^2)[(2/\pi)K_0((m^2 Q)^{\frac{1}{2}})\theta(Q) - N_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))\theta(-Q)]F_4(u_1). \quad (4.9)$$

See Fig. 3. For $m^2 < 0$, we get the solutions

$$\delta(u_2)\theta(u_1^2 - 4)J_0((m^2 Q)^{\frac{1}{2}}/(u_1^2 - 4))G(u_1), \quad (4.10)$$

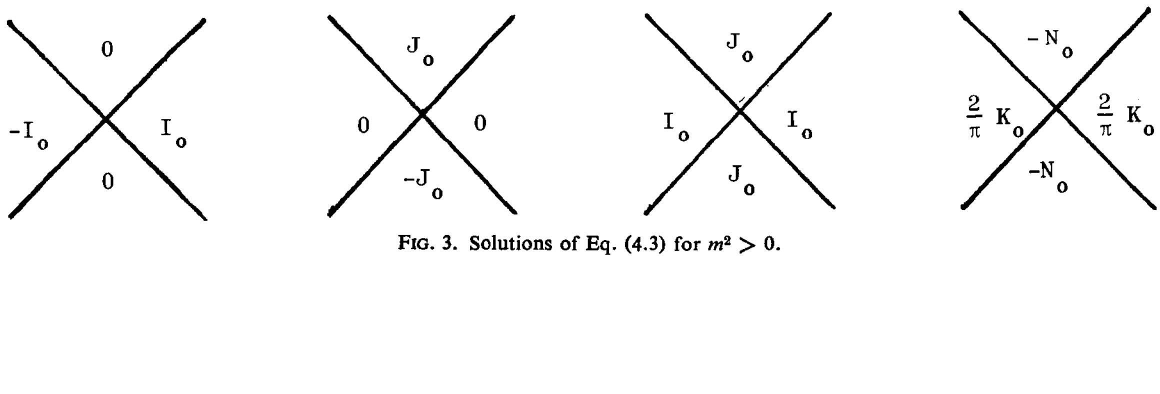


FIG. 3. Solutions of Eq. (4.3) for $m^2 > 0$.

$$\delta(u_2)\theta(4-u_1^2)J_0((m^2Q)^{\frac{1}{2}}/(u_1^2-4))\theta(Q)\epsilon(\hat{\xi}^0)G_1(u_1), \tag{4.11}$$

$$\delta(u_2)\theta(4-u_1^2)[N_0((m^2Q)^{\frac{1}{2}}/(u_1^2-4))\theta(Q) - (2/\pi)K_0((m^2Q)^{\frac{1}{2}}/(u_1^2-4))\theta(-Q)]G_2(u_1). \tag{4.12}$$

See Fig. 4. (All Bessel functions have real arguments.)

For any other complex value of m^2, (4.11) and (4.12) are the only solutions.

If we let $u_1^2 - 4$ be 0 with $\{\hat{z}_{\mu\nu} \neq 0\}$, we have to look at the action of $\mathbf{P}^2$ on (3.19). From (4.3), we see that, this time, the transversal order is lowered, so that we cannot have eigendistributions of this type. We should also look at the action of $\mathbf{P}^2$ on (3.23) and (3.24), but we can see that the action of $\mathbf{W}^2$ raises their transversal order, so that they are of no interest to us now.

Let us mention that

$$\mathbf{P}^2\left[\delta(u_2)\delta(u_1^2-4)\delta\left(\frac{\hat{\eta}^0}{(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)T(\lambda)\right] = -\delta(u_2)\delta(u_1^2-4)\delta\left(\frac{\hat{\eta}^0}{(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}}\right)\frac{\partial^2 T}{\partial\lambda^2},$$

so that

$$\delta(u_2)\delta(u_1^2-4)\delta[\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}](ae^{im\lambda}+be^{-im\lambda}) \tag{4.13}$$

(a and b any constants) is an eigendistribution of $\mathbf{P}^2$.

We do not look at the action of $\mathbf{P}^2$ on (3.26) – (3.30) because such distributions cannot be eigendistributions of $\mathbf{W}^2$.

B. Simultaneous Central Eigendistributions of $\mathbf{P}^2$ and $\mathbf{W}^2$

We shall now look for those among the eigendistributions of $\mathbf{P}^2$ which are also eigendistributions of $\mathbf{W}^2$. Our last calculation showed that, outside $u_1^2 - 4 = 0$, the solutions of $\mathbf{P}^2 - m^2 = 0$ depended only on the variables u_1, u_2 [by $\delta(u_2)$] and $Z = Q/(u_1^2 - 4)$.

Now, in the same way we used the fact that $\mathbf{P}^2$ did not act on the $\hat{z}_{\mu\nu}$ variables, we shall use the fact that $\mathbf{W}^2$ does not act on u_2 and Z. More precisely, a calculation valid in some open set U where $u_1^2 - 4 \neq 0$ and $Q \neq 0$, for distributions $\delta(u_2)T$ such that

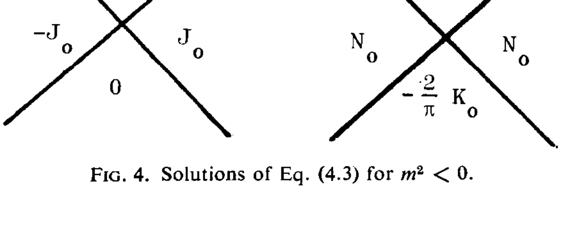


FIG. 4. Solutions of Eq. (4.3) for $m^2 < 0$.

$\delta(u_2)(\mathbf{P}^2 - m^2)T = 0$, gives

$$4\mathbf{W}^2[\delta(u_2)T(u_1, Z)] = m^2\delta(u_2)\left((u_1^2-4)\frac{\partial^2 T}{\partial u_1^2} + Su_1\frac{\partial T}{\partial u_1} + \frac{3u_1^2-4}{u_1^2-4}T\right); \tag{4.14}$$

i.e., $\mathbf{W}^2$ acting on such distributions is reduced to a second-order differential operator in u_1.

We are thus led to replace the products $\theta(\pm Q) \times \theta(\pm(u_1^2-4))$ in formulas (4.6)–(4.12) with the equivalent and suitably fitted products $\theta(\pm Z)\theta(\pm(u_1^2-4))$.

The solution of the eigenvalue problem in the whole group will be the cumbersome part of this paragraph.

In U, Eq. (4.14) allows us to determine the distributions $F(u_1)$ and $G(u_1)$ of formulas (4.6)–(4.12). Indeed, the eigenequation $\mathbf{W}^2 - w^2 = 0$ then leads to

$$(u_1^2-4)T''_{u_1{}^2} + 5u_1T'_{u_1} + (3u_1^2-4)/(u_1^2-4)T = (4w^2/m^2)T. \tag{4.15}$$

This is a Fuchsian equation of first type with two singular-regular points at $u_1 = \pm 2$. It has two solutions for each right or left neighborhood of these two points. Again, by analogy with $SL(2, R)$, we call these regular and singular. They have the behavior

$$R(u_1) = A(u_1)/(u_1^2-4)^{\frac{1}{2}}, \quad S(u_1) = B(u_1)/(u_1^2-4), \tag{4.16}$$

with $A(u_1)$ and $B(u_1)$ holomorphic functions.

We now have to see if the distributions (4.6)–(4.12), where the F and G are replaced by $R(u_1)$ and $S(u_1)$, are eigendistributions on the whole $\bar{P}$. For this calculation, we use the method of extension given in Sec. I [practically, we replace, for the derivations, $\theta(\pm(u_1^2-4))$ by $\theta(\pm u(_1^2-4)-\epsilon)$ and then take the limit $\epsilon \to 0$]. The details of these tedious calculations are given in Appendix C. The results are (in the open sets $u_1 > -2$ or $u_1 < 2$) the following:

1. With $F, G = R(u_1) = A_1(u_1)/(u_1^2-4)^{\frac{1}{2}}$:

if $m^2 > 0$:

$$(\mathbf{W}^2 - w^2)(4.6) = (i/m)A(\pm 2)\delta(u_2)\delta(u_1^2-4) \times \delta(\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}})\sinh m\lambda\,\epsilon(\hat{\xi}^0),$$

$$(\mathbf{W}^2 - w^2)(4.7) = 0, \tag{4.17}$$

$$(\mathbf{W}^2 - w^2)(4.8) = (i/m)A(\pm 2)\delta(u_2)\delta(u_1^2-4) \times \delta(\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}})\cosh m\lambda,$$

$$(\mathbf{W}^2 - w^2)(4.9) = 0;$$

if $m^2 < 0$:

$$(\mathbf{W}^2 - w^2)(4.10) = (1/m)A(\pm 2)\delta(u_2)\delta(u_1^2 - 4) \times \delta(\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}) \cos m\lambda,$$
$$(\mathbf{W}^2 - w^2)(4.11) = (i/m)A(\pm 2)\delta(u_2)\delta(u_1^2 - 4) \times \delta(\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}}) \sin m\lambda\epsilon(\hat{\xi}^0),$$
$$\mathbf{W}^2 - w^2)(4.12) = -(i/m)A(\pm 2)\delta(u_2)\delta(u_1^2 - 4) \times \delta(\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha))^{\frac{1}{2}} \cos m\lambda. \quad (4.18)$$

2. With $F, G = S(u_1) = B(u_1)/(u_1^2 - 4)$:

if $m^2 > 0$:

$$\begin{aligned}(\mathbf{W}^2 - w^2)(4.6) &= 0,\\ (\mathbf{W}^2 - w^2)(4.7) &= \prod_{\alpha<\beta} \delta(\hat{z}_{\alpha\beta})\Delta_m^1(\mathbf{x}^2),\\ (\mathbf{W}^2 - w^2)(4.8) &= \prod_{\alpha<\beta} \delta(\hat{z}_{\alpha\beta})\Delta_m^2(\mathbf{x}^2),\\ (\mathbf{W}^2 - w^2)(4.9) &= \prod_{\alpha<\beta} \delta(\hat{z}_{\alpha\beta})\Delta_m^3(\mathbf{x}^2);\end{aligned} \quad (4.19)$$

if $m^2 < 0$:

$$\begin{aligned}(\mathbf{W}^2 - w^2)(4.10) &= 0,\\ (\mathbf{W}^2 - w^2)(4.11) &= 0,\\ (\mathbf{W}^2 - w^2)(4.12) &= \prod_{\alpha<\beta} \delta(\hat{z}_{\alpha\beta})\Delta_{im}^4(\mathbf{x}^2).\end{aligned} \quad (4.20)$$

We were not able to calculate the distributions $\Delta^i(\mathbf{x}^2)$ [which satisfy $(\mathbf{P}^2 \mp m^2)\Delta^i = 0$], so we cannot assert they are different from 0 and, for the three Δ_m, linearly independent.

On the other hand, we can calculate that $\mathbf{W}^2$ acting on any central distribution with support $\{\hat{\eta}^0 = 0, u_1^2 - 4 = 0\}$ or $\{\hat{z}^{\mu\nu} = 0 \; \forall \; \mu, \nu\}$ raises its transversal order. Thus, the only simultaneous eigendistributions of $\mathbf{P}^2$ and $\mathbf{W}^2$ are those lines or combinations of lines in (4.17), (4.18), (4.19), and (4.20) which lead to a right-hand side 0; i.e., with $R(u_1)$

$$(4.7),\ (4.9),\ (4.10) - i(4.12),$$

with $S(u_1)$

$$(4.6),\ (4.7),\ (4.11). \quad (4.21)$$

C. Spectrum of $\mathbf{W}^2$. Character of the Representations $m^2 \neq 0$

1. Explicit Expression of R and S

The differential equation (2.2) can be solved explicitly with a change of variables suited to the sets $|u_1| \gtrless 2$ and a change of function:

$$u_1 > 2 \quad \text{or} \quad u_1 < -2, \quad u_1 = \pm 2 \cosh \tfrac{1}{2}\eta, \quad \eta > 0, \quad g(\eta) = f/\sinh 2\tfrac{1}{2}\eta,$$
$$|u_1| < 2, \quad u_1 = 2 \cosh \tfrac{1}{2}\varphi, \quad 0 < \varphi < 2\pi, \quad g(\varphi) = f/\sin^2 \tfrac{1}{2}\varphi.$$

The equation then becomes

$$g'' \pm (\tfrac{1}{4} - w^2/m^2)g = 0, \quad (4.22)$$

and we see that the solutions of (2.2) are

$$\begin{aligned}|u_1| > 2, \quad R(u_1) &= \sinh \sigma\eta/\sinh^2 \tfrac{1}{2}\eta,\\ S(u_1) &= \cosh \sigma\eta/\sinh^2 \tfrac{1}{2}\theta,\\ |u_1| < 2, \quad R(u_1) &= \sin \sigma\varphi/\sin^2 \tfrac{1}{2}\varphi,\\ S(u_1) &= \cos \sigma\varphi/\sin^2 \tfrac{1}{2}\varphi\end{aligned}$$

with $\sigma = (\frac{1}{4} - \frac{1}{2}w^2)^{\frac{1}{2}}$ (there is, of course, no fixed choice of normalization for R and S across their singularities).

2. Spectrum of $\mathbf{W}^2$

Exactly as for $SL(2, R)$, the existence of a center with two elements in $\bar{P}$ enforces the "parity" condition

$$\langle \chi, \varphi_- \rangle = \pm \langle \chi, \varphi \rangle, \quad \varphi_-(\mathbf{x}, \hat{z}) = \varphi(\mathbf{x}, -\hat{z}).$$

Through the transformation $\hat{z}^{\mu\nu} \to -\hat{z}^{\mu\nu}$, Q and $\hat{\eta}^0$ are invariant, and $\hat{\xi}^0$ and u_1 take opposite signs. We see that the solutions (2.8) do not mix with one another, so that we must have

$$R(u_1) = \pm R(-u_1), \quad S(u_1) = \pm S(-u_1).$$

This condition enforces 2σ to be an integer for those solutions which are not 0 for $u_1^2 < 4$. With $2\sigma = 2j + 1$, $j = 0, \frac{1}{2}, 1, \cdots$, or $2\sigma = 2K - 1$, $K = \frac{1}{2}, 1, \cdots$ (we shall choose the definition with j for $m^2 > 0$ and with K for $m^2 < 0$ so as to get, at the end, the notations of the literature for the representations of $\bar{P}$), we thus have

$$w^2 = -m^2 j(j+1) \quad \text{or} \quad w^2 = m^2 K(1 - K).$$

For the solution (1.9) with $S(u_1)$, it only links the disconnected parts with supports $u_1 > 2$, < -2, which therefore must appear with coefficients of the same absolute value.

3. The Characters of $\bar{P}$, $m^2 \neq 0$

We now form, with the central eigendistributions we have found for each value of m^2 and w^2, the linear combinations which can be characters of irreducible representations, i.e., which possess definite parity properties and, eventually, for real values of m^2 and w^2, which are of positive type and extremal.

As in Sec. 1, however, we are just able to select bounded, Hermitian, and extremal distributions, and say they are candidates to be of positive type. Only comparisons with the known UIR's of $\bar{P}$ and with the results of Schrader[12] allow us to fix our last unknown coefficient.

We refer to Sec. 2 for the notations used relative to $SL(2, R)$.

(1) $m^2 > 0$: We have three solutions: (4.7) and (4.9) with R and (4.6) with S. Necessarily,

$$w^2 = -m^2 j(j+1).$$

The solution (4.6), unbounded and odd in $\mathbf{x}$, can at most be the character of nonunitary representation

$$\chi_{m,j} = \delta(u_2) I_0((m^2Q)^{\frac{1}{2}}/(4-u_1^2))\theta(Q)\epsilon(\xi^0) \times \theta(4-u_1^2)\cos(2j+1)\tfrac{1}{2}\varphi/\sin^2\tfrac{1}{2}\varphi.$$

With the other two solutions, we can form two Hermitian [i.e., satisfying $\chi(g) = \chi(g^{-1})$; here $g \to g^{-1}$ entails $Q \to Q$, $u_1 \to u_1$, $\xi_0 \to \xi_0$, $\eta_0 \to -\eta_0$] extremal distributions which are bounded:

$$(4.9) \pm ia(m,j)\,(4.7).$$

They can be the characters of the unitary "physical" representations $[m, j, \pm]$ of $\bar{P}$, induced by the rotation group. Indeed, comparison with Schrader's work[12] gives us $a(m,j) = 1$, and we have

$$\begin{aligned}\chi^{\pm}_{m,j} = \delta(u_2)[&(2/\pi)K_0((m^2Q)^{\frac{1}{2}}/(4-u_1))\theta(Q) \\ &+ iH_0^1((m^2Q)^{\frac{1}{2}}/(u_1^2-4))\theta(\pm\eta^0)\theta(-Q) \\ &- iH_0^2((m^2Q)^{\frac{1}{2}}/(u_1^2-4))\theta(-Q)\theta(\mp\eta^0)] \\ &\times \theta(4-u_1^2)\sin(2j+1)\tfrac{1}{2}\varphi/\sin^2\tfrac{1}{2}\varphi.\end{aligned}$$

All these three characters have parity $(-1)^{2j}$.

(2) $m^2 < 0$:

(i) $w^2 \neq m^2K(1-K)$: We have only two solutions built with (4.10) and S:

$$\begin{aligned}\chi^{\pm}_{im,\sigma} = \delta(u_2)&J_0((m^2Q)^{\frac{1}{2}}/(u_1^2-4)) \\ &\times [\theta(u_1-2)\cosh\sigma\eta/\sinh^2\tfrac{1}{2}\eta \\ &\pm \theta(-u_1-2)\cosh\sigma\eta/\sinh^2\tfrac{1}{2}\eta].\end{aligned}$$

If w^2 is not real or if $w^2/m^2 < -\frac{3}{4}$, $\chi^{\pm}_{im,\sigma}$ can be a character of a nonunitary representation of $\bar{P}$ "induced" by a nonunitary representation of $SL(2, R)$ (for $w^2/m^2 < -\frac{3}{4}$, χ is not bounded).

If $w^2/m^2 > \frac{1}{4}$, we find the characters of the representations $[im, \sigma, \pm]$ of $\bar{P}$ induced by the principal series of $SL(2, R)$.

For $0 < w^2/m^2 < \frac{1}{4}$, we have the characters of the representations induced by the complementary series of $SL(2, R)$, with $\chi^{\pm}_{im,\sigma}$ corresponding to a unitary representation.

For $-\frac{3}{4} \leq w^2/m^2 < 0$, $\chi^{\pm}_{im,\sigma}$, though bounded, can only be a character of a nonunitary representation.

(ii) $w^2 = m^2K(1-K)$: We have a 4-dimensional space of solutions: the two solutions of (i), (4.11) with S, and (4.10) $- i$(4.12) with R. These last two solutions have parity $(-1)^{2k}$, so that we can isolate the solution of α with parity $(-1)^{2k+1}$:

$$\begin{aligned}\chi^{\nearrow}_{im,K} = \delta(u_2)&J_0((m^2Q)^{\frac{1}{2}}/(u_1^2-4)) \\ &\times [\theta(u_1-2)\cosh(2K-1)\tfrac{1}{2}\eta/\sinh^2\tfrac{1}{2}\eta \\ &+ (-1)^{2K+1}\cosh(2K-1)\tfrac{1}{2}\eta/\sinh^2\tfrac{1}{2}\eta].\end{aligned}$$

For $K > 1$, it is unbounded and probably corresponds to the character of nonunitary representation of $\bar{P}$ "induced" by the nonunitary representation of $SL(2, R)$ we have found at this point. For $K = \frac{1}{2}$, it is the character of the unitary representation of $\bar{P}$ induced by the unitary representation $C^0_{\frac{1}{4}}$ of the complementary series.

The solution (4.10) $- i$(4.12) with R corresponds to the representations of $\bar{P}$ "induced" by the finite-dimensional representations of $SL(2, R)$. It is unitary only for $K = \frac{1}{2}$:

$$\begin{aligned}\chi^{f}_{im,K} = \delta(u_2)\Bigg\{&\theta(u_1^2-4)J_0\left(\frac{(m^2Q)^{\frac{1}{2}}}{u_1^2-4}\right) \\ &\times \left(\theta(u_1)\frac{\sinh(2K-1)\frac{1}{2}\eta}{\sinh^2\frac{1}{2}\eta} + (-1)^{2K}\theta(-u_1)\frac{\sinh(2K-1)\frac{1}{2}\eta}{\sinh^2\frac{1}{2}\eta}\right) \\ &- \theta(4-u_1^2)\left[N_0\left(\frac{(m^2Q)^{\frac{1}{2}}}{u_1^2-4}\right)\theta(Q) - \frac{2}{\pi}K_0\left(\frac{(m^2Q)^{\frac{1}{2}}}{4-u_1^2}\right)\theta(-Q)\right]\frac{\sin(2K-1)\frac{1}{2}\varphi}{\sin^2\frac{1}{2}\varphi}\Bigg\}.\end{aligned}$$

Last, we can form two bounded combinations which are the characters of the unitary representations $[im, K, \pm]$ of $\bar{P}$ induced by the two discrete series of $SL(2, R)$:

$$\begin{aligned}\chi^{\pm}_{im,K} = \delta(u_2)\Bigg\{&\theta(u_1^2-4)J_0\left(\frac{(m^2Q)^{\frac{1}{2}}}{u_1^2-4}\right) \\ &\times \left(\theta(u_1)\frac{e^{-(2K-1)\frac{1}{2}\eta}}{\sinh^2\frac{1}{2}\eta} + (-1)^{2K}\theta(-u_1)\frac{e^{-(2K-1)\frac{1}{2}\eta}}{\sinh^2\frac{1}{2}\eta}\right) \\ &\pm i\theta(4-u_1^2)\left[N_0\left(\frac{(m^2Q)^{\frac{1}{2}}}{u_1^2-4}\right)\theta(Q) - \frac{2}{\pi}K_0\left(\frac{(m^2Q)^{\frac{1}{2}}}{4-u_1^2}\right)\theta(-Q)\right]\frac{e^{\pm i(2K-1)\frac{1}{2}\varphi}}{\sin^2\frac{1}{2}\varphi}\Bigg\}.\end{aligned}$$

[We obtained, in fact, a coefficient $a(im, K)$ before $\theta(4-u_1^2)$, which we set equal to 1 with the help of Schrader.[12]]

(3) m^2 complex: We have only one solution: (4.11) with S and still the spectral condition for w^2:

$$w^2 = -m^2 j(j+1).$$

It might be a character of nonunitary representation

$$\chi_{m,j} = \delta(u_2) J_0\left(\frac{(m^2 Q)^{\frac{1}{2}}}{u_1^2 - 4}\right)\theta(Q)\delta(\xi^0) \times \theta(4 - u_1^2)\frac{\cos(2j+1)\frac{1}{2}\varphi}{\sin^2\frac{1}{2}\varphi}.$$

4. Frobenius' Formula

Again, it is worthwhile here to note how nicely it appears that the representations of $\bar{P}$ are induced. We have already stated Frobenius' formula for finite groups and seen its generalization for $SL(2, R)$.

Here, our inducing representations are those of the little groups $R^4 \times SU_2$ or $R^4 \times SL(2, R)$. The corresponding characters are of the form

$$\chi_w = e^{i\mathbf{p}\cdot\mathbf{x}},$$

where χ_w is a character of SU_2 or $SL(2, R)$, $\mathbf{x} \in R^4$, $\mathbf{p}^2 = m^2$.

Now, leaving aside $\chi^{\pm}_{im,k}$, the characters of $\bar{P}$, we find the structure

$$\chi_{w^2,m^2} = \delta(u_2)\frac{\chi_w}{\sin\frac{1}{2}\varphi, \sinh\frac{1}{2}\eta} \times B\left(\left(\frac{m^2 Q}{u_1^2 - 4}\right)^{\frac{1}{2}}\right),$$

where B is any Bessel function; i.e., they have for support the elements of $\bar{P}$ which are conjugates of elements of the inducing group which belong to the support of the character of the inducing representation; they contain the character χ_w time a "multiplicity" factor $1/\sin\frac{1}{2}\varphi$ or $1/\sinh\frac{1}{2}\eta$; they also contain, in a certain sense, $e^{i\mathbf{p}\cdot\mathbf{x}}$ with a multiplicity (Bessel functions being integral $e^{i\mathbf{p}\cdot\mathbf{x}}$).

The structure of $\chi^{\pm}_{im,k}$ is not as nice, due to the fact that, for $u_1 < 2$, the trace of a matrix of $SL(2, \mathbb{R})$ does not specify its conjugation class as it does in $SL(2, \mathbb{C})$.

D. Characters of the Representations $m^2 = 0$, $w^2 \neq 0$

1. Central Eigendistributions of $\mathbf{P}^2$

There is no major change here. Calculation runs similarly to that of Sec. 4C with, instead of the two Bessel functions J_0 and N_0, a constant and $\log|Q|$. We need only notice that the solutions of $\mathbf{P}^2 = 0$ are now defined up to the addition of distributions of the form $T(u_1, u_2)$.

Leaving those aside, we get zero for eigendistributions in $u_1^2 - 4 > 0$. For eigendistributions in $u_1^2 - 4 < 0$, we get

$$\begin{aligned} &\delta(u_2)\theta(Q)\epsilon(\hat{\xi}^0)T(u_1),\\ &\delta(u_2)\theta(-Q)\epsilon(\hat{\eta}^0)T(u_1), \qquad (4.23)\\ &\delta(u_2)\log|Q|T(u_1),\end{aligned}$$

with T an arbitrary distribution. Outside $\{\hat{z}^{\mu\nu} = 0\}$ we have, in addition,

$$\delta(u_2)\delta(u_1^2 - 4)T(Q, \epsilon(\hat{\eta}^0), \epsilon(u_1)). \qquad (4.24)$$

2. Simultaneous Eigendistributions of $\mathbf{P}^2$ and $\mathbf{W}^2$

The part played in Sec. 4B by the variable Z is now played by Q itself. Indeed, we see that $\mathbf{W}^2$ acting on distribution (4.23) [and also, of course, on distributions of type $T(u_1, u_2)$] gives 0, so that these will not interest us for the moment. We see that $\mathbf{W}^2$ acting on distribution (4.24) leads to the eigenequation

$$2Q^3\frac{\partial^4 T}{\partial Q^4} + 11Q^2\frac{\partial_3 T}{\partial Q^3} + 11Q\frac{\partial^2 T}{\partial Q^2} + \frac{\partial T}{\partial Q} - \frac{\omega^2}{8}T = 0. \qquad (4.25)$$

With $x = \sqrt[4]{-4w^2 Q}$ (we choose the determination of argument in $[0, \pi/2]$), we find that four independent solutions of (4.25), outside $Q = 0$, are

$$B_0(x)B_0'(ix), \qquad (4.26)$$

where $B_0, B_0' = J_0$ or N_0.

Calculation shows that the only global solutions are those whose behavior as $Q \to 0$ is

$$a,\ a\epsilon(\hat{\eta}^0),\ (-Q)^{\frac{1}{2}}\epsilon(\hat{\eta}^0),\ \log|Q|,$$

i.e., the four independent solutions (see Fig. 5)

$$J_0(x)J_0(ix), \qquad (4.27)$$

$$J_0(x)J_0(ix)\epsilon(\hat{\eta}_0), \qquad (4.28)$$

$$[J_0(x)N_0(ix) - N_0(x)J_0(ix)]\epsilon(\hat{\eta}_0), \qquad (4.29)$$

$$J_0(x)N_0(ix) + N_0(x)J_0(ix). \qquad (4.30)$$

See Fig. 5.

These solutions give rise to eight simultaneous eigendistributions of $\mathbf{P}^2$ and $\mathbf{W}^2$. Indeed, the parity requirement enforces the form

$$\delta(u_2)[(4.27)\text{–}(4.30)][\delta(u_1 - 2) \pm \delta(u_1 + 2)].$$

3. Characters of the Representations

We now extract, from the two sets of distributions which have definite parity, the linear combinations which are characters of unitary representations of $\bar{P}$. Once more, we can only guess that the remaining solutions are characters of nonunitary representations.

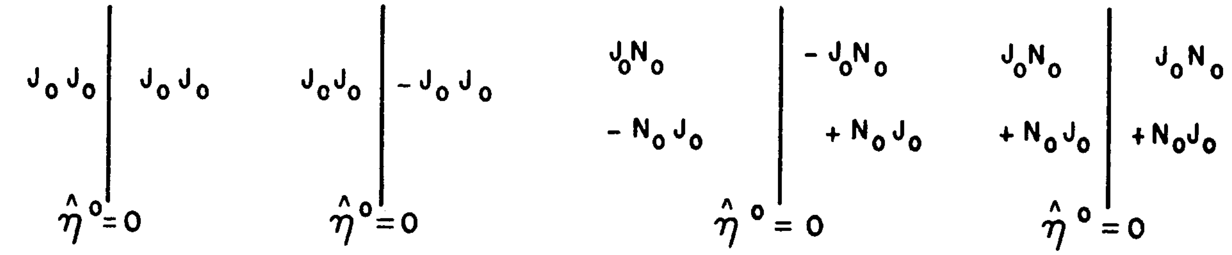


FIG. 5.

We restrict the real values of w^2 and search for bounded combinations of (4.27)–(4.30). If $w^2 > 0$, then $x > 0$ and, clearly, the only bounded product of the form (4.26) is $B_0(x)H_0^1(ix)$. Now, these cannot be obtained from (4.27)–(4.30), so we can conclude there is no unitary representations with $w^2 > 0$.

Similarly, we easily see that, for $w^2 < 0$, the only bounded products of the form (4.26) are

$$H_0^1(x)J_0(ix), \quad H_0^1(x)N_0(ix), \quad J_0(x)H_0^1(ix).$$

We thus find four global bounded solutions

$$\delta(u_2)[\delta(u_1 - 2) \pm \delta(u_1 + 2)] \times \{2iJ_0(x)J_0(ix) - [J_0(x)N_0(ix) + N_0(x)J_0(ix)]\},$$

$$\delta(u_2)[\delta(u_1 - 2) \pm \delta(u_1 + 2)] \times [J_0(x)N_0(ix) - N_0(x)J_0(ix)]\epsilon(\hat{\eta}^0).$$

Moreover, both the first line, being real, and the second, pure imaginary, possess Hermitian symmetry [through $g \to g^{-1}: \epsilon(\hat{\eta}^0) \to -\epsilon(\hat{\eta}^0)$].

The condition of extremality then gives us the form

$$\chi_{0,w}^{\epsilon_1,\epsilon_2} = \delta(u_2)[\delta(u_1 - 2) + \epsilon_1\delta(u_1 + 2)] \times \{2iJ_0(x)J_0(ix) - [J_0(x)N_0(ix) + N_0(x)J_0(ix) + \epsilon_2\alpha[J_0(x)N_0(ix) - N_0(x)J_0(ix)]\epsilon(\hat{\eta}^0)\}.$$

The coefficient α, which is real, should be determined by the fact that χ is of positive type and extremal. Then $\chi_{0,w}^{\epsilon_1\epsilon_2}$ would be the characters of the mass-0, continuous-spin, representations of $\bar{P}$. $\epsilon_1 = \mp 1$ tells us whether the representation is faithful or not for the center of $SL(2, C)$; ϵ_2 is the "sign of the energy."

4. *Frobenius' Formula*

Except that $\chi_{0,w}$ has for support $u_2 = 0$, $u_1 = \pm 2$, nothing expresses the fact that the corresponding representation is induced. But this is not a surprise: the inducing representation is a representation of $R^4 \times E_2$ which "kills" the rotations of the Euclidean group. The inducing character is $\delta(\varphi)J_0(w\,|z|)e^{i\mathbf{p}\cdot\mathbf{x}}$ [for an element

$$\begin{pmatrix} e^{i\varphi} & z \\ 0 & e^{-i\varphi} \end{pmatrix} \text{ of } E_2,$$

$\mathbf{x} \in R^4$, $\mathbf{p}^2 = 0$]. But elements with $\varphi = 0$ and different z are in the same class of conjugation of $SL(2, C)$, so that the multiplicative multiplicity factor of Frobenius formula is here partly replaced by an integral over z, which gives rise to the second Bessel function of the character [by means of the integral representation

$$H_0^\alpha(x)H_0^\beta(ix) \simeq \int_{\gamma^\alpha} e^{x^2/2z}H_0^\beta(z)\,\frac{dz}{z},$$

where γ^α is a suitable path, $\alpha, \beta = 1, 2$].

E. Characters of Helicity Representations, $m^2 = w^2 = 0$

1. *Generalities*

The first part of the last paragraph showed us that the space of central solutions of

$$\mathbf{P}^2T = 0, \tag{4.31}$$

$$\mathbf{W}^2T = 0, \tag{4.32}$$

in which we must pick up the characters of helicity representations, was infinite dimensional. This infinity has two origins: First, the arbitrary distribution in formula (4.23); second, the arbitrary additive term of the form $T(u_1, u_2)$.

But we know that, for unitary irreducible representations with $m^2 = w^2 = 0$, the operators $U(W_\mu)$ and $U(P_\mu)$ are proportional:

$$U(W_\mu) = i\lambda U(P_\mu)\ \forall\ \mu.$$

So the character of the representation must satisfy the supplementary equation

$$(W_\mu - i\lambda P_\mu)\chi = 0, \tag{4.33}$$

where W_μ and P_μ are now the differential operators. This condition, as we shall see, removes the first ambiguity.

In order to remove the second ambiguity (that is, to obtain well-defined solutions), we must solve Eqs. (4.31), (4.32), and (4.33) in a space other than $\mathfrak{D}'(\bar{P})$, namely in $\mathfrak{D}'_0(\bar{P})$, the dual space of

$$\mathfrak{D}_0(\bar{P}) = \left\{\varphi \in \mathfrak{D}(\bar{P}), \int_{R^4}\varphi(\mathbf{x}, \Lambda)\,d\mathbf{x} = 0, \quad \forall\, \Lambda \in \bar{\mathfrak{L}}_0\right\},$$

equipped with the induced topology (distributions which do not depend on $\mathbf{x}$ vanish on $\mathfrak{D}_0$).

The fact that the characters of the "zero-mass, finite-helicity" unitary representations cannot be distributions can be shown directly: If U is a strongly continuous unitary representation such that, for each $\varphi \in \mathfrak{D}$, the operator $U(\varphi)$ has a trace and thus is compact, then, for each $f \in L^1(G)$, $U(f)$ is compact (because $\mathfrak{D}$ is dense in L^1 and the representation of L^1 is norm continuous). It is known[3] that the "zero-mass, finite-helicity" representations do not have this property. By different methods, Schrader[12] had already pointed out that the characters of these representations could only be defined on $\mathfrak{D}_0$.

2. The Space $\mathfrak{D}_0'(\bar{P})$

Let $i:\mathfrak{D}_0 \to \mathfrak{D}$ be the canonical injection; the transposed application ${}^t i:\mathfrak{D}' \to \mathfrak{D}_0'$ is linear and continuous. Now there exists a (noncanonical) linear, continuous map $p:\mathfrak{D} \to \mathfrak{D}_0$ such that $p \circ i = \mathbf{1}_{\mathfrak{D}_0}$.

For some $f \in \mathfrak{D}(R^4)$ such that

$$\int_{R^4} f(\mathbf{x}) d^4x = 1,$$

we define

$$(p\varphi)(\mathbf{x}, \Lambda) = \varphi(\mathbf{x}, \Lambda) - f(\mathbf{x}) \int_{R^4} \varphi(\mathbf{y}, \Lambda) d^4y.$$

The transposed application is such that ${}^t i \circ {}^t p = \mathbf{1}_{\mathfrak{D}_0'}$; so ${}^t i$ is surjective and $\mathfrak{D}_0'$ is the quotient of $\mathfrak{D}'$ by the subspace $\operatorname{Ker} \mathfrak{D}_0 \subset \mathfrak{D}'$ of the distributions which vanish on $\mathfrak{D}_0$ (i.e., constant with respect to $\mathbf{x}$).

The partial derivation is naturally defined in $\mathfrak{D}_0'$ and is the transform by ${}^t i$ of the derivation in $\mathfrak{D}'$.

The product by a C^∞ function $\alpha(\mathbf{x}, \Lambda)$ depending effectively on $\mathbf{x}$ is not defined everywhere in $\mathfrak{D}_0'$ ($\varphi \in \mathfrak{D}_0 \not\Rightarrow \alpha\varphi \in \mathfrak{D}_0$). However, it is defined on elements of $\mathfrak{D}_0'$ which are derivatives with respect to $\mathbf{x}$:

$$\langle \alpha \partial_\mu T, \varphi \rangle = -\langle T, \partial_\mu(\alpha\varphi) \rangle, \; \partial_\mu(\alpha\varphi) \in \mathfrak{D}_0.$$

Then it is the transform by ${}^t i$ of the multiplication by α in $\mathfrak{D}'$.

3. Resolution of (1.9), (1.10), (4.31), (4.32), and (4.33) *in* $\mathfrak{D}_0'$

The differential operators which appear in the equations

$$\mathcal{T}_\mu T = 0, \tag{1.9}$$

$$\mathfrak{M}_{\mu\nu} T = 0, \tag{1.10}$$

$$(\mathbf{P}^2) T = 0, \tag{4.31}$$

$$(\mathbf{W}^2) T = 0, \tag{4.32}$$

$$(W_\mu - i\lambda P_\mu) T = 0 \tag{4.33}$$

have a meaning as operators in $\mathfrak{D}_0'$, and we have to solve this system of equations in $\mathfrak{D}_0'$. From the definition of $\operatorname{Ker} \mathfrak{D}_0$, we see that it is equivalent to solve

$$\partial_\rho \mathcal{T}_\mu T = 0, \tag{4.34a}$$

$$\partial_\rho \mathfrak{M}_{\mu\nu} T = 0, \tag{4.34b}$$

$$\partial_\rho (\mathbf{P}^2) T = 0, \tag{4.34c}$$

$$\partial_\rho (\mathbf{W}^2) T = 0, \tag{4.34d}$$

$$\partial_\rho (W_\mu - i\lambda P_\mu) T \to 0, \tag{4.34e}$$

in $\mathfrak{D}'$, each solution being well defined up to a distribution in $\operatorname{Ker} \mathfrak{D}_0$.

The central distributions already found are obviously solutions of the new "class equations" (4.34a) and (4.34b). But we have now other solutions, namely (mod $\operatorname{Ker} \mathfrak{D}_0$)

$$\delta(u_2)\theta(4 - u_1^2) T_\pm(u_1) \log |\eta^\mu \pm (4 - u_1^2)^{\frac{1}{2}} \xi^\mu| \tag{4.35}$$

(which are independent of μ, mod $\operatorname{Ker} \mathfrak{D}_0$; we shall take $\mu = 0$ in the following).

It is quite remarkable that these solutions depend on variables which are no longer constant on the conjugation classes of $\bar{P}$.

We note that the sum of two distributions of the form (4.35) with opposite signs in the argument of the log, and with $T_+(u_1) = T_-(u_1) = T(u_1)$, is equal (mod $\operatorname{Ker} \mathfrak{D}_0$) to the "old" solution

$$\delta(u_2)\theta(4 - u_1^2) T(u_1) \log |Q|.$$

Distributions (4.23) remain solutions of (4.34c) and (4.34d), and so do distributions (4.24) with $T(Q)$ solution of (4.25) with $w^2 = 0$; that is,

$$\delta(u_2)\delta(u_1 \pm 2)[a_\pm \log |Q| + b_\pm \epsilon(\eta^0) + c_\pm (-Q)^{\frac{1}{2}} \epsilon(\eta^0)]. \tag{4.36}$$

New solutions appear, namely

$$\delta(u_2) T(u_1) Q. \tag{4.37}$$

Last, distributions of the form (4.35) are solutions of (4.34c) and (4.34d); as for $m^2 \neq 0$, $\mathbf{W}^2$ raises the transversal order of the other central distributions.

We now look for solutions of (4.34c) among distributions (4.23), (4.35), (4.36), and (4.37).

For (4.35), outside $u_1 = \pm 2$, we have

$$(P_\mu)\delta(u_2)\theta(4 - u_1^2)\alpha_\pm(u_1) \log |\eta^0 \pm (4 - u_1^2)^{\frac{1}{2}} \xi^0| = \delta(u_2)\theta(4 - u_1^2)\alpha_\pm(u_1) \frac{\xi^{0\alpha} \xi_{\mu\alpha} \pm (4 - u_1^2)^{\frac{1}{2}} \xi^0{}_\mu}{\eta^0 \pm (4 - u_1^2)^{\frac{1}{2}} \xi^0}$$

and [using the relation $Q = (\hat{\eta}^\mu \hat{\xi}^\nu - \hat{\eta}^\nu \hat{\xi}^\mu)/\hat{z}^{\mu\nu}$, $\forall\, \mu \neq \nu$ if $u_2 = 0$]

$$2(W_\mu)\delta(u_2)\theta(4 - u_1^2)\alpha_\pm(u_1) \log |\hat{\eta}^0 \pm (4 - u_1^2)\hat{\xi}^{\frac{1}{2}0}|$$
$$= \delta(u_2)\theta(4 - u_1^2)\left(\frac{\pm u_1}{(4 - u_1^2)^{\frac{1}{2}}} \alpha_\pm(u_1) \mp (4 - u_1^2)^{\frac{1}{2}}\alpha'_\pm(u_1)\right) \frac{\hat{z}^{0\alpha}\hat{z}_{\mu\alpha} \pm (4 - u_1^2)^{\frac{1}{2}}\hat{z}^0\mu}{\hat{\eta}^0 \pm (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0},$$
$$\bmod \operatorname{Ker} \mathfrak{D}_0.$$

So (4.34c) gives

$$u_1\alpha_\pm(u_1) - (4 - u_1^2)\alpha'_\pm(u_1) = \pm 2i\lambda(4 - u_1^2)^{\frac{1}{2}}\alpha_\pm(u_1),$$

and with $u_1 = 2\cos\varphi$, $0 \leq \varphi \leq \pi$,

$$\alpha_\pm(u_1) = \alpha_\pm e^{\pm 2i\lambda\varphi}/\sin\varphi, \quad \alpha_\pm \text{ const.}$$

Now the only global solution is the combination

$$S_\lambda = -\delta(u_2)\theta(4 - u_1^2) \times \{(e^{2i\lambda\varphi}/\sin\varphi)\log|\hat{\eta}^0 + (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0| + (e^{-2i\lambda\varphi}/\sin\varphi)\log|\hat{\eta}^0 - (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0|\}. \quad (4.38)$$

For (4.23), we have (outside $\hat{\xi}^\mu = 0$, $\forall\, \mu$ and $u_1^2 = 4$)

$$(P^\mu)\delta(u_2)T(u_1, Q) = -2\delta(u_2)T'_Q(u_1, \varphi)\hat{\eta}^\mu,$$

$$(W^\mu)\delta(u_2)T(u_1, Q) = \delta(u_2)\{2u_1QT''_{Q^2}(u_1, Q) + 3u_1T'_Q(u_1, Q) + (u_1^2 - 4)T''_{u_1Q}(u_1, Q)\}\hat{\xi}^\mu.$$

The vectors $\hat{\eta}$ and $\hat{\xi}$ are proportional if and only if $Q = 0$, where

$$\hat{\eta}^\mu = 2(\hat{\eta}^0)\epsilon(\hat{\xi}^0)(4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^\mu,$$

and so $T'_Q(u_1, Q)$ must vanish for $Q \neq 0$ and solutions must be linear combinations of

$$T_\pm = \tfrac{1}{2}\delta(u_2)\theta(4 - u_1^2)\beta_\pm(u_1)\{\theta(Q)\epsilon(\hat{\xi}^0)^{\pm}\theta(-Q)\epsilon(\hat{\eta}^0)\}.$$

Then

$$(P^\mu)T_\pm = 2\delta(u_2)\theta(4 - u_1^2)\delta(Q)\beta_\pm(u_1) \times \{\theta(\hat{\xi}^0)\theta(\mp\hat{\eta}^0) - \theta(-\hat{\xi}^0)\theta(\pm\hat{\eta}^0)\}\hat{\eta}^\mu,$$

$$(W^\mu)T_\pm = -\delta(u_2)\theta(4 - u_1^2)\delta(Q)\{u_1\beta_\pm(u_1) + (u_1^1 - 4)\beta'_\pm(u_1)\} \times \{\theta(\hat{\xi}^0)\theta(\mp\hat{\eta}^0) - \theta(-\hat{\xi}^0)\theta(\pm\hat{\eta}^0)\}\hat{\xi}^\mu.$$

(4.34c) gives

$$u_1\beta_\pm(u_1) - (4 - u_1^2)\beta'_\pm(u_1) = \pm 2i\lambda(4 - u_1^2)^{\frac{1}{2}}\beta_\pm(u_1),$$

with the solutions

$$\beta_\pm(u_1) = \beta_\pm e^{\pm 2i\lambda\varphi}/\sin\varphi,$$

where $\beta_\pm$ are any constant. There is only one global solution of (4.34c):

$$T_\lambda = \delta(u_2)\theta(4 - u_1^2)\{(\sin 2\lambda\varphi/\sin\varphi)\theta(Q)\epsilon(\hat{\xi}^0) - i(\cos 2\lambda\varphi/\sin\varphi)\theta(-Q)\epsilon(\hat{\eta}^0)\}.$$

The only solutions of (4.34c) of the form (4.36) or (4.37) are

$$Y_\pm = \delta(u_2)\{\delta(u_1 - 2) \pm \delta(u_1 + 2)\}(-Q)^{\frac{1}{2}}\epsilon(\hat{\eta}^0{}_\alpha)$$
$$= \delta(u_2)\{\delta(u_1 + 2) \pm \delta(u_1 + 2)\}\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}^0{}_\alpha)^{\frac{1}{2}},$$

which satisfy $(P^\mu)Y_\pm = (W^\mu)Y_\pm = 0$, mod Ker $\mathfrak{D}_0$.

4. *Characters of the Unitary Representations, $m^2 = 0$, $w^2 = 0$*

First, the characters must have definite symmetry property under translations by $(0, -1) \in \bar{P}$. The solutions of the differential equations which have such a property are

S_λ and T_λ if 2λ is a real integer [with "parity" $(-1)^{2\lambda}$],

$Y_\pm$ (with "parity" ± 1).

For 2λ integer we have a 3-dimensional vector space of solutions spanned by S_λ, T_λ, and $Y_{(-1)^{2\lambda}}$, in which we have to find the characters of the two known unitary representations. They are of the form

$$\chi_\lambda^\pm = S_\lambda \pm xT_\lambda + iyY_{(-1)^{2\lambda}}, \quad (4.39)$$

where x and y (real because of Hermitian symmetry) must be such that $\chi_\lambda^\pm$ is extremal of positive type.

Before going on, we must note a property of the space $\mathfrak{D}_0$: The linear span of $\mathfrak{D}_0^+$ (the conus of functions of positive type in $\mathfrak{D}_0$) is not dense in $\mathfrak{D}_0$; the set of positive linear combinations of functions of the form $\alpha * \tilde{\alpha}$, $\alpha \in \mathfrak{D}$, is dense in $\mathfrak{D}^+$, so that any $\varphi \in \mathfrak{D}_0^+$ can be approximated by such functions. Now, if $\varphi = \alpha * \tilde{\alpha}$ ($\alpha \in \mathfrak{D}$), $\varphi \in \mathfrak{D}_0$ implies $\alpha \in \mathfrak{D}_0$, and so

$$\int \varphi(x, \Lambda)a_\mu x^\mu\, dx = 0 \quad \forall\, \Lambda \in \bar{\mathfrak{L}}_0$$

for any linear function $a_\mu x^\mu$; thus, any distribution linear with respect to $\mathbf{x}$ vanishes on $\mathfrak{D}_0^+$ and also on the linear span of $\mathfrak{D}_0^+$, but not on $\mathfrak{D}_0$. (The linear span of $\mathfrak{D}_0^+$ is the closure of $\mathfrak{D}_0^2 = \{\varphi = \alpha * \beta,\ \alpha, \beta \in \mathfrak{D}_0\}$.)

Distributions $Y_\pm$ are linear in $\mathbf{x}$ and so vanish on $\mathfrak{D}_0^+$. The fact that (4.39) is of positive type or not does not depend on the value of y: We have defined the character on too small a space to determine it completely by the properties given in Sec. 1. On the other hand, the calculation of x can theoretically be done, but, as in the case $m \neq 0$, we were not able to perform it. Thus, we can only write

$$\chi_\lambda^\pm = \delta(u_2)\theta(4 - u_1^2)\Big[-\Big(\frac{e^{2i\lambda\varphi}}{\sin\varphi}\log|\hat{\eta}^0 + (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0| + \frac{e^{-2i\lambda\varphi}}{\sin\varphi}\log|\hat{\eta}^0 - (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0|\Big) \pm x\Big(\frac{\sin 2\lambda\varphi}{\sin\varphi}\theta(+Q)\epsilon(\hat{\xi}^0) - i\frac{\cos 2\lambda\varphi}{\sin\varphi}\theta(-Q)\epsilon(\hat{\eta}^0)\Big)\Big] + iy\delta(u_2)\{\delta(u_1 - 2) + (-1)^{2\lambda}\delta(u_1 + 2)\}\frac{\hat{\eta}^0}{(-\hat{z}^{0\alpha}\hat{z}_\alpha^0)^{\frac{1}{2}}}. \tag{4.40}$$

The χ_λ are (with the right x and y) the characters of the helicity representations $[0, \lambda, \pm]$ of $\bar{P}$.

(We note that S_λ and, therefore, $\chi_\lambda^\pm$ is not bounded, but an element of $\mathfrak{D}_0'$ can be of positive type and not bounded.)

5. *The Remaining Solutions*

For the sake of completeness, we list the other solutions, which might be related to the characters of some nonunitary representations.

1. We have seen that

$$Y_\pm = \delta(u_2)[\delta(u_1 + 2) \pm \delta(u_1 - 2)]\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}_\alpha^0)^{\frac{1}{2}}$$

are solutions of (4.34a) to (4.34e); moreover, they vanish on $\mathfrak{D}_0^\pm$ and so are, in the limit, "of positive type." It is only because there are no other unitary representations of $\bar{P}$ than the helicity ones that we can assert that $Y_\pm$ are at most the restriction to $\mathfrak{D}_0'$ of the character of some nonunitary representation.

2. Last, we have solutions of (4.34a) to (4.34d): For 2λ real integer,

$$\delta(u_2)\theta(4 - u_1^2)\Big(\frac{e^{-2i\lambda\varphi}}{\sin\varphi}\log|\hat{\eta}^0 + (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0| - \frac{e^{2i\lambda\varphi}}{\sin\varphi}\log|\hat{\eta}^0 - (4 - u_1^2)^{\frac{1}{2}}\hat{\xi}^0|\Big)$$

and

$$\delta(u_2)\theta(4 - u_1^2)\Big(\frac{\cos 2\lambda\varphi}{\sin\varphi}\theta(Q)\epsilon(\hat{\xi}^0) + i\frac{\sin 2\lambda\varphi}{\sin\varphi}\theta(-Q)\epsilon(\hat{\eta}^0)\Big),$$

$$\delta(u_2)[\delta(u_1 + 2) \pm \delta(u_1 - 2)](a_\pm \log|Q| + b_\pm\epsilon(\hat{\eta}^0));$$

with f even or odd,

$$\delta(u_2)Qf(u_1).$$

6. *Frobenius' Formula*

The same remarks as for the characters of the representations with $m^2 \neq 0$ can be made here. The character of the inducing representation [the little group is again $R^4 \times E(2)$, the inducing representation being now trivial for the translation of $E(2)$] is

$$e^{-2i\lambda\varphi}e^{i\mathbf{p}\cdot\mathbf{x}},$$

with $\mathbf{p}^2 = 0$, $\mathbf{x} \in R^4$.

As a conclusion to this paragraph, we mention the help we found in studying first the characters of the 0-mass representations of the group $R \cdot R^2$, where R acts on the 2-dimensional translation group R^2 through the matrices

$$\begin{pmatrix}\cosh\eta & \sinh\eta\\ \sinh\eta & \cosh\eta\end{pmatrix}.$$

For this group, the characters can be obtained as well by means of the usual integral-kernel method as by our differential procedure. They possess a structure very similar to (4.40).

F. Characters of $SL(2, \mathbb{C})$

From the central distributions $T(u_1, u_2)$ which do not depend on $\mathbf{x}$, we can extract the characters of the representations of $SL(2, C)$ (which are also representations of $\bar{P}$). They are the solutions of the eigenequations

$$A\chi = a\chi, \tag{4.41}$$

$$B\chi = b\chi, \tag{4.42}$$

where A and B are the Laplace operators of $SL(2, \mathbb{C})$

$$A = \tfrac{1}{2}M^{\mu\nu}M_{\mu\nu}, \quad B = \tfrac{1}{4}\epsilon^{\lambda\mu\nu\rho}M_{\lambda\mu}M_{\nu\rho},$$

$M^{\alpha\beta}$ being the differential operator given in (4.1), in which the derivatives with respect to $\mathbf{x}$ now vanish.

With $u = u_1 + iu_2$ and $a_{1,2} = a \pm ib$, Eqs. (4.41) and (4.42) take the form

$$(u^2 - 4)\frac{\partial^2\chi}{\partial u^2} + 3u\frac{\partial\chi}{\partial u} + 4q_1\chi = 0,$$

$$(\bar{u}^2 - 4)\frac{\partial^2\chi}{\partial\bar{u}^2} + 3\bar{u}\frac{\partial\chi}{\partial\bar{u}} + 4q_2\chi = 0. \tag{4.43}$$

Now, if we set $u = 2\cosh(\eta + i\varphi)$, $\eta \geq 0$, and $0 \leq \varphi < 2\pi$, we find that two independent global solutions of (4.43) are [with $\sigma_i = (1 - 4q_i)^{\frac{1}{2}}$]

$$R = \sinh\sigma_1(\eta + i\varphi)\sinh\sigma_2(\eta - i\varphi)/|\sinh(\eta + i\varphi)|^2,$$

$$S = \cosh\sigma_1(\eta + i\varphi)\cosh\sigma_2(\eta - i\varphi)/|\sinh(\eta + i\varphi)|^2.$$

[The singular points of (4.41) and (4.42), $\eta = 0$, $\varphi = 0, \pi$, do not divide the set of the parameters into disconnected parts: This comes from the fact that $SL(2, \mathbb{C})$ is a semisimple complex group.]

Our usual "parity" requirement ($\varphi \to \varphi + \pi$) forces $\sigma_1 + \sigma_2$ or $\sigma_1 - \sigma_2$ to be an integer.

1. *One and only one of the numbers* $\sigma_1 + \sigma_2$ *and* $\sigma_1 - \sigma_2$ *is a real integer.*

σ_1 and σ_2 are only defined up to the sign. We choose them in such a way that

$$\sigma_1 + \sigma_2 = m > 0, \quad m \text{ an integer.}$$

(a) $m > 0$: σ_1 and σ_2 are now completely defined (with their sign); we put $\sigma_1 - \sigma_2 = i\rho \neq 0$ (ρ is a well-defined complex number). Then we have one (global) solution

$$\chi_{m,\rho} = \cos(\rho\eta + m\varphi)/|\sinh(\eta + i\varphi)|^2 \; (= S - R).$$

(b) $m = 0$: i.e., $q_1 = q_2$; we have chosen $\sigma_1 = -\sigma_2$, but the sign of σ_1 is arbitrary. We put $\sigma_1 - \sigma_2 = i\rho \neq 0$. ρ is a complex number defined up to the sign. We have one solution:

$$\chi_{0,\rho} = \cos \rho\eta/|\sinh(\eta + i\varphi)|^2.$$

We note that $\chi_{m,\rho}$ has parity $(-1)^m$.

Unitary representations

Unitary implies $q_1 = \bar{q}_2$.

$m > 0$: We have that $\sigma_1 = \bar{\sigma}_2$, $\rho \neq 0$ is any real number, and $\chi_{m,\rho}$ is the well-known character of the corresponding representation (m, ρ) of the principal series.

$m = 0$: $q_1 = q_2$, $\sigma_1 = -\sigma_2$.

If $1 - 4q_i < 0$, σ_i is pure imaginary. One can choose ρ real positive, and one has the character of the representation $(0, \rho)$ of the principal series.

If $1 - 4q_i > 0$, σ_i is real, $-i\rho$ can be chosen real positive.

If $0 < q_i < \frac{1}{4}$, $\chi_{0,\rho}$ is bounded, it is the character of the representation $(0, \rho)$ of the complementary series $(0 < -i\rho < 2)$.

If $q_i < 0$, $\chi_{0,\rho}$ is not bounded and is the character of a nonunitary representation.

2. *Both* $\sigma_1 + \sigma_2$ *and* $\sigma_1 - \sigma_2$ *are real integers.*

This condition implies that σ_1 and σ_2 are both integers or both half-integers.

(a) σ_1 *and* σ_2 *are half-integers:* One can choose σ_1 and σ_2 positive. $S + R$ and $S - R$ have opposite parity, and so we have the two solutions

$$\chi_+ = \cosh[(\sigma_1 + \sigma_2)\eta + i(\sigma_1 - \sigma_2)\varphi]/|\sinh(\eta + i\varphi)|^2 \quad (= S + R) \text{ parity } (-1)^{\sigma_1+\sigma_2+1},$$

$$\chi_- = \cosh[(\sigma_1 - \sigma_2)\eta + i(\sigma_1 + \sigma_2)\varphi]/|\sinh(\eta + i\varphi)|^2, \quad (= S - R) \text{ parity } (-1)^{\sigma_1+\sigma_2}.$$

χ_- is bounded for $\sigma_1 = \sigma_2$. We then have

$$\chi_- = \chi_{m,0},$$

where $\chi_{m,0}$ is the character of the representation $(m, 0)$ of the principal series (with $m = \sigma_1 + \sigma_2$, odd).

χ_+ is bounded for $\sigma_1 = \sigma_2 = \frac{1}{2}$ and we have $\chi_+ = \chi_{0,i}$, the character of a representation of the complementary series.

(b) σ_1 *and* σ_2 *are integers:* R and S are global solutions with the same parity $(-1)^m$, $m = \sigma_1 + \sigma_2$, so that there is a solution (R), which is a continuous function on G. It is the character

$$\chi^f_{j_1 j_2} = \sinh(2j_1 + 1)(\eta + i\varphi) \times \sinh(2j_2 + 1)(\eta - i\varphi)/|\sinh(\eta + i\varphi)|$$

of the finite-dimensional representation D_{j_1, j_2}:

$$\sigma_i = 2j_i + 1 \quad \text{with} \quad j_i = 0, \tfrac{1}{2}, 1 \cdots.$$

(If one of the σ's is zero, the corresponding R is zero.) We have $\chi^f_{00} = 1$, the character of the trivial representation which is the only finite-dimensional unitary representation!

On the other hand,

$$\chi^f_{\frac{1}{2},0} = 2\cosh(\eta + i\varphi) = u.$$

Now we have to find a second linear combination of R and S which is the character of the other (known) representation corresponding to the given values of q_1 and q_2.

If $\sigma_1 = \sigma_2$, there is a unique bounded linear combination of R and S, namely,

$$\chi_{m,0} = \cos(\sigma_1 + \sigma_2)\varphi/|\sinh(\eta + i\varphi)\eta^2|,$$

which is the character of the unitary representation $(m, 0)$ of the principal series ($m = \sigma_1 + \sigma_2$, even).

If $\sigma_1 \neq \sigma_2$, we have not yet found any argument to determine what linear combination of R and S is the character of the corresponding nonunitary representation.

ACKNOWLEDGMENTS

The authors are grateful to Professor J. Lascoux, who introduced them to the problems related with characters and gave them precious advices and criticisms. They also thank Professor H. Joos, Professor G. Rideau, and Dr. R. Schrader for helpful discussions.

APPENDIX A

Calculation of the limits of the distributions

$$\delta(u - 2 \pm \eta)\delta'(u - 2 \pm \eta), \quad \eta > 0.$$

(1) Limit of $\delta(u - 2 - \eta)$:

This distribution is defined by

$$\langle \delta(u - 2 - \eta), \varphi \rangle = \int \varphi \frac{dv\, dr}{|s|}\bigg|_{u=2+\eta} = \int_{C_\eta} \varphi \, d\omega_\eta, \quad \varphi \in \mathfrak{D}(G),$$

where the last integral is the sum of φ over the 1 - sheet hyperboloid C_η $(u = 2 + \eta)$, with its invariant measure $d\omega_\eta = dv\, dr/|s|$.

Now

$$\lim_{\eta\to+n}\int_{C_0}\varphi\,d\omega_\eta=\int_{C_0}\varphi\,d\omega_0 \qquad \text{(A1)}$$

[integral over the cone C_0 ($u = 2$) with the measure $d\omega_0 = dv\,dr/|s|$]. Therefore, the distribution $\delta(u - 2 - \eta)$ has a limit defined by (A1). We write

$$\lim_{\eta\to 0}\delta(u-2-\eta)=\delta(u-2).$$

(2) Limit of $\delta'(u - 2 - \eta)$:
Using the change of variables

$$r=\rho\cos\theta,\quad v=\rho\sin\theta,$$

we can write

$$\langle\delta'(u-2-\eta),\varphi\rangle = -\frac{d}{du}\int_{\rho>(u^2-4)^{\frac12}}\frac{\varphi_++\varphi_-}{(\rho^2-u^2+4)^{\frac12}}\rho\,d\rho\,d\theta\Big|_{u=2+\eta},$$

$$\varphi_+=\varphi|_{s>0},\quad \varphi_-=\varphi|_{s<0}.$$

We have

$$I_\pm=\int_{\rho>(u^2-4)^{\frac12}}\frac{\varphi_\pm}{(\rho^2-u^2+4)^{\frac12}}\rho\,d\rho\,d\theta$$
$$=\int d\theta[\varphi_\pm(\rho^2-u^2+4)^{\frac12}]|^{\infty}_{(u^2-4)^{\frac12}}-\int_{\rho>(u^2-4)^{\frac12}}\frac{\partial\varphi_\pm}{\partial\rho}(\rho^2-u^2+4)^{\frac12}\,d\rho\,d\theta.$$

The integrated part is equal to zero. Therefore,

$$\frac{d}{du}I_\pm=-\int_{\rho>(u^2-4)^{\frac12}}\frac{\partial^2\varphi_\pm}{\partial u\partial\rho}(\rho^2-u^2+4)^{\frac12}\,d\rho\,d\theta + u\int_{\rho>(u^2-4)^{\frac12}}\frac{\partial\varphi_\pm}{\partial\rho}\frac{d\rho\,d\theta}{(\rho^2-u^2+4)^{\frac12}}+A.$$

(A, the derivative of the lower bound, will not contribute by compensation.)

With

$$\frac{\partial\varphi}{\partial\rho}=\cos\theta\frac{\partial\varphi}{\partial r}+\sin\theta\frac{\partial\varphi}{\partial v}+\frac{\rho}{s}\frac{\partial\varphi}{\partial s},$$

$$\frac{\partial^2\varphi}{\partial u\partial\rho}=-\frac{u}{s}\left(\cos A\frac{\partial^2\varphi}{\partial r\partial s}+\sin\theta\frac{\partial\varphi^2}{\partial v\partial s}+\frac{\rho}{s}\frac{\partial^2\varphi}{\partial s^2}-\frac{\rho}{s^2}\frac{\partial\varphi}{\partial s}\right),$$

we have

$$\frac{d}{du}I_\pm=u\int_{\rho>(u^2-4)^{\frac12}}\left(\frac{\partial^2\varphi_\pm}{\partial r\partial s}\cos\theta+\frac{\partial^2\varphi_\pm}{\partial v\partial s}\sin\theta\right)d\rho\,d\theta + u\int_{\rho>(u^2-4)^{\frac12}}\frac{\partial^2\varphi_\pm}{\partial s^2}\frac{\rho\,d\rho\,d\theta}{(\rho^2-u^2+4)^{\frac12}} + u\int_{\rho>(u^2-4)^{\frac12}}\left(\frac{\partial\varphi_\pm}{\partial r}\cos\theta+\frac{\partial\varphi_\pm}{\partial v}\sin\theta\right)\times\frac{d\rho\,d\theta}{(\rho^2-u^2+4)^{\frac12}}.$$

The derivatives of φ with respect to v, r, and s are infinitely differentiable and therefore bounded, and have compact supports. The first two terms are then bounded by a constant A, independent of η, and the third is bounded by

$$B\int_{(4\eta)^{\frac12}}^{C}\frac{d\rho}{(\rho^2-4\eta)^{\frac12}}=B\cosh^{-1}\frac{C}{2\sqrt{\eta}},$$

where B and C are constants (C is an upper bound of ρ in the support of φ).

Therefore, though the distribution $\delta'(u - 2 - n)$ has no limit in $\mathfrak{D}'(G)$, the distribution $\eta^\alpha\delta'(u - 2 - \eta)$ has the limit zero for $\alpha > 0$:

$$\lim_{\eta\to0}\eta^\alpha\delta'(u-2-\eta)=0\quad\forall\alpha>0.$$

(3) Limit of $\delta(u - 2 + \eta)\theta(\pm s)$:
This distribution is defined by

$$\langle\delta(u-2+\eta)\theta(\pm s),\varphi\rangle=\int_{s\gtrless0}\varphi\frac{dv\,dr}{|s|}\Big|_{u=2-\eta}=\int_{C_{-\eta}^{\pm}}\varphi\,d\omega_{-\eta},$$

an integral over the upper or lower sheet of the hyperboloid $C_{-\eta}(u = 2 - \eta)$.

Now

$$\lim_{\eta\to0}\int_{C_{-\eta}^{\pm}}\varphi\,d\omega_{-\eta}=\int_{C_0^{\pm}}\varphi\,d\omega_0$$

integral over the upper or lower sheet of the cone C_0. We write

$$\lim_{\eta\to0}\delta(u-2-\eta)\theta(\pm s)=\delta(u-2)\theta(\pm s).$$

(4) Limit of $\delta'(u - 2 + \eta)\theta(\pm s)$:
With the previous notations,

$$\langle\delta'(u-2+\eta)\theta(\pm s),\varphi_0\rangle=-\frac{d}{du}\left(\int_{\rho>0}\frac{\varphi_\pm\cdot}{(\rho^2-u^2+4)^{\frac12}}\rho\,d\rho\,d\theta\right)\Big|_{u=2-\eta},$$

$$J_\pm=\int_{\rho>0}\frac{\varphi_+\rho.d\rho\,d\theta}{(\rho^2-u^2+4)^{\frac12}}$$
$$=\int d\theta[\varphi_\pm(\rho^2-u^2+4)^{\frac12}]|_0^\infty-\int_{\rho>0}\frac{\partial\varphi_\pm}{\partial\rho}(\rho^2-u^2+4)^{\frac12}\,d\varphi\,d\theta.$$

The first term is

$$\int d\theta[\varphi_\pm(\rho^2-u^2+4)^{\frac12}]|_0^\infty = -2\pi\varphi_\pm(u,\rho=0)(4-u^2)^{\frac12}.$$

Therefore, neglecting terms in $O(\sqrt{\eta})$, we have

$$-\frac{d}{du}[-2\pi\varphi_\pm(u,\rho=0)(4-u^2)^{\frac{1}{2}}]|_{u=2-\eta}$$

$$=4\pi(\sqrt{\eta})\frac{\partial\varphi_\pm}{\partial u}(2-\eta,0)-2\pi\varphi_\pm(2-\eta,0)\frac{1}{\sqrt{\eta}},$$

$$\simeq\mp4\pi\frac{\partial\varphi_\pm}{\partial s}(2-\eta,0)-2\pi\varphi\pm(2-\eta,0)\frac{1}{\sqrt{\eta}},$$

$$\frac{\partial\varphi}{\partial u}=-\frac{u}{s}\frac{\partial\varphi}{\partial s}.$$

The second term is developed as in (2), and is bounded by

$$A+B\int_0^C\frac{d\rho}{(\rho^2+4\eta)^{\frac{1}{2}}}=A+B\sinh^{-1}\frac{C}{2\sqrt{\eta}}.$$

Thus, $\eta^\alpha\, dJ_\pm/du|_{2-\eta}$ has a limit only if $\alpha\geq\frac{1}{2}$, and

$$\lim_{\eta\to0}(\sqrt{\eta})\frac{d}{du}J_\pm\bigg|_{u=2-\eta}=2\pi\varphi_\pm(0,0)=2\pi\varphi(2).$$

Therefore,

$$\lim_{\eta\to0}(\sqrt{\eta})\delta'(u-2+\eta)\theta(\pm s)=-2\pi\delta(v)\delta(r)\delta(s).$$

APPENDIX B

The basic results are, of course, the equations of the group manifold. Some others appear only on submanifolds. Finally, some are just consequences of the properties of the $\epsilon^{\lambda\mu\nu\rho}$ tensor.

On the whole group manifold, we have

$$\hat{z}^{\lambda\mu}\hat{z}_{\lambda\nu}=(u_1^2-u_2^2-4)g_\nu^\mu+z^{\rho\mu}z_{\rho\nu},$$
$$\hat{z}^{\lambda\mu}z_{\lambda\nu}=-u_1u_2g_\nu^\mu,$$
$$\tfrac{1}{2}\epsilon^{\lambda\mu\nu\rho}\hat{z}_{\nu\rho}=-z^{\lambda\mu},$$
$$(yzz\cdots zy)=0,$$

where the bracket represents a completely contracted quantity, y is any vector, and the number of z is odd. This gives, for instance,

$$x^\mu\xi_\mu=x^\mu\hat{\xi}_\mu=0.$$

Furthermore, on $u_2=0$ only, we have

$$z_\lambda^\alpha z_\mu^\beta\epsilon^{\lambda\mu\nu\rho}=z^{\alpha\pi}\hat{z}^{\nu\rho},$$
$$z^{\alpha\beta}\hat{z}_{\alpha\beta}=0,\quad \xi^\alpha\hat{\xi}_\alpha=0,$$
$$Q=\xi^2=\hat{\eta}^\mu\eta^\nu+(u_1^2-4)\xi^\mu\xi^\nu/\hat{z}^{\mu\alpha}\hat{z}_\alpha^\nu,\quad \forall\,\mu,\nu,$$
$$=\hat{\eta}^\mu\xi^\nu-\hat{\eta}^\nu\xi^\mu/\hat{z}^{\mu\nu},\quad \forall\,\mu\neq\nu.$$

APPENDIX C

We give here the general line we followed for the calculation of the limit of the distributions

$$\delta(u_2)F[(4.6)\text{–}(4.12)]\times(\mathbf{W}^2-w^2)[R,S(u_1)\theta(\pm(u_1^2-4)-\epsilon)],\quad\text{(C1)}$$

where the F are the $\mathbf{x}$ depending parts of distributions (4.10) (for $u_1^2-4>\epsilon$) or (4.6)–(4.9), (4.11), and (4.12) (for $4-u_1^2>\epsilon$).

(I) Parametrization of $SL(2,C)$.

Here, we introduce briefly another parametrization of $SL(2,\mathbb{C})$, which possesses no formal covariance, but leads us for our present purpose to easier calculations. We define the trivectors

$$a_k=\hat{z}_{0k},$$
$$b_k=z_{0k}.$$

The group manifold is then

$$u_1u_2+\vec{a}\vec{b}=0,$$
$$u_1^2-u_2^2+\vec{a}^2-\vec{b}^2-4=0,$$

and we get

$$\xi^0=\vec{a}\vec{x},\quad \hat{\eta}^0=\vec{x}\cdot(\vec{b}\wedge\vec{a})-\vec{a}^2x^0.$$

Now, outside the center of the group ($a=b=0$) and on $u_2=0$, we can take as new space coordinate system the unit vectors

$$\hat{a},\ \hat{b},\ \text{and}\ \hat{n}=\hat{a}\wedge\hat{b}.$$

We then have ($x,y,z=\hat{a}\vec{x},\hat{b}\vec{x},\hat{n}\vec{x}$):

$$Q=-[(ax^0+bz)^2+(b^2-a^2)x^2]\quad\text{and}\quad\lambda=x.\quad\text{(C2)}$$

(II) The distribution (C1) are defined by the integrals

$$I_{R,S}^\pm[(1.5)\text{ to }(1.11)]=\int\delta(u_2)F(1.5)\text{ to }(1.11)R,\ S(u_1)\theta(\pm(u_1^2-4)-\epsilon)\times(\mathbf{W}^2-w^2)\varphi(x^\lambda,\hat{z}^{\mu\nu})d^4xd^6\hat{z}/u_1,$$

where $\varphi(x^\lambda,z^{\mu\nu})\in\mathfrak{D}(\bar{P})$.

Using Eq. (4.15), of which R and S are solutions, by partial integration we find

$$I_{R,S}^{\pm,\epsilon}=\int\delta(u_2)F\,D_{R,S}^{\pm,\epsilon}(u_1)\varphi(x^\lambda,\hat{z}^{\mu\nu})d^4x\frac{d^6\hat{z}}{u_1},\quad\text{(C3)}$$

where the distribution $D_{R,S}^{\pm,\epsilon}$ is ($g=R$ or S)

$$D_{R,S}^{\pm,\epsilon}=\pm6\delta(u_1^2-4\mp\epsilon)g((4\pm\epsilon)^{\frac{1}{2}})+4\epsilon\delta(u_1^2-4\mp\epsilon)g'((4+\epsilon)^{\frac{1}{2}})+4\epsilon\delta'(u_1^2-4\mp\epsilon)g((4\pm\epsilon)^{\frac{1}{2}}).\quad\text{(C4)}$$

(In the following, all our open sets will exclude the point $u_1=-2$.)

With the parametrization defined above, we can also write

$$I^{\pm,\epsilon}_{R,S} = \int F D^{\pm,\epsilon}_{R,S} \times \varphi(x_0, x, y, z, a, b, \Omega_a, \varphi_{ab}, (b^2 - a^2 + 4)^{\frac{1}{2}})d^4x \times \frac{a\, da\, d\Omega_a}{2(b^2 - a^2 + 4)^{\frac{1}{2}}}\, b\, db\, d\varphi_{ab}.$$

(III) For $I^{+,\epsilon}$ $(u_1^2 - 4 = b^2 - a^2 > \epsilon)$, we define

$$(b^2 - a^2)^{\frac{1}{2}} X_3 = ax_0 + bz,$$

so that

$$Z = Q/(u_1^2 - 4) = -(x^2 + X_3^2).$$

(1) Now if we calculate (C3) for R, with $D_R^{+,\epsilon}$ replaced by $\delta(b^2 - a^2 - \epsilon)R(4 + \epsilon)^{\frac{1}{2}}$, we find that this integral has a limit in ϵ which is

$$L_1^R = \tfrac{1}{4}A(2)\int J_0[m^2(x^2 + X_3)]^{\frac{1}{2}} \times \varphi(x_0, x, y, -x_0, a, a, \Omega_a, \varphi_{ab}, 2) \times dx_0\, dx\, dy\, dX_3 a\, da\, d\Omega_a\, d\varphi_{ab}.$$

[Here, $F = J_0(m^2Z)^{\frac{1}{2}}$.]

The variable X_3 does not appear any more in φ, and so we can perform the corresponding integration with the result

$$L_1^R = \frac{1}{4m} A(2)\int \cos(-m^2x^2)^{\frac{1}{2}} \times \varphi(x_0, x, y, -x_0, a, a, \Omega_a, \varphi_{ab}, 2) \times dx_0\, dx\, dya\, da\, d\Omega_a\, d\varphi_{ab}.$$

With $L_1^R = \langle \mathfrak{L}_y^R, \varphi\rangle$, going back to covariant formulation, we see that

$$\mathfrak{L}_1^R = \tfrac{1}{2}\delta(u_2)\delta(u_1^2 - 4)\delta(\hat{\eta}^0/(-\hat{z}^{0\alpha}\hat{z}_\alpha^0)^{\frac{1}{2}}) \cos m\lambda A(u_1).$$

The complete calculation of $I_R^{+,\epsilon}$ gives

$$I_R^{+,0} = 2\langle \mathfrak{L}_1, \varphi\rangle;$$

i.e., our limit is central and an eigendistribution of $\mathbf{P}^2$ [see (4.13)], as it has to be.

(2) If we calculate (C3) similarly for S, with $D_S^{+,\epsilon}$ replaced by $\delta(b^2 - a^2 - \epsilon)S((4 + \epsilon)^{\frac{1}{2}})$, we find that the integral is divergent in the limit $\epsilon = 0$. Neglecting terms which go to 0 with ϵ, we find

$$L_1^S = L_1^R/\sqrt{\epsilon} + M_1,$$

where the integral M_1, which defines a distribution with support $a = b = 0$ [center of $SL(2, \mathbb{C})$], is in fact 0. The complete calculation of $I_S^{+,\epsilon}$ $(\delta' = -\partial\delta/\partial\epsilon)$ gives

$$I_S^{+,0} = 0.$$

(IV) For $I^{-,\epsilon}$ $(4 - u_1^2 = a^2 - b^2 > \epsilon)$, we define

$$(a^2 - b^2)^{\frac{1}{2}} t = ax_0 + bz,$$

so that

$$Z = t^2 - x^2, \quad \epsilon(\hat{\xi}^0) = \epsilon(x), \quad \text{and} \quad \epsilon(\hat{\eta}^0) = -\epsilon(t).$$

(1) For R, the calculation goes as in (III), each term of $D_R^{-,\epsilon}$ having separately a limit L_i^R, the support of which is

$$u_2 = u_1^2 - 4 = \hat{\eta}_0^0/(-\hat{z}_\alpha^0\hat{z}^{0\alpha})^{\frac{1}{2}} = 0.$$

Usual integrals of Bessel functions are needed to get results (2.4) and (2.5).

(2) For S, we again get developments of the form $L_i^S = L_i^R/\sqrt{\epsilon} + M_i$. The L_i^R are the same integrals which appear for R. We were not able to calculate the M_i (except in cases where they are 0 by symmetry), whose supports are again the center of $SL(2, \mathbb{C})$. We were not able to perform the four necessary integrations, though we know the result had to be of the form $\Delta(\mathbf{x}^2)$, solution of $\mathbf{P}^2 = m^2 = 0$.

[1] F. A. Berezin, *Laplace Operators on Semisimple Lie Groups* (American Mathematical Society, Providence, R.I., 1962), Translation Ser. 2, Vol. 21.

[2] Harish-Chandra, *The Characters of Semisimple Lie Groups* (American Mathematical Society, Providence, R.I., 1956), Translation Ser., Vol. 83.

[3] E. Nelson and W. F. Stinespring, Am. J. Math. **81**, 556 (1959).

[4] L. Schwartz, *Théorie des distributions, Tome II* (Hermann, Paris, 1957).

[5] J. Dixmier, *Les C*-algèbres et leurs représentations* (Gauthier-Villars, Paris, 1964).

[6] G. W. Mackey, "Infinite-Dimensional Group Representations," Colloquium at Stillwater, 1961.

[7] V. Bargmann, Ann. Math. **48**, 568 (1947).

[8] I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966).

[9] I. M. Gel'fand, Usp. Fiz. Nauk **76**, 29 (1962).

[10] G. W. Mackey, Ann. Phys. **55**, 101 (1952); **58**, 193 (1953).

[11] H. Joos and R. Schrader, "On the Primitive Characters of the Poincaré Group," in *Lectures in Theoretical Physics, Vol. VIIa: Lorentz Group*, W. E. Brittin and A. O. Barut, Eds. (Gordon and Breach, New York, 1964).

[12] R. Schrader, to be published.

Mathematics of the N/D Method

D. H. LYTH
University of Lancaster, Department of Physics, Lancaster, England

(Received 30 June 1969)

A complete mathematical account is given of the N/D method, subject to certain restrictions on the given functions L and ρ (left-hand cut contribution and phase-space factor). The restrictions are the weakest possible if the phase shift is to be Hölder continuous. Any asymptotic behavior $\rho \sim x^{\beta}$, $0 < \beta < 2$, is allowed. No analyticity is assumed for L. Exhaustive existence and uniqueness theorems are given: Any allowed function F $(= e^{i\delta} \sin \delta/\rho)$ possesses an N/D decomposition, and the integral equation for N acts in a space of L_2 functions; this equation satisfies the Fredholm alternative theorems [though its kernel has a continuous spectrum if $\delta(\infty)/\pi$ is not integral]; any L_2 solution of the equation yields an allowed function F; *all* allowed solutions may be obtained by varying the CDD parameters (which enter linearly); each solution has a uniqueness index κ; there will usually be a κ parameter infinity of solutions with each $\kappa \geq 0$, and none with $\kappa < 0$, but precise conditions on L and ρ are given in order that there will be a negative κ solution.

INTRODUCTION

This paper aims to give a complete mathematical account of the N/D method, subject to some mild restrictions on the functions involved. It will be shown in a later paper[1] how to use the results obtained in physics.

As is well known, the N/D method tries to find a function of the form

$$F(x) = e^{i\delta(x)} \sin \delta(x)/\rho(x),$$

with ρ given, which may be written in the form

$$F(x) = L(x) + R(x),$$

where L is given and R has to be analytic with only the physical cut (say, $1 < x < \infty$). In other words, one tries to predict F (or, alternatively, δ) in terms of L and ρ.

The most important respects in which the present treatment goes beyond previous ones[2–13] are the following:

(a) Results are established rigorously for functions ρ having asymptotic behavior $\rho \sim x^{\beta}$ with β in the interval $0 < \beta < 2$. In the case $\beta \geq 2$ (for which no direct solution is possible), it is explained how to solve the problem by transforming it to one of the above type. Previous treatments consider only special values of β (0, $\frac{1}{2}$, or 1) (except for Ref. 10 and part of Ref. 5, which treat only particular aspects of the problem).

(b) The conditions assumed for L are likewise extremely general, in fact, the most general compatible with the conditions assumed for ρ and δ. This feature is shared by Ref. 11.

(c) It is shown explicitly that it is enough to consider only those solutions of the integral equation for N which are in a certain L_2 space. This is essential if one is to have a tractable mathematical problem.

(d) The case where the integral equation has more than one L_2 solution (i.e., the homogeneous equation has L_2 solutions) is dealt with; no earlier work does this except for a brief treatment in Refs. 4 and 6.

(e) It is proved that all solutions of the N/D equations with the L_2 restriction actually do give solutions of the problem stated above; again, the only previous references doing this are Refs. 9 and 12, which treat very special cases.

(f) Actual existence theorems are given, the solutions being classified according to a uniqueness index. In this connection a totally new concept is introduced, which is called the "index of the problem."

To enable the contents of this paper to be easily surveyed, the main results have been collected in ten theorems. The less important results are labeled lemmas.

1. THE PROBLEM

The real functions $L(x)$ and $\rho(x)$, defined in the interval $1 < x < \infty$, are supposed to be given. It is required to find a (complex) function $F(x)$, defined in this interval, which satisfies the following two requirements:

(I) It may be written in the form

$$F(x) = L(x) + R_{+}(x), \tag{1}$$

where R_{+} is the limit, as x tends to the real axis from above, of a function $R(x)$ having the following properties: (a) R is a real analytic function in the complex plane cut along $1 < x < \infty$, except for a finite number of poles; (b) near the end points $x = 1$ and $x = \infty$ it has the bounds

$$R < \text{const}\,(x-1)^{-p}, \tag{2}$$

$$R < \text{const}\,x^{q}, \tag{3}$$

respectively, for some p and q.

(II) It satisfies the condition

$$\text{Im}\,[F^{-1}(x)] = -\rho(x). \tag{4}$$

Restrictions on L, ρ, and F

In order to make any progress, it is necessary to restrict the class of given functions L and ρ to be considered and also to restrict the class of allowed solutions.

In most earlier work, L is assumed to be real analytic function except for a cut along $-\infty < x < x_L$ with $x_L < 1$. Then, if L is sufficiently well behaved to satisfy a dispersion relation, it will be completely specified by its imaginary part on the cut; in that case, (1) may be replaced by

$$\text{Im}\,F(x) = \text{Im}\,L(x) \quad (\text{given}) \tag{5}$$

for $-\infty < x < x_L$. There is then a considerable symmetry between (4) and (5), and, from a purely mathematical point of view, the left- and right-hand cuts play similar roles, these roles being interchanged if one considers F^{-1} instead of F. This symmetry carries through to the N/D decomposition described below, and, instead of writing an integral equation for N and a dispersion relation for D, one can write an integral equation for D and a dispersion relation for N, as was indeed first done historically.[2]

Analyticity assumptions for L will not be used here, however, for the following reasons:

(i) Although from a physical point of view one certainly expects F to be analytic except for left- and right-hand cuts, the polynomial boundedness assumption necessary to ensure that L actually satisfies a dispersion relation is without foundation even from the most liberal S-matrix-plus-phenomenology viewpoint[14] (the polynomial boundedness assumption for R is all right, however, as will be seen in the next section, Theorems 3 and 4, *et seq.*).

(ii) If one wants to include inelasticity by the Frye–Warnock method,[3] one has to consider a modified amplitude which has no simple analyticity properties.[15–20]

(iii) The imaginary part of L on the left-hand cut is not a convenient function to consider because it cannot be directly related to F in the (physical) region $1 < x < \infty$, whereas it will be seen shortly (Theorems 3, 4) that L itself (evaluated for $1 < x < \infty$) can be so related.

(iv) There is not actually any mathematical advantage to be gained from exploiting analyticity of L. In order to make any progress in solving the problem, one has to abandon any symmetry between the left- and right-hand cuts, at least with the presently available N/D method.[16]

Instead of analyticity, we shall rely simply on Hölder continuity.[17] A variable $u = 1/x$ will be used, so that our functions are defined in the interval $0 < u < 1$. A function f is said to be Hölder continuous in the interval $a < u < b$ if

$$|f(u') - f(u)| < \text{const}\,|u' - u|^{\mu} \tag{6}$$

for all $a < u < b$ and $a < u' < b$ and some *Hölder index* $\mu > 0$.

If $a \neq 0$, (6) is equivalent to

$$|f(x') - f(x)| < \text{const}\,|x' - x|^{\mu} \tag{7}$$

(with a different constant), so that then there is no distinction between Hölder continuity in u and in x. However, if $a = 0$, it is not enough simply to redefine the constant; rather, (6) is equivalent to the two requirements

$$|f(x') - f(x)| < \frac{\text{const}}{(xx')^{\mu}}\,|x' - x|^{\mu} \tag{8}$$

and

$$|f(x) - f(\infty)| < \text{const}\,x^{-\mu}. \tag{9}$$

It is clear that Hölder continuity (H.c.) in every interval $a < u < b$, for every choice $0 < a < b < 1$, is not equivalent to Hölder continuity in $0 < u < 1$ because the constant necessary in (6) might go to infinity as $a \to 0$ or $b \to 1$. These distinctions will be expressed by phrases like "f is H.c. near and at $u = 0$" or "f is H.c. except near and at $u = 0$."

In using Hölder continuity, we are following Ref. 3 and, more closely, Ref. 10. It is useful for the following two basic reasons:

(i) If a quantity is Hölder continuous, then so is a dispersion integral over it; closely related to this is the fact that the kernel occurring in the integral equation for N transforms Hölder continuous functions into one another. These properties are not true of, say, differentiable functions.

(ii) Hölder continuity is a physically reasonable requirement on the various quantities arising, whereas they may, for example, fail to possess derivatives at points corresponding to s-wave 2-body thresholds.[10]

Armed with the concept of Hölder continuity, we can now give the conditions assumed for L, ρ, and F. Here and throughout the paper, μ will indicate some fixed number in the range $0 < \mu < 1$, the same for all the functions involved.

(*a*) *Conditions on* ρ*:*

(i) There exist real positive quantities α and β such that for x near $x = 1$ and $x = \infty$, respectively,[18]

$$\rho \sim (x - 1)^{\alpha}, \tag{10}$$

$$\rho \to x^{\beta}. \tag{11}$$

(ii) ρ is H.c. except near and at $u = 0$, $\rho/(x-1)^\alpha$ is H.c. near and at $u = 1$, and $x^{-\beta}\rho$ is H.c. near and at $u = 0$.

(iii) $\rho(x)$ is nonzero except at $u = 1$.

(*b*) *Conditions on L:* Define the integer b by

$$\beta = b + c, \quad 0 \leq c < 1. \tag{12}$$

Then, for some constants a_i and A,

$$L(x) = \sum_{i=1}^{b} a_i x^{-i} + A x^{-\beta} + \bar{L}(x) \tag{13}$$

or

$$L(x) = \sum_{i=1}^{b} a_i x^{-i} + A \log x/x^b + \bar{L}(x) \tag{14}$$

for $\beta \neq$ integer or $\beta =$ integer, respectively, where

$$\bar{L} < \text{const}\, x^{-\beta-\mu}, \tag{15}$$

and $x^\beta \bar{L}$ is H.c. in $0 < u < 1$. (If $0 < \beta < 1$, then $b = 0$ and there is no summation.) In addition, A satisfies the inequality

$$A_{\min} < A < A_{\max}, \tag{16}$$

where, for $\beta \neq$ integer,

$$A_{\min} = -\tfrac{1}{2} \cot \tfrac{1}{2}\pi c, \tag{17}$$

$$A_{\max} = \tfrac{1}{2} \tan \tfrac{1}{2}\pi c \tag{18}$$

[c being defined by Eq. (12)] and, for $\beta =$ integer,

$$A_{\min} = 0, \tag{19}$$

$$A_{\max} = 1. \tag{20}$$

(*c*) *Conditions on F:* Because of condition (4), F may be expressed in terms of its phase δ by

$$F = e^{i\delta} \sin \delta/\rho. \tag{21}$$

The phase δ may be defined completely by specifying that

$$0 \leq \delta(1) < \pi \tag{22}$$

and that any jumps in δ consist of *increases* of less than π.

The conditions to be imposed on F are most conveniently expressed as conditions on δ. They are as follows:

(i) δ is H.c. in any interval not including the end points $u = 0$ and $u = 1$.

(ii) Near an end point (call it $u = a$), *either* δ is H.c. near and at $u = a$, or δ has the logarithmic behavior

$$\delta \sim 1/\log|u - a| \tag{23}$$

in such a way that

$$\tan \delta = f(u)/[\log|u - a| + g(u)], \tag{24}$$

f and g being H.c. near and at $u = a$. Solutions for which δ satisfied these conditions will be called *allowed solutions.*

This completes our catalog of the conditions on ρ, L, and F (or δ). The definitions of the quantities α, β, b, and A are important, and in what follows these symbols will often be used without explanation, denoting always these same quantities. They are, of course, fixed once ρ and L are fixed.

The conditions may look unduly involved, but in fact they are about the simplest that could be devised, as will now be explained.

First, look at the conditions on ρ. Physically this quantity will be equal to a kinematic factor times an inelasticity factor. The former will vanish at $x = 1$ (because one wants phase shifts which have the correct threshold behavior) and will diverge at $x = \infty$ (at least if one does not consider negative-angular momenta). The latter is finite (equal to 1) at $x = 1$ and is certainly nonzero at $x = \infty$; thus it does not alter the qualitative behavior. Hence, the limiting behavior assumed for ρ near $x = 1$ and $x = \infty$ is the simplest possible. It will become clear that we cannot get away without some smoothness requirement such as Hölder continuity.

Regarding δ, it is expected to approach finite limits at $x = 1$ and $x = \infty$, so that the simplest requirement to make is just that δ be H.c. for $0 < u < 1$. It is shown below that the rather elaborate conditions given for L essentially *follow* (Theorems 3 and 4); i.e., δ cannot be H.c. for $0 < u < 1$ unless L satisfies essentially these conditions.

Finally, the possibility that δ approaches its limits (at $x = 1$ and ∞) logarithmically has to be allowed because *in certain special cases* solutions having this behavior arise rather naturally, and it would be awkward to have to exclude them (Sec. 6). However, this only happens in special cases (namely, when α is integral or when β is integral and $A = 0$), and it must be emphasized that, generally speaking, the conditions assumed for L and ρ are such as to exclude such a logarithmic approach (Theorem 1).

2. A DISPERSION RELATION FOR R

The problem stated above is more commonly expressed as that of satisfying a dispersion relation, and the essential equivalence between the two formulations will now be established. Some other useful results will also be obtained.

Because of the behavior (10) of ρ, it is not obvious that F will be sufficiently well behaved near $x = 1$ (threshold) to allow R to satisfy a dispersion relation.

We start, therefore, with the following results. *Throughout this paper, except for Theorems 3 and 4, F will indicate an allowed solution of the problem and* ρ *and L will be assumed to satisfy the conditions above. Also, it is to be understood that* $u = x^{-1}$, *with "H.c." meaning "Hölder continuous in u."*

Lemma 1: Define the integer a by

$$\alpha = a + c, \quad 0 \le c < 1. \tag{25}$$

Then, the function

$$I(x) = \frac{1}{\pi}\int_1^\infty dx'\left(\frac{x'-1}{x'}\right)^a \frac{\operatorname{Im} F(x')}{x'-x} \tag{26}$$

satisfies the following conditions:

(a) It is real analytic in the complex plane cut along $1 < x < \infty$, and, as x tends to the cut from above, it approaches the limits

$$\operatorname{Re} I(x) = \frac{P}{\pi}\int_1^\infty dx'\left(\frac{x'-1}{x'}\right)^a \frac{\operatorname{Im} F(x')}{x'-x}, \tag{27}$$

$$\operatorname{Im} I(x) = \operatorname{Im} F(x). \tag{28}$$

(b) These limits are H.c. except, possibly, near the end points $u = 0$ and $u = 1$.

(c) Near $u = 0$ $(x = \infty)$: (i) If δ does not have the logarithmic behavior, then

$$I(x) = \sum_{i=1}^{b} c_i x^{-i} + \begin{Bmatrix}\cot \pi\delta(\infty) \\ \text{or} \quad \log x\end{Bmatrix} x^{-\beta} + \bar{I}(x), \tag{29}$$

where $x^\beta \bar{I}$ is H.c. near and at $u = 0$ and vanishes there (the two possibilities in the bracket are for $\beta \ne$ integer and $\beta =$ integer, respectively). (ii) If δ has the logarithmic behavior, then

$$I(x) = \sum_{i=1}^{b} c_i x^{-i} + \bar{I}, \tag{30}$$

where

$$\bar{I} = o(x^{-\beta}), \qquad \beta \ne \text{integer}, \tag{31}$$

or

$$\bar{I} \sim \log x\, x^{-\beta}, \quad \beta = \text{integer}. \tag{32}$$

(d) Near $u = 1$ $(x = 1)$: (i) If δ does not have the logarithmic behavior, then[18]

$$\bar{I} \to \cot \pi\delta(1)\cdot(x-1)^{a-\alpha}, \tag{33}$$

or, if it does have, then

$$I \sim \log(x-1)\cdot(x-1)^{a-\alpha}, \quad \alpha = \text{integer}, \tag{34}$$

or

$$I = o[(x-1)^{a-\alpha}], \qquad \alpha \ne \text{integer}. \tag{35}$$

The proof of this formidable statement is given in Appendix A (Lemmas A5, A8, and A9). Using it, we shall now prove the following lemma.

Lemma 2: The requirements on R under Part I of the problem (Sec. 1) are completely equivalent to the requirement that R satisfies

$$R(x) = \left(\frac{x}{x-1}\right)^a \frac{1}{\pi}\int_1^\infty dx'\left(\frac{x'-1}{x'}\right)^a \frac{\operatorname{Im} F(x')}{x'-x} + \sum_{i=1}^{a}\frac{c_i}{(x-1)^i} + \sum_i \frac{R_i}{x_i - x}, \tag{36}$$

where the second summation contains just the poles of R (including higher order ones, although this is not indicated explicitly) and the first summation is taken to be absent if $a = 0$.

Proof: The fact that (36) implies the requirements under Part I of the problem (Sec. 1) follows immediately from the previous lemma. To prove the converse, consider the difference Δ between the left-hand and right-hand sides of (36). By Lemma 1, Δ is real analytic in the plane cut along $1 < x < \infty$ and has zero imaginary part there. Hence,[19] it can only have singularities at $x = 1$ and $x = \infty$. But the bounds (2) and (3) on R, also (c) and (d) of Lemma 1, imply that the singularities here are at worst poles, not essential singularities. On the other hand, (13), (14), (15), (21), (10), and (11) substituted into (1) imply that on the *real axis* $(1 < x < \infty)$ R cannot even go to infinity as fast as a first-order pole near $x = 1$ and must vanish at $x = \infty$; also, by (c) and (d) of Lemma 1, the integral cannot go like a pole of order higher than a near $x = 1$, and it too vanishes at $x = \infty$. Hence, by a suitable choice of the constants c_i, Δ can be a function which has no singularities anywhere and vanishes at infinity; thus, it is identically zero. This ends the proof.

If (36) is substituted into (1), powerful restrictions on F and L are implied. Some of these are summarized in the following important theorem.

Theorem 1: Near the points $x = 1$ and $x = \infty$, δ has the following properties:

(a) Near $x = \infty$, δ cannot have the logarithmic behavior, unless both $\beta =$ integer and $A = 0$.

(b) There is the following connection between β, A, and $\delta(\infty)$:

$$A = \sin^2\delta(\infty)[\cot\delta(\infty) - \cot\pi\beta], \quad \beta \ne \text{integer}, \tag{37}$$

$$A = \sin^2\delta(\infty), \quad \beta = \text{integer}. \tag{38}$$

(c) Near $x = 1$, δ cannot have the logarithmic behavior unless α = integer.

(d) If it does not, then either $\delta(1) = \pi(\alpha - a)$, or $\delta(1) = 0$, a being defined as in Lemma 1.

(e) If the latter, then either

$$\delta \sim (x - 1)^{\alpha - m} \tag{39}$$

for some integer m, $0 < m < \alpha$, or else

$$\delta = O([x - 1]^{\alpha}). \tag{40}$$

Proof: Parts (a), (b), (c), and (d) follow directly upon requiring that the left-hand and right-hand sides of (1) balance, bearing in mind Lemmas 1 and 2. The same is true of case (e) if $0 < \alpha < 1$ (when, of course, the only possibility is $m = 0$, and $a = 0$ in Lemma 1). For the case $\alpha \geq 1$, (e) is proved in Appendix B by an iteration procedure. The idea is to first assume only that δ is H.c., then to show that (1) implies a more restrictive behavior, then to assume this behavior, and show that (1) implies a more restrictive behavior still, and so on, until the result required is obtained.

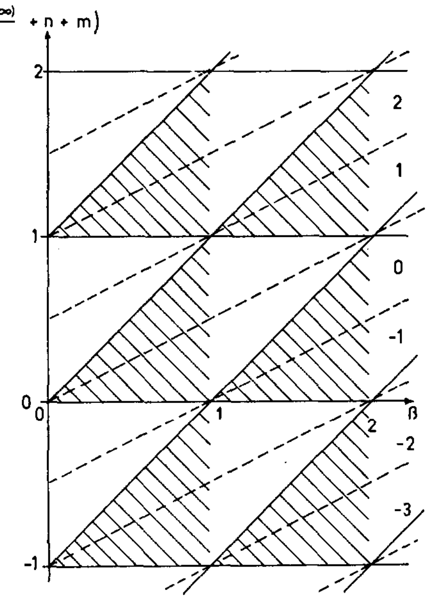


FIG. 1. The uniqueness index κ. Full lines are $A = 0$, dotted lines are $A = A_{\max}$ or $A = A_{\min}$ [Eqs. (37) and (38)]. Cross-hatched regions are $A > 0$. The numbers on the right are the values of κ in the strips bounded by the dotted lines [Eqs. (51) and (52) *et seq.*].

Discussion: Various remarks about Theorem 1 are in order. Parts (a) and (c) tell us that, generally speaking, δ has to be H.c. even near and at the end points; it is only exceptionally that a logarithmic approach can be tolerated.

Part (b) has motivated the restriction (16) on A. However, it tells us much more; namely, that once the given functions L and ρ are known, allowed solutions can have only two possible values for $\delta(\infty)$ (mod π), corresponding to the solutions of (37) or (38). For a given value of β, the allowed values δ_1 and δ_2 lie in the ranges (mod π)

$$\tfrac{1}{2}\beta - \tfrac{1}{2} \leq \delta_1/\pi \leq \tfrac{1}{2}\beta \tag{41}$$

and

$$\tfrac{1}{2}\beta - 1 \leq \delta_2/\pi \leq \tfrac{1}{2}\beta - \tfrac{1}{2}. \tag{42}$$

As A *increases* from its minimum to its maximum value, δ_1 *increases* through its range, and δ_2 *decreases* through its range when b is even, or vice versa, when b is odd. This situation is illustrated in Fig. 1, the other features of which will be explained later.

A special case of interest is when $A = 0$. In this case, the allowed values of $\delta(\infty)$ are given by

$$\delta_1/\pi = \text{integer} \tag{43}$$

and

$$\delta_2/\pi = \text{integer} + \beta. \tag{44}$$

It is clear that the behavior $\delta(\infty)/\pi$ = integer can only occur when $A = 0$. This is the behavior of δ for well-behaved potentials, and perhaps also for physical scattering amplitudes with a suitable choice of inelasticity.[11,14,20] For this reason, the case $A = 0$ is of special importance physically, and it will become clear that it is also the simplest case mathematically.

Part (d) of Theorem 1 is a consequence of the assumption that L is bounded near $x = 1$. From a purely mathematical point of view, it would have been more natural to allow L to have near $x = 1$ a term $\sim (x - 1)^{m-\alpha}$ [analogous to the expression (13) or (14) valid near $x \sim \infty$], and then any value of $\delta(1)$ would have been allowed; but, physically, one requires that $\delta(1) = 0$ and (consequently) that L is bounded near $x = 1$.

Part (e) of Theorem 1 suggests that F may actually be better behaved near $x = 1$ than the behavior (10) for ρ might suggest; in fact, from the proof of part (e) in Appendix B, the following improvement of Lemma 2 is obvious.

Theorem 2: Let us associate with every allowed solution F an integer m defined as follows (cf. Theorem 1): (i) If $\delta = O([x - 1]^{\alpha})$, let $m = 0$; (ii) if δ has a behavior of the type $\delta \sim (x - 1)^{\alpha - m}$, $0 < m < \alpha$, define m by this relation; (iii) if δ has the logarithmic behavior or is nonzero near $x = 1$, let $m = a$ where a is related to α by (25). Then, the requirements on R, under part (I) (Sec. 1) of the problem, are completely equivalent to the requirement that R satisfy

$$R(x) = \left(\frac{x}{x-1}\right)^m \frac{1}{\pi}\int_1^\infty dx' \left(\frac{x'-1}{x'}\right)^m \frac{\operatorname{Im} F(x')}{x'-x} + \sum_{i=1}^{m} \frac{C_i}{(x-1)^i} + \sum_i \frac{R_i}{x_i - x}, \quad (45)$$

where the second summation includes just the poles of R (including any higher-order ones) and the first summation is understood to be absent if $m = 0$.

Discussion: If $m = 0$, Theorem 2 tells us that part (I) of the problem (Sec. 1) could have been replaced simply by the requirement that F satisfy the dispersion relation

$$\operatorname{Re} F(x) = L(x) + \frac{P}{\pi}\int_1^\infty dx' \frac{\operatorname{Im} F(x')}{x'-x} + \sum_i \frac{R_i}{x_i - x}, \quad (46)$$

and the problem is more usually stated in this form. It will become clear that the case $m = 0$ is the usual one, the case $m > 0$ corresponding in a loose sense to F acquiring an mth order pole at threshold (though, of course, since $x = 1$ is a branch point, this statement is not strictly meaningful). The fact that m will usually be zero is, of course, commonly exploited in physical applications, where one imposes a physical requirement that $\delta = O((x-1)^{l+\frac{1}{2}})$ (l being the orbital angular momentum) by choosing a function ρ with $\alpha = l + \frac{1}{2}$.

However, from a mathematical point of view it greatly complicates matters to exclude cases where $m > 0$, and therefore they will not be excluded here.

To end this section, we give two results which are important in applications (see Ref. 1).

The first follows from Lemma A5 of Appendix A.

Theorem 3: Let ρ satisfy the conditions of Sec. 1, let δ be H.c. in $0 < u < 1$, and near $x = 1$ let $\delta = O([x-1]^\alpha)$ with $[\delta/(x-1)^\alpha]$ H.c. near and at $u = 1$.

Define

$$F = e^{i\delta} \sin \delta/\rho, \quad (47)$$

and define L_0 by

$$\operatorname{Re} F(x) = L_0(x) + \frac{P}{\pi}\int_1^\infty dx' \frac{\operatorname{Im} F(x')}{x'-x}. \quad (48)$$

Then L_0 satisfies the conditions (b) of Sec. 1, and F is an allowed solution of the problem with $L = L_0$.

The second result is obvious from the proof of Lemma 2.

Theorem 4: Under the conditions of Theorem 3, the most general possible choice of L for which F will be an allowed solution of the problem is

$$L = L_0 - \sum_i \frac{R_i}{x_i - x} + \Phi(x), \quad (49)$$

where the summation includes an arbitrary number of poles (of any order) not on the interval $1 \le x \le \infty$ and Φ is a function whose only singularities are at $x = 1$ and $x = \infty$. A "canonical" choice $\Phi = 0$ follows if R is required to satisfy the dispersion relation

$$R(x) = \frac{1}{\pi}\int_1^\infty dx' \frac{\operatorname{Im} F(x')}{x'-x} + \sum_i \frac{R_i}{x_i - x}. \quad (50)$$

Discussion: In physical applications, one assumes that L has no singularities except for a left-hand cut; this then requires that any poles of F be included in R and that the function Φ is an entire function (no singularities except at infinity). *If* F actually satisfies an unsubtracted dispersion relation, then the choice $\Phi = 0$ may be motivated by the requirement that L and R correspond respectively to the left-hand and right-hand cut contributions to F. However, as there is no justification for such an assumption,[14] it is best to require only that R satisfies the dispersion relation (which Theorem 2 tells us is always possible provided that F has the appropriate continuity properties in the physical region $1 < x < \infty$).

3. UNIQUENESS THEOREMS

Before beginning an investigation of the problem of Sec. 1 using the N/D method, we establish two results using a simpler technique (Ref. 3). The first result will be a useful preliminary to the N/D investigation; the second is not strictly relevant for the present paper, but is given because it is such an easy and interesting extension of the first. First, we need a definition.

Definition 1: Associate with each allowed solution of the problem an integer κ, called the *uniqueness index* by: (i) If δ does not have the logarithmic behavior near $x = \infty$, then

$$2[\delta(\infty)/\pi + n + m] - \beta + 1 = \kappa + c, \quad 0 < c \le 1; \quad (51)$$

(ii) if δ does have the logarithmic near $x = \infty$, then

$$2\,[\delta(\infty)/\pi + n + m] - \beta + 1 = \kappa. \quad (52)$$

[Note that, in this second case, part (a) of Theorem 1 assures us that κ is still an integer.] Here n is the number of poles of R, counting each pole ξ times where ξ is its order, and m is defined as in Theorem 2.

Discussion: The appearance of the combination $(n + m)$ is related to the remark (following Theorem 2) that m is the order of a pole of F at the point $x = 1$.

It is clear from (37) or (38) that, once A and ρ are fixed, there is in general a one-to-one correspondence between κ and the quantity $(\delta(\infty)/\pi + n + m)$. The only case where the correspondence is not one-to-one is if $\beta =$ integer and $A = 0$; in this case, κ is equal to $(\delta(\infty)/\pi + n + m) - \beta + 1$ if δ has the logarithmic behavior at infinity, otherwise it is equal to this quantity minus one.

In Fig. 1 the values of κ are indicated in various strips of the β vs $(\delta(\infty)/\pi + n + m)$ plane. Every allowed solution corresponds to a point in the interior of a strip, except if $\beta =$ integer and $A = 0$ when it corresponds to a point lying on the boundary between two strips; in that case, the point is deemed to lie in the strip immediately above or below it according to whether δ has the logarithmic behavior at infinity or not, as has just been explained.

Note that the exclusion of points on the boundary arises because of the strict inequality (16) on A. The exclusion of the case where A achieves its upper or lower bound was, in fact, made partly to simplify the definition of κ.

The results to be proved will now be stated. It is understood that in all cases L and ρ are given, and they satisfy the conditions above.

Lemma 3: The problem has at most one allowed solution with $\kappa \leq 0$ and at most κ-parameter infinity of allowed solutions with each $\kappa > 0$. If it has an allowed solution with $\kappa = -r \leq -1$, then all other allowed solutions have $\kappa > +r$.

Theorem 5: Define a *pole-free* solution as an allowed solution for which $n = m = 0$ and a *ghost-free* solution as an allowed solution for which $m = 0$ and all the poles of R are simple and have real positions and positive residues. Then, if there is a pole-free solution with $\kappa \leq -1$, there is no other pole-free solution whatever; and if there is a ghost-free solution with $\kappa \leq -2$, there is no other ghost-free solution whatever.

Proof of Lemma 3 and Theorem 5: Let F_1 and F_2 be two different allowed solutions. Consider the function

$$P(x) \equiv \exp\left(-\frac{1}{\pi}\int_1^\infty dx'\,\frac{x}{x'}\,\frac{\delta_1(x') + \delta_2(x')}{x' - x}\right) \times \pi_i(x - x_{i1})^{\xi_{i1}}\pi_i(x - x_{i2})^{\xi_{i2}} \times (x-1)^{m_1+m_2} \times \Delta(x), \tag{53}$$

where

$$\Delta \equiv R_1 - R_2. \tag{54}$$

Here the subscripts 1 and 2 refer to the two different solutions, x_i are the positions of the poles of R, and ξ_i are their orders. It will be shown that P is a polynomial. The argument is very similar to that following Lemma 1 above.

First, consider the exponential. It follows from Appendix A that the exponential has the following properties: (a) It is a real analytic function in the plane cut along $1 < x < \infty$, and its phase tends to the well-defined limit $-(\delta_1 + \delta_2)$ as x moves on to the cut from above. (b) Near $x = 1$ it has the bound

$$\exp \sim (x-1)^{-\epsilon}, \quad \text{all} \quad \epsilon > 0. \tag{55}$$

(c) Near $x = \infty$, it has the behavior

$$\exp \sim x^{\delta_1(\infty)/\pi + \delta_2(\infty)/\pi} \tag{56}$$

with an extra factor $\log x$, if either δ_1 or δ_2 has the logarithmic behavior. (d) It is bounded for all other values of x.

Next, consider the function Δ. From the requirements placed on R in defining the problem, this is a real analytic function in the plane cut along $1 \leq x \leq \infty$, except for a finite number of poles, and at the end points it is bounded by const $(x-1)^{-p}$ and const x^q, respectively (for some p and q). As x tends to the cut from above, it tends to the limit

$$(F_1 - F_2) = (e^{i\delta_1}\sin\delta_1 - e^{i\delta_2}\sin\delta_2)/\rho = e^{i(\delta_1+\delta_2)}\sin(\delta_1 - \delta_2)/\rho; \tag{57}$$

its phase just above the cut is therefore $+(\delta_1 + \delta_2)$.

The results of the last two paragraphs imply that P is analytic everywhere except for possible poles at $x = 1$ and $x = \infty$; but, from (55), (56), and (57), there cannot be even a first-order pole at $x = 1$; hence, P is a polynomial.

From (56) and (57), the order of P is not greater than k where

$$k = \sum_{i=1}^{2} [\delta_i(\infty)/\pi + n_i + m_i] - \beta. \tag{58}$$

In the case where $\delta_1^{(\infty)}/\pi$ and $\delta_2^{(\infty)}/\pi$ are both integers and where neither δ_1 nor δ_2 has the logarithmic behavior at $x = \infty$, one can make the vital improvement

$$k = \sum_{i=1}^{2} [\delta_i(\infty)/\pi + n_i + m_i] - \beta - \mu, \tag{59}$$

where μ (the Hölder index) comes from the vanishing of $\sin(\delta_1 - \delta_2)$ at infinity.

From the definition of κ, it is now easily verified that the requirement $k \geq 0$ leads in all cases to

$$\sum_{i=1}^{2} \kappa_i > 0. \quad (60)$$

[To get the strict inequality, one has to note that (59) applies whenever the equality applies in (51).] Hence, there is at most one allowed solution with $\kappa \leq 0$, and, if there is an allowed solution with $\kappa = -r < 0$, then all other allowed solutions have $\kappa > r$.

Next, observe that P must have at least l zeros [coming from the sine term in (57)] where

$$l = \delta_1(\infty)/\pi - \delta_2(\infty)/\pi - 1. \quad (61)$$

For a pole-free solution, this then implies that

$$\delta_1(\infty)/\pi + \delta_2(\infty)/\pi - \beta \geq \delta_1(\infty)/\pi - \delta_2(\infty)/\pi - 1,$$

i.e.,

$$2\delta_2(\infty)/\pi - \beta + 1 \geq 0, \quad (62)$$

the equality being forbidden if (59) holds.

Using the definition of κ, we see that it then follows that

$$\kappa_2 > -1. \quad (63)$$

(If one of the solutions has the logarithmic behavior near $x = \infty$, but not the other, then, to get the strict inequality, one has to choose the labeling so that it is F_2 which has the logarithmic behavior.) Hence, there cannot be two different pole-free solutions with $\kappa \leq -1$, which proves the first half of Theorem 2. To prove the second half, note that,[3] for a ghost-free solution, l may be increased by $|n_1 - n_2| - 1$.

It remains to prove the statement of Lemma 3 to the effect that there is at most a κ-parameter infinity of solutions with each $\kappa > 0$. To do this, we need only observe that no two solutions with the same value for κ can have the same values for the set of κ numbers $\{R'(0), R''(0), \cdots, R^{(\kappa)}(0)\}$; for, if they had, P would have to be of order at least κ, in which case the arguments leading to (60) would lead instead to $2\kappa > 2\kappa$.

Discussion: Lemma 3 is merely a weaker form of Theorem 10 below; it is proved at this stage only because it will be needed in the course of proving that theorem.

Theorem 5 tells one that, with some additional restrictions, the problem may sometimes have only one solution. It is usually applicable to physical amplitudes because these will, in fact, be ghost free, and even pole free if there are no bound states (in the channel under consideration), and will, moreover, have $\kappa \leq -1$, and even $\kappa \leq -2$, except for small angular momenta.[1] However, it will be seen (Theorem 10 below) that these negative values for κ are possible only because L and ρ satisfy special conditions (see also Ref. 15); as soon as one makes any approximations for L or ρ, the value of κ will generally go up[1] to $\kappa = 0$, and the theorem will become inapplicable. Thus, it is not possible to utilize the theorem in approximate calculations.

4. EXISTENCE OF N/D DECOMPOSITION

In this section, it is shown by explicit construction that any allowed solution F may be written in the form $F = N/D$, where N and D satisfy the "N/D equations" below. Besides serving the obvious purpose of assuring one that all solutions of the problem may be found by the N/D method, this result is also of direct use in physical applications.[1]

The treatment here is along fairly well-worn lines, but two points should be noted. First, we show that it is an L_2 solution of the (symmetrized) integral equation which is involved with a suitable N/D decomposition. Without some such assurance, one could not proceed to a discussion of the solution of this equation; even for very well-behaved kernels (e.g., of the Hilbert–Schmidt type), nothing whatever appears to be known to mathematicians about the solution of integral equations *in general*, but only within the framework of some suitable function space. The second point is that, if parameters appear in the equations, N and D depend *linearly* on them; this makes it possible to obtain far more complete results than in the older treatments, where the *position* of a CDD pole was involved. For further discussion of this point, see the end of this section. The main result is the following theorem.

Theorem 6: Let $b_1, b_2, \cdots$ be any sufficiently long sequence of real quantities satisfying

$$-\infty < b_i < 1, \quad b_i \neq 0, \quad b_i \neq b_j, \quad i \neq j. \quad (64)$$

Then F may be written in the form

$$F = N/D, \quad (65)$$

where $n \equiv (\rho/u)^{\frac{1}{2}} N$ is in $L_2(0, 1)$ and N and D satisfy one of the following pairs of equations:

$$\text{(a)}\quad D(x) = 1 - \frac{1}{\pi}\int_1^{\infty} dx' \frac{x}{x'}\frac{\rho(x')N(x')}{x'-x} + \sum_{i=1}^{\chi}\frac{x}{b_i}\frac{d_i}{b_i - x}, \tag{66}$$

$$N(x) = B(x) + \int_1^{\infty} dx' K(xx')N(x'), \tag{67}$$

where

$$B(x) = L(x) - \sum_{i=1}^{\chi}\left(L(x)\frac{x}{b_i}\frac{d_i}{b_i - x} + \frac{k_i}{b_i - x}\right), \tag{68}$$

$$K(xx') = \frac{x'L(x') - xL(x)}{x' - x}\frac{\rho(x')}{x'}, \tag{69}$$

$$\text{(b)}\quad D(x) = -\frac{1}{\pi}\int dx' \frac{x}{x'}\frac{\rho(x')N(x')}{x'-x} + \sum_{i=1}^{\chi}\frac{x}{b_i}\frac{d_i}{b_i - x}, \tag{70}$$

$$N(x) = B_0(x) + \int_1^{\infty} dx' K(xx')N(x'), \tag{71}$$

where

$$B_0(x) = -\sum_{i=1}^{\chi}\left(L(x)\frac{x}{b_i}\frac{d_i}{b_i - x} + \frac{k_i}{b_i - x}\right). \tag{72}$$

In these equations, χ is a positive integer and the d_i and k_i are real quantities. It is to be understood that the summations may be absent.

Comment: It will be convenient in what follows to define χ to be precisely the number of pole terms in the above equations, i.e., (a) never to have $d_i = k_i = 0$ for any i and (b) set $\chi = 0$ if there are no pole terms.

Definition 2: χ is the number of pole terms in the N/D equations.

Thus the $\chi = 0$ equations for the two cases are

$$\text{(a)}\qquad D(x) = 1 - \frac{1}{\pi}\int_1^{\infty} dx' \frac{x}{x'}\frac{\rho(x')N(x')}{x'-x}, \tag{73}$$

$$N(x) = L(x) + \frac{1}{\pi}\int_1^{\infty} dx' K(xx')N(x'), \tag{74}$$

$$\text{(b)}\qquad D(x) = -\frac{1}{\pi}\int_1^{\infty} dx' \frac{x}{x'}\frac{\rho(x')N(x')}{x'-x}, \tag{75}$$

$$N(x) = \frac{1}{\pi}\int dx' K(xx')N(x'). \tag{76}$$

Note that these last equations, and only these, are homogeneous in D and N.

Proof of Theorem 6: For any solution, define x_i, ξ_i, n, and m as in Sec. 3; i.e., x_i are the positions of the poles of R, n is their number counting each one ξ_i times where ξ_i is its order, and m is defined in Theorem 2. Then, define $\hat{D}(x)$ by

$$\hat{D}(x) = \exp\left(-\frac{1}{\pi}\int_1^{\infty} dx' \frac{x}{x'}\frac{\delta(x')}{x'-x}\right) \times (1-x)^m \pi_i \begin{cases}(1 - x/x_i)^{\xi_i} \\ \text{or}\quad (x_i - x)^{\xi_i},\end{cases} \tag{77}$$

where the first and second alternatives for the product apply if R has no pole at $x = 0$ or has one, respectively.

It follows from Appendix A that $\hat{D}$ has the following properties: (i) $\hat{D}(0) = 1$ or 0, for the two cases just mentioned respectively; (ii) $\hat{D}(x)$ is real analytic in the plane cut along $1 < x < \infty$; (iii) the phase of $\hat{D}$ on the upper side of the cut is $-\delta$; (iv) $\hat{D}$ is finite for all x except possibly $x = 1$ and $x = \infty$ and is nonzero except possibly at these points and $x = x_i$; (v) near $x = \infty$,

$$\hat{D} \sim x^{\delta(\infty)/\pi + n + m}, \tag{78}$$

unless δ has the logarithmic behavior near $x = \infty$ when there is an extra factor $\log x$; (vi) near $x = 1$,

$$\hat{D} = O((x-1)^m \log x). \tag{79}$$

Next, define a function $\hat{N}$, on the interval $1 < x < \infty$ only, by

$$F(x) = \hat{N}(x)/\hat{D}_+(x), \tag{80}$$

where $\hat{D}_+$ denotes the value of $\hat{D}$ on the upper side of the cut. From (iii) above, $\hat{N}$ is real, and hence, from (4),

$$\operatorname{Im}\hat{D}_+ = -\rho\hat{N}. \tag{81}$$

Now, define $N = \Phi\hat{N}$ and $D = \Phi\hat{D}$, where Φ is a rational function which has poles only at some number χ of the points b_i mentioned in the theorem and which satisfies, *if R has no pole at $x = 0$*,

$$\Phi(0) = 1 \quad\text{or}\quad 0. \tag{82}$$

(This restriction is not required if R has a pole at $x = 0$.)

By making Φ fall off fast enough as $x \to \infty$, we can convert (78) into

$$D(x) = O(x^{1-\epsilon}). \tag{83}$$

The other statements (i)—(vi) remain true for D, so that the Eq. (66) or (70) for D [according to whether $D(0) = 1$ or 0, respectively] follows from a straightforward application of Cauchy's theorem to the function D/x.

To derive the integral equation for N, consider the function

$$X(x) = R(x)D(x). \quad (84)$$

The value X_+ of X on the upper edge of the cut is, from (1) and (80),

$$X_+ = N - LD_+, \quad (85)$$

so that

$$\begin{aligned} \operatorname{Im} X_+ &= -L \operatorname{Im} D_+ \\ &= L\rho N. \end{aligned} \quad (86)$$

By a suitable choice of Φ, one can ensure that near $x = 1$ and $x = \infty$, $X = O((x-1)^{\epsilon-1})$ and $O(x^{-\epsilon})$, respectively, some $\epsilon > 0$, and then Cauchy's theorem gives the relation

$$X(x) = \frac{1}{\pi}\int_1^\infty dx' \frac{\rho(x')N(x')L(x')}{x'-x} + \sum_{i=1}^{\chi} \frac{k_i}{b_i - x} \quad (87)$$

(the zeros of D ensure that X does not have poles at the positions of the poles of R). Then, the integral equation (67) or (71) follows on substituting (87) and the equation for D into (85).

Finally, property (iv) of D and (21) ensure (i) that N is finite for $1 < x < \infty$ and (ii) that by a suitable choice of Φ the behavior of N near the end points can be made sufficiently good, so that

$$\begin{aligned} (\rho/u)^{\frac{1}{2}}N &= O(u^{\epsilon-\frac{1}{2}}), & u \to 0, \\ &= O((u-1)^{\epsilon-\frac{1}{2}}), & u \to 1. \end{aligned} \quad (88)$$

Hence, it follows that $(\rho/u)^{\frac{1}{2}}N$ is indeed in $L_2(0, 1)$.

Special Choices of the Points b_i: In general, the pole positions b_i will be thought of in this paper as being fixed once and for all at arbitrary finite values, in the same way as the normalization point for D is fixed at $x = 0$. If there is more than one allowed solution corresponding to a given value of χ (as is usually the case), then the coefficients k_i and d_i are to be thought of as varying with (i.e., parametrizing) the different solutions, but the b_i are to be thought of as being held fixed at their arbitrarily chosen values.

In earlier treatments a different view was adopted; the points b_i were chosen to be the positions of either poles of R or zeros of F, and different solutions corresponding to the same values of χ were parametrized by the numbers b_i, plus *one* of the sets d_i or k_i (or some linear combination of these). Such approaches, besides being awkward because of the nonlinear dependence on the numbers b_i, are less general because they only cover solutions with enough (χ) poles or zeros as the case may be. They will, however, be briefly mentioned here for completeness.

If the b_i are chosen to be the positions of some χ simple poles of R, it is clear from (77) that D ceases to have poles at the points b_i, i.e., $d_i = 0$ for all i. In that case, it is clear that the N/D equations for arbitrary χ may be obtained from the $\chi = 0$ equations by the replacement

$$L \to L - \sum_{i=1}^{\chi} \frac{k_i}{b_i - x}. \quad (89)$$

What has, in fact, been done is to transfer χ poles from R to L (see Theorem 4 and the following remarks) and to parametrize solutions by the positions and residues of these poles.

The other choice, of making the points b_i zeros of F, only makes sense if these points are in the interval $1 < x < \infty$ (unless one is willing to assume analyticity for L and F, which is not done here). Because zeros of F imply zeros of $\hat{N}$, this does not, in fact, cause any difficulty, and it is easy to verify that with this choice the equation for D is unchanged, but that the term B in the Eq. (67) for N becomes

$$B(x) = L(x) + \sum_{i=1}^{\chi} d_i \frac{b_i L(b_i) - xL(x)}{b_i - x}\frac{1}{b_i} \quad (90)$$

[the first term being absent in case (b)].

Because of a somewhat remote connection with Ref. 20, zeros of F which can be chosen to be the points b_i are called "CDD zeros," the resulting poles of D are called "CDD poles," and the ambiguity arising if one tries to solve the problem without knowing the numbers χ, d_i, and b_i is called the CDD ambiguity.

It is interesting to note that the N/D equations resulting from this second choice of b_i may be obtained for arbitrary χ from the $\chi = 0$ equations by the replacement

$$\rho(x) \to \rho(x) + \sum_{i=1}^{\chi} \frac{d_i}{N(b_i)} \delta(x - b_i). \quad (91)$$

This may be compared with the first choice, i.e., replacement (89), which, if L is analytic except for a left-hand cut, may be written

$$\operatorname{Im} L(x) \to \operatorname{Im} L(x) - \sum_{i=1}^{\chi} k_i \delta(x - b_i). \quad (92)$$

One is again reminded of the mathematical symmetry between left-hand and right-hand cuts, F and F^{-1}, N and D, etc., mentioned above [Eq. (5) *et seq.*].

5. THE INTEGRAL EQUATION

In this section the integral equation for N is studied for the case $0 < \beta < 2$. The situation is radically different for the two cases $0 < \beta < 2$ and $\beta \geq 2$, and it is seen in Sec. 8 that in the latter case one can expect solutions only for special choices of the inhomogeneous term. Only the case $0 < \beta < 2$ is considered for the next three sections.

In addition to this dependence on the value of β, there is also an essential difference between the cases $A = 0$ and $A \neq 0$, the former being much the simpler.

In order to discuss the equation, it is convenient to rewrite it so that the kernel is symmetric and also to work in the variable $u = 1/x$. This may be achieved by the transformations

$$n = (\rho/u)^{\frac{1}{2}}N, \tag{93}$$

$$b = (\rho/u)^{\frac{1}{2}}B, \tag{94}$$

$$\begin{aligned} k(u, u') &= K\left(\frac{\rho(u')}{u'}\right)^{\frac{1}{2}}\left(\frac{\rho(u)}{u}\right)^{\frac{1}{2}} \\ &= -\frac{1}{\pi}\frac{L(u')/u' - L(u)/u}{u' - u}[\rho(u')u'\rho(u)u]^{\frac{1}{2}}. \end{aligned} \tag{95}$$

Then the integral equation becomes

$$n(u) = b(u) + \int_0^1 du'k(uu')n(u'). \tag{96}$$

Because of Theorem 6 it is only necessary to consider solutions n which are in $L_2(0, 1)$, and it is clear from (68) and (72) that b is also in $L_2(0, 1)$. The integral equation may therefore be regarded as an operator equation in $L_2(0, 1)$ (which will simply be called L_2).

The function spaces $L_2(a, b)$ are dealt with exhaustively in the literature.[21,22] It will be convenient to establish the following notation:

(i) For any function f in L_2, its *norm* is

$$\|f\| \equiv \int_0^1 du\, |f(u)|^2. \tag{97}$$

(ii) For any two such functions, their *scalar product* is

$$(f, g) \equiv \int_0^1 du f^*(u)g(u). \tag{98}$$

(iii) For any operator k acting in L_2, its *bound* is

$$\|k\| = \sup_{\|f\|=1} \|kf\|. \tag{99}$$

(iv) A *bounded* operator is one whose bound is finite.

The following result will also be needed (besides some others which will be introduced as required).

(v) An operator transforms *every* L_2 function into another such function if and only if it is bounded.

If an operator H is sufficiently well behaved, then one of the following statements (called the *Fredholm alternatives*) is true for equations of the form

$$n = b + Hn.$$

(i) The equation $n = b + Hn$ has exactly one L_2 solution for every b in L_2. (Hence, the homogeneous equation $n = Hn$ has no nonzero L_2 solution.) From (v) above this is equivalent to saying that $(1 - H)^{-1}$ is a bounded operator in L_2.

(ii) The homogeneous equation $n = Hn$ has a finite number of linearly independent solutions n_i, and the equation $n = b + Hn$ has an L_2 solution if and only if the scalar products (n_i, b) vanish. In this case the most general L_2 solution is clearly $n = \bar{n} + \sum_i c_i n_i$, where $\bar{n}$ is any particular L_2 solution and the c_i are arbitrary numbers.

With these preliminaries, the main result of this section can now be stated in the following theorem.

Theorem 7: For $0 < \beta < 2$, Eq. (96) either has a unique L_2 solution for every choice of the parameters d_i and k_i, provided the points b_i are chosen in accordance with the restrictions of Theorem 6, or else it has an L_2 solution only for those choices satisfying a certain number r of linear conditions; in this second case, the homogeneous form of Eq. (96) has r linearly independently L_2 solutions (apart from the zero solution) so that Eq. (96) itself has $(r + 1)$ linearly independent L_2 solutions. The second case occurs only if the given functions L and ρ take on exceptional values, in the following sense: If this case occurs for some values $L = L_0$ and $\rho = \rho_0$ which give $k = k_0$ [Eq. (95)], it will not occur for any values which give $k = (1 + \epsilon)k_0$, with $|\epsilon|$ in some interval $0 < |\epsilon| < \Delta$.

Actually, this theorem is a corollary of the following more detailed statement, which brings out the difference between the cases $A = 0$ and $A \neq 0$.

Lemma 4: Let $0 < \beta < 2$, and consider the modified equation

$$n = b + \lambda kn, \tag{100}$$

where λ varies over the complex plane. Then, (i) if $A = 0$, the first Fredholm alternative holds for all λ [i.e., $(1 - \lambda k)^{-1}$ is a bounded L_2 operator], apart from certain isolated points on the real axis at which the second alternative holds. (ii) If $A \neq 0$, define $\lambda_0 = A_{\max}/A$ or $A_{\min}/A$ according to whether $0 < \beta \leq 1$ or $1 < \beta < 2$, respectively, and define the exceptional line to be $\lambda_0 < \lambda < \infty$ or $-\infty < \lambda < \lambda_0$ for the two cases (Figs. 2 and 3). Then, the first

Fredholm alternative holds for all λ not on the exceptional line, apart from certain isolated points on the real axis where the second alternative holds.

Discussion: The appearance of the quantity λ_0 may be understood as follows. A change $k \to \lambda k$ may be effected by making a change $L \to \lambda L$; the new L thus obtained will satisfy condition (16) only if λ lies in the range $(A_{\min}/A) < \lambda < (A_{\max}/A)$; one of the end points of this "allowed" range is λ_0 (Figs. 2 and 3). Of course, there is no hint at this stage of the distinction between the cases $0 < \beta \leq 1$ and $1 < \beta < 2$.

One sees by taking $\lambda = 1$ (Figs. 2 and 3) that Theorem 7 is indeed a corollary of Lemma 4.

Proof of Lemma 4: The proof of Lemma 4 is quite lengthy and will occupy the rest of this section, as well as Appendix C and D.

(i) Case $A = 0$: According to a well-known result (Ref. 21, pp. 579, 609), the lemma will follow immediately if it can be shown that, for some integer n, the nth power of the operator k is compact (Ref. 22, p. 206, where compactness is called "complete continuity"). This will be shown to be the case. The two facts upon which the proof will rest are: (a) An operator A is certainly compact if it is of the Hilbert–Schmidt type, i.e., if it may be expressed

$$Af \leftrightarrow \int_0^1 du' A(uu') f(u')$$

with

$$\int_0^1 du \int_0^1 du' \, |A(uu')|^2 < \infty, \tag{101}$$

(Ref. 22, p. 179); (b) an operator K is certainly compact if it may be written, for all sufficiently small ϵ,

$$K = A(\epsilon) + B(\epsilon), \tag{102}$$

where $\|B\| \to 0$ as $\epsilon \to 0$ and where A is compact for all sufficiently small ϵ (Ref. 22. p. 178).

To apply these results, consider the decomposition

$$k(uu') = a(uu') + b(uu'), \tag{103}$$

$$a(uu') \equiv \theta(u - \epsilon)\theta(u' - \epsilon)k(uu'), \tag{104}$$

$$b(uu') \equiv 1 - \theta(u - \epsilon)\theta(u' - \epsilon)k(uu'). \tag{105}$$

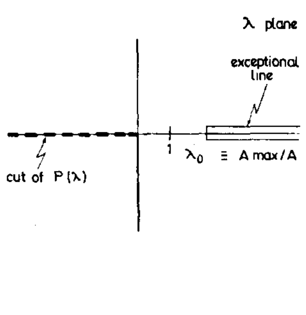
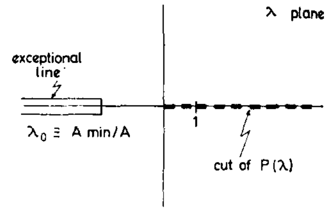


FIG. 2. The exceptional line of Lemma 4, and the cut of $P(\lambda)$ of Appendix D, for the case $0 < \beta \leq 1$.

FIG. 3. As Fig. 2, for the case $1 < \beta < 2$.



Clearly, b is zero unless either u or u' lies between 0 and ϵ. It will be shown that $\|b\| \to 0$ as $\epsilon \to 0$ and that the nth power of a is Hilbert–Schmidt for some integer n.

It is easily proved[23] using Schwart's inequality (Ref. 22, p. 41) that, for *any* positive function $S(u)$,

$$\begin{aligned} \|b\| &< \sup_{0<u<1} \int_0^1 du' \, |b(uu')| \frac{S(u)}{S(u')} \\ &\leq \sup_{0<u<\epsilon} \int_0^1 du' \, |k(uu')| \frac{S(u)}{S(u')} \\ &\quad + \sup_{\epsilon<u<1} \int_0^\epsilon du' \, |k(uu')| \frac{S(u)}{S(u')}. \end{aligned} \tag{106}$$

For the choice $S = u^{\frac{1}{2}|1-\beta|+\delta}$, with δ sufficiently small and positive, it is shown in Appendix C that *for* $0 < \beta < 2$ *and* $A = 0$ the right-hand side of (106) tends to zero as $\epsilon \to 0$.

To prove that a^n is Hilbert–Schmidt, use the bound

$$|k(uu')| < \frac{C(\epsilon)}{|u' - u|^\mu}, \quad \text{for} \quad \epsilon < u < 1, \epsilon < u' < 1, \tag{107}$$

which is valid for any $\epsilon > 0$ because of L being H.c. (except near and at $u = 0$). From this bound, it follows (Ref. 24) that for some n the quantity

$$\int_0^1 du_1 \int_0^1 du_2 \cdots \int_0^1 du_{n-1} a(uu_1) a(u_2 u_3) \cdots a(u_{n-1} u')$$

is a bounded function of u and u'. This proves that a^n is Hilbert–Schmidt.

Now we can apply the facts (a) and (b) above to the decomposition

$$k^n \equiv K = A(\epsilon) + B(\epsilon), \tag{108}$$

where

$$A \equiv a^n, \tag{109}$$

$$B \equiv a^{n-1}b + a^{n-2}ba + \cdots + b^n. \tag{110}$$

It has just been proved that A is Hilbert–Schmidt, and so it remains to show that $\|B\| \to 0$ as $\epsilon \to 0$. Since every term in the expression for $\|B\|$ is bounded by some product $\|a\|^{n-i} \|b\|^i, 1 \leq i < n$, and since $\|b\| \to 0$ as $\epsilon \to 0$, this will follow if it can be shown that $\|a\|$ is bounded for all (sufficiently small) ϵ. To show this, we first apply the analog of Eq. (106) to show that k is bounded,

$$\|k\| < \sup_{0<u<1} \int_0^1 du' \, |k(uu')| \left(\frac{u}{u'}\right)^{\frac{1}{2}|1-\beta|+\delta} < \text{const} \quad (111)$$

(from Appendix 2), and then write

$$\begin{aligned}\|a(\epsilon)\| &= \|k - b(\epsilon)\| \\ &\leq \|k\| + \|b(\epsilon)\| \\ &< \text{const}\end{aligned} \quad (112)$$

(for all sufficiently small ϵ).

(ii) $A \neq 0$: It has to be shown that $(1 - \lambda k)^{-1}$ is a bounded operator in L_2 throughout the λ plane, apart from the exceptional line plus some isolated points on the real axis where the second Fredholm alternative holds.

This will be done by splitting k up into two parts, a "singular" part being associated with the exceptional line and a "nonsingular" part which by itself would not involve the exceptional line.

Let us consider first the decomposition

$$k = k_s + k_0, \quad (113)$$

$$k_s \equiv -\frac{A}{\pi}\frac{u'^{-\gamma} - u^{-\gamma}}{u' - u}(u'u)^{\frac{1}{2}\gamma}, \quad (114)$$

or

$$k_s \equiv -\frac{A}{\pi}\frac{\log u' - \log u}{u' - u}, \quad (115)$$

for $\beta \neq 1$ and $\beta = 1$, respectively. Here

$$\gamma \equiv 1 - \beta. \quad (116)$$

From (95), k_s gives the leading term of k as u or u' tend to infinity. One has (assuming $0 < \beta < 2$)

$$k_0 = C + D, \quad (117)$$

$$C \equiv -\frac{1}{\pi}\frac{\bar{L}(u')/u' - \bar{L}(u)/u}{u' - u}\left(\frac{\rho(u')}{u'}\frac{\rho(u)}{u}\right)^{\frac{1}{2}}, \quad (118)$$

$$D \equiv \{[u'^{\beta}\rho(u')u^{\beta}\rho(u)]^{\frac{1}{2}} - 1\}k_s, \quad (119)$$

where $\bar{L}$ is defined by (13) or (14), being the remainder of L after its leading terms have been subtracted off.

It is first shown that $(k_0)^n$ is compact for some integer n. First, observe that the reasoning used above for the case of $A = 0$ applies equally well to C (since the term proportional to A has been removed), so that C^n is compact for some integer n. For D, one has from (11)

$$|D(uu')| < \text{const}\,(u'^{\frac{1}{2}\mu} + u^{\frac{1}{2}\mu})\,|k_s(uu')|. \quad (120)$$

We write, as in (103) above,

$$D = a(\epsilon) + b(\epsilon), \quad (121)$$

$$b = \theta(u - \epsilon)\theta(u' - \epsilon)D(uu'), \quad (122)$$

and it is clear that, for all ϵ in $0 < \epsilon < 1$,

$$|a(uu')| < \text{const} \quad (123)$$

and hence that a is Hilbert–Schmidt and so compact. Also, the same arguments as those following (103) above show that $\|b\| \to 0$ as $\epsilon \to 0$ (always keeping the restriction $0 < \beta < 2$, i.e., $-1 < \gamma < 1$). This proves that D is compact.

Now, we write

$$(k_0)^n = E + F, \quad (124)$$

$$E \equiv C^n, \quad (125)$$

$$F = C^{n-1}D + \cdots + D^n. \quad (126)$$

It has just been shown that E is compact, and the same is true of F because each of its terms is compact, being the product of a bounded operator and the compact operator D. Hence, $(k_0)^n$ is compact.

Now we consider k_s. In Refs. 5 and 7 a formal expression is given for the operator $(1 - \lambda k_s)^{-1}$,

$$(1 - \lambda k_s)^{-1}f \leftrightarrow f(u) + \int_0^1 du' R(u, u'; \lambda) f(u'), \quad (127)$$

where R is an analytic function of λ apart from a cut along the exceptional line. This leads us to conjecture that $(1 - \lambda k_s)^{-1}$ is a bounded operator in L_2, except for λ on the exceptional line. However, as the methods used to obtain (127) involve higher functions, it is difficult to prove this statement rigorously and, instead, a slightly different approach will be used which involves only a Fourier sine transform.[25] The method is described fully in Appendix D and is only outlined here in the text.

Instead of considering k_s, we define a new "singular" part by

$$k_s = g + \bar{k}_s \quad (128)$$

in such a way that g is compact and $\bar{k}_s$ can be diagonalized by a Fourier sine transform. Then, it is seen by direct inspection that $(1 - \lambda\bar{k}_s)^{-1}$ *is a bounded* L_2 *operator*. Since Fourier transforms in L_2 have been widely investigated,[26] it proves easy to justify the steps rigorously by quoting standard theorems.

Using this result, let us now consider instead of (113) the decomposition

$$k = \bar{k}_s + \bar{k}_0, \quad (129)$$

where

$$\bar{k}_0 = k_0 + g. \quad (130)$$

Since g is compact, $(\bar{k}_0)^n$ is compact and hence $(1 - \lambda\bar{k}_0)^{-1}$ *is a bounded operator in* L_2 *throughout the* λ *plane, apart from some isolated points where the second Fredholm alternative holds.*

It is therefore now very plausible that the sum $(\bar{k}_s + \bar{k}_0)$ behaves in the way asserted in the lemma. To prove this rigorously, observe that the equation

$$n = b + \lambda kn \tag{131}$$

is equivalent to

$$n = \hat{b} + T(\lambda)n, \tag{132}$$

where

$$\hat{b} \equiv (1 - \lambda\bar{k}_s)^{-1}b \tag{133}$$

and

$$T(\lambda) \equiv (1 - \lambda\bar{k}_s)^{-1}\lambda\bar{k}_0. \tag{134}$$

Clearly, for all λ not on the exceptional line, $\hat{b}$ is in L_2 whenever b is. Thus, we have to show that $[1 - T(\lambda)]^{-1}$ is a bounded operator in L_2 throughout the λ plane, apart from the exceptional line plus some isolated real points where the second Fredholm alternative holds for $T(\lambda)$.

This may be proved using the following (Ref. 21, p. 592).

Lemma 5: Let $S(\lambda)$ be an analytic operator-valued function defined on a connected domain D, and let S be compact for each λ in D. Then $(1 - S)^{-1}$ is either bounded for no λ in D or is bounded everywhere except at a countable number of isolated points.

To apply this, we take D to be the entire λ plane excluding the exceptional line and take $S = T^n$. The boundedness of $(1 - S)^{-1}$ will imply that of $(1 - T)^{-1}$ since

$$(1 - T)^{-1} = (1 - S)^{-1}(1 + T + \cdots + T^{n-1}). \tag{135}$$

That T (and therefore S) is analytic in D is a standard result (Ref. 21, p. 566, Lemma 2). Since $S(0) = 0$, $(1 - S)^{-1}$ is certainly bounded for $\lambda = 0$. Hence, we deduce that $(1 - T)^{-1}$ is bounded in the entire λ plane, apart from the exceptional line and a countable number of isolated points.

At these points the second Fredholm alternative must hold (since T^n is compact). Finally, since the original operator k is Hermitian, these points must be on the real axis.

6. SOLUTIONS OF THE PROBLEM

At this stage, it has been shown that allowed solutions of the problem, *if they exist*, may be found using the N/D equations, and further that one need only consider solutions of Eq. (96) which are in L_2 (Theorem 6). It has also been shown that there is one (and usually only one) L_2 solution of Eq. (96) for each choice of the parameters d_i and k_i (Theorem 7). In this section, it is shown that every L_2 solution of Eq. (96) will actually yield an allowed solution of the problem. Knowing this, one is assured that the N/D method finds those, and only those, functions $F\,(= N/D)$ which are allowed solutions of the problem. In addition, the N/D solutions will be classified according to their uniqueness index.

In earlier treatments, the possibility that N/D solutions could fail to solve the original problem was ignored. (The only exceptions of this are Refs. 10 and 13, where, however, much stronger conditions on L than the present ones are imposed.) This was probably because, on the one hand, the result to be proved is very plausible *provided* that N is reasonably smooth, but, on the other hand, it is difficult to *prove* any smoothness rigorously, since all one knows initially is that n is in L_2.

Because of the difficulties involved, even the present treatment is incomplete when $A \neq 0$; in this case, a certain degree of smoothness of N (to be made precise below) will have to be *assumed*. For this reason, the theorems from now on will all be given subject to the restriction $A = 0$. Nevertheless, both their statements and method of proof go through unchanged for the case $A \neq 0$ provided the above mentioned assumption is allowed, and *we conjecture that the stated restriction $A = 0$ is not, in fact, necessary* (i.e., that the smoothness assumption does, in fact, follow from our initial requirements).

The result which is proved in the first half of this section is then the following theorem.

Theorem 8: For $0 < \beta < 2$ and $A = 0$, every L_2 solution of Eq. (96) yields [via Eqs. (93), (66) or (70), and (65)] an allowed solution of the problem.

Proof: The vital step is the following lemma concerning the smoothness of N and D.

Lemma 6: For $0 < \beta < 2$ and $A = 0$, every L_2 solution of Eq. (96) yields a function D with the following properties: (a) D is analytic in the plane cut along $1 < x < \infty$ except for poles at the points b_i and on the upper side of the cut $\operatorname{Im} D = -\rho N$. (b) $\operatorname{Re} D$ and $\operatorname{Im} D$ are H.c. in every subinterval of $0 < u < 1$ which does not include the end point $u = 0$. (c) Near an end point $u = u_0$, *either*

$$D(u) \sim (u - u_0)^c, \tag{136}$$

with $c > -1$, $\operatorname{Re} D/|u - u_0|^c$ and $\operatorname{Im} D/|u - u_0|^c$ being H.c. near and at u_0, *or*

$$\operatorname{Im} D(u) \sim (u - u_0)^r, \tag{137}$$

$$\operatorname{Re} D(u) = \text{const}\,(u - u_0)^r \log(u - u_0) + f(u), \tag{138}$$

where r is an *integer* ≥ 0 with $\text{Im}\, D/|u - u_0|^r$ and $f/|u - u_0|^r$ H.c. near and at $u = u_0$. The second alternative is only possible for $u_0 = 1$ if $\alpha =$ integer and is only possible for $u_0 = 0$ if both $\beta = 1$ and $A = 0$. (d) If $D = 0$ (i.e., $\text{Re}\, D = \text{Im}\, D = 0$) for $u = u_1$, $0 < u_1 < 1$,

$$\bar{N} = [(u - u_2)/(u - u_1)]N \tag{139}$$

and

$$\bar{D} = [(u - u_2)/(u - u_1)]D \tag{140}$$

are solutions of the N/D equations [part (a) or (b) of Lemma 6] with $(\rho/u)^{\frac{1}{2}}\bar{N}$ in L_2, for any choice of u_2. If N and D satisfy the N/D equations with $\chi = \chi_0 > 0$, $\bar{N}$ and $\bar{D}$ will satisfy them with $\chi = \chi_0 - 1$, provided that u_2 is taken to be one of the points b_i in the equations for N and D; otherwise, the value of χ is the same for $\bar{N}$ and $\bar{D}$ as for N and D.

Proof of Lemma 6: The difficult part is to prove (b) and (c), and this is done in Appendix E. The method is essentially to iterate Eq. (96); for instance, if $b = 0$, one has

$$n = kn = k(kn) = k[k(kn)] = \cdots . \tag{141}$$

It is shown that k maps the class L_2 onto a smaller class S_1 (say), and that it also maps S_1 onto a still more narrow class S_2, S_2 onto S_3, etc., until one finally arrives at a class of functions sufficiently well behaved that the results under (b) and (c) follow. This iterative proof of parts (b) and (c) does not work for $A \neq 0$, but, if parts (b) and (c) are assumed, the rest of the proof of the theorem applies also to the case $A \neq 0$.

Knowing the properties of Im D under (b) and (c), part (a) is a standard consequence Appendix A. Part (d) follows easily from the identity

$$(u' - u)^{-1} - (u' - u_1)^{-1} = (u - u_1)(u' - u_1)^{-1}(u' - u)^{-1}, \tag{142}$$

except for the statement that $(\rho/u)^{\frac{1}{2}}\bar{N}$ must be in L_2; this may be proved by the techniques of Appendix D just mentioned.

Once Lemma 6 is established, the rest of the proof of Theorem 8 is effected by unwinding the steps followed in Sec. 4 in setting up the N/D equations. First, we prove the following lemma.

Lemma 7: Under the conditions of Lemma 6, the phase $-\delta$ of D satisfies conditions (i) and (ii) of Sec. 1.

Proof of Lemma 7: First, it will be shown that, because of part (d) of Lemma 6, the number of zeros of D in the interval $0 < u < 1$ must be finite. For every zero of D in this interval, one can find, if $\chi > 0$, a solution of Eq. (96) with the next lowest value of χ. Eventually, one will either run out of zeros or reach $\chi = 0$. In the latter case, for every remaining zero of D in this interval, one can produce a new linearly independent L_2 solution of Eq. (96), with $\chi = 0$, and hence with b equal to either 0 or $(\rho/u)^{\frac{1}{2}}L$. From Theorem 7 there is only a finite number of such solutions. This proves the required result.

If all the zeros in $0 < u < 1$ are divided out, one obtains a new function $\bar{D}$ which still has the properties (a)–(c) of Lemma 6 [since $(\rho/u)^{\frac{1}{2}}\bar{N}$ is an L_2 solution of Eq. (96)]; also the phase of $\bar{D}$ just above the cut is still δ. One may then use part (b) of Lemma 6 as follows: Just above the cut, we have

$$\bar{D}^2 = |\bar{D}|^2\, e^{-2i\delta}, \tag{143}$$

so that

$$e^{-2i[\delta(u')-\delta(u)]} = \left(1 + \frac{\bar{D}^2(u') - \bar{D}^2(u)}{\bar{D}^2(u)}\right) \times \left(1 + \frac{|\bar{D}(u)|^2 - |\bar{D}(u')|^2}{|\bar{D}(u')|^2}\right). \tag{144}$$

Because $|\bar{D}|^2$ never vanishes for $0 < u < 1$ and Re $\bar{D}$ and Im $\bar{D}$ are H.c. except near and at the end points, it follows that, for all u and u' satisfying $a < u < u' < b$ ($a > 0$ and $b < 1$),

$$|e^{-2i[\delta(u')-\delta(u)]} - 1| < \text{const}\, |u' - u|^{\mu}. \tag{145}$$

(Use Lemmas, A3 and A4 of Appendix A.) Hence, there exists a number Δ such that, for all u and u' satisfying $|u' - u| < \Delta$,

$$|e^{2i[\delta(u')-\delta(u)]} - 1| < 2. \tag{146}$$

For all such u and u', one then has

$$|\delta(u\ ) - \delta(u)| < \text{const}\, |e^{2i[\delta(u')-\delta(u)]} - 1|, \tag{147}$$

and, hence,

$$|\delta(u') - \delta(u)| < \text{const}\, |u' - u|^{\mu}, \tag{148}$$

for all u, u' satisfying both $a < u < u' < b$ ($a > 0$, $b < 1$) and $|u' - u| < \Delta$. However, since (148) implies that δ is bounded for $a < u < b$, one can remove the last restriction with (if necessary) a redefinition of the constant.

This proves that δ is H.c. except near and at $u = 0$ and $u = 1$. If near either of these points $\bar{D}$ has the first type of behavior in (c) of Lemma 6, one can divide by $|u - u_0|^c$ so that the above arguments are valid even near and at the end point; then one will deduce that δ is H.c. near and at the end point. Otherwise, it is clear from

$$\tan \delta = (\text{Im}\, D/\text{Re}\, D) \tag{149}$$

and part (c) of Lemma 6 that δ has the logarithmic behavior of Sec. 1. This proves Lemma 7.

Next, let us prove the following lemma.

Lemma 8: Under the conditions of Lemma 6, D has only a finite number of zeros in the complex plane.

Proof: Define a function P by

$$D(x) = P(x) \times \exp\left(-\frac{1}{\pi}\int_1^\infty dx' \frac{x}{x'}\frac{\delta(x')}{x'-x}\right)\Big/\prod_{i=1}^{\chi}\left(1-\frac{x}{b_i}\right). \tag{150}$$

It will be shown that P is a polynomial. As was stated in Sec. 4, it follows from Appendix A that (because δ satisfies the conditions of Sec. 1) the exponential is real analytic in the plane cut along $1 < x < \infty$ with phase $-\delta$ just above the cut. P is real analytic in the cut plane with zero imaginary part on the cut and so[19] is actually singular at most at $x = 1$ and $x = \infty$. Then, from the bounds (78), (79) and (136), (131), and (138), there is no singularity at $x = 1$ and at worst a pole at $x = \infty$, i.e., P is a polynomial.

Finally, we need the following result which follows in a straightforward manner from part (b) of Lemma 6 (see Appendix A).

Lemma 9: Under the conditions of Lemma 6, define

$$X(x) = \frac{1}{\pi}\int_1^\infty dx' \frac{L(x')\rho(x')N(x')}{x'-x}. \tag{151}$$

Then X is real analytic in the plane cut along $1 < x < \infty$, just above the cut $\operatorname{Im} X = L\rho N$, and near $x = \infty$ and $x = 1$, it satisfies the bounds $X <$ const x^p and $X <$ const $(x-1)^{-q}$, respectively, for some p and q.

Now we can put all the pieces together. From the integral equation (67) or (71) and Lemmas 6 and 9,

$$N(x) = L(x)D_+(x) + X_+(x), \tag{152}$$

where the $+$'s mean that the functions are evaluated on the upper side of the cut. We define

$$F(x) = N(x)/D_+(x), \quad \text{for } 1 < x < \infty, \tag{153}$$

and

$$R(x) = X(x)/D(x), \quad \text{for } x \text{ in the cut plane}. \tag{154}$$

Then, clearly,

$$F(x) = L(x) + R_+(x) \tag{155}$$

and R satisfies all the conditions under part (I) of the problem. Finally,

$$\operatorname{Im}(F^{-1}) = -\rho(x)N(x)/N(x) = -\rho(x), \tag{156}$$

so that part (II) of the problem is also satisfied.

This ends the proof of Theorem 8. In the course of the proof, it has emerged [Eq. (150) *et seq.*] that the construction of N and D in Sec. 4 is the most general one possible, at least if one requires that $n = (\rho/u)^{\frac{1}{2}}N$ is in L_2. Using this fact, we shall now investigate in detail the connection between the uniqueness index of a solution and its N/D decomposition.

First, let us associate with any N/D decomposition a pair of integers q and p as follows.

Definition 3: The *number of coincident zeros q* is defined as the number of zeros of Φ [Sec. 4 following Eq. (81)], counting each zero ξ times where ξ is its order.

Comment: Actually q is also the number of coincident zeros of D and χ, provided Φ has no zero at $x = 1$; this is the reason for the terminology. If N is analytic, it is also the number of coincident zeros of D and N (see Ref. 4).

Definition 4: The *asymptotic class* of D is defined to be the integer p for which the following are true:

(i) If $\beta = 1$ and $A = 0$,

$$D \sim \log x \cdot x^{-\frac{1}{2}p}, \quad p \text{ even}, \tag{157}$$

or

$$D \sim x^{-\frac{1}{2}(p-1)}, \quad p \text{ odd} \tag{158}$$

(i.e., let $p = 0, 1, 2 \cdots$ for $D \sim \log x$, const, $\log x/x$, $1/x$, $\log x/x^2, \cdots$).

(ii) If β has any other value in the range $0 < \beta < 2$ under consideration or if $A \neq 0$,

$$D \sim x^{c-\frac{1}{2}p} \tag{159}$$

which c lies in the range

$$\tfrac{1}{2}(\beta - 1) < c < \tfrac{1}{2}\beta. \tag{160}$$

[The strict inequalities are permissible because of (16).]

Lemma 10: If $(\rho/u)^{\frac{1}{2}}N$ is in L_2, the asymptotic class p of D is *nonnegative.*

Proof: If D has the asymptotic behavior $D \sim x^k$ with k nonintegral, then $N = -\operatorname{Im} D/\rho$ must have the behavior $N \sim x^{k-\beta}$; for $(\rho/u)^{\frac{1}{2}}N$ to be in L_2, one must therefore have $k - \frac{1}{2}\beta < 0$, and it is easy to verify that this implies $p \geq 0$. If $D \sim x^k$ or $x^k \log x$ with k integral, the requirement $p \geq 0$ follows from the requirement that $k \leq 0$ or ≤ -1, respectively (which is necessary if D is to satisfy the once subtracted dispersion relation).

With these definitions, it is easy to establish the following result.

Lemma 11: For $0 < \beta < 2$ and $A = 0$, a solution $F = N/D$ with $n = (\rho/u)^{\frac{1}{2}}N$ in L_2 has uniqueness index

$$\kappa = 2\chi - 2q - p, \tag{161}$$

where κ, χ, p, and q are defined in Definitions 1, 2, 3, and 4.

Proof: This follows immediately upon comparing the behavior

$$D \sim x^{(n+m+\delta(\infty)/\pi)+(q-\chi)}, \tag{162}$$

obtained from (78), with the definitions of p and κ.

Remembering that p and q are nonnegative, we see that Lemma 11 together with Theorems 6 and 8 implies the following important result.

Theorem 9: For $0 < \beta < 2$ and $A = 0$, one may obtain *all* allowed solutions of the problem with index $\kappa \leq 2\chi$ by solving the N/D equations for all possible values of the 2χ parameters k_i and d_i [restricting oneself to solutions $n = (\rho/u)^{\frac{1}{2}}N$ in L_2 and holding the points b_i at any fixed values satisfying the restrictions of Theorem 6]. Conversely, *every* function obtained by this procedure will be an allowed solution with $\kappa \leq 2\chi$.

Discussion: This result together with Theorem 7 tells us that, provided only solutions with κ not exceeding some (even) maximum are required, these can be obtained in terms of only a finite number of parameters. If L and ρ are not such that the second Fredholm alternative holds (Theorem 7), these are just the parameters k_i and d_i. Otherwise, there will be additional parameters relating to the nonuniqueness of the solution of the integral equation; but, in that case there will also be restrictions on the k_i and d_i; hence, it is not yet clear how many independent parameters will remain.

Even when the number of parameters involved in N and D has been ascertained, one cannot immediately deduce that the same number is involved in the solution $F = N/D$, because the N/D decomposition is not unique.

These problems, and related ones, are tackled in the next section.

7. EXISTENCE THEOREMS

First, we need another definition. It makes sense in view of Lemma 3.

Definition 5: For any pair of functions L and ρ (satisfying the conditions of Sec. 1), the *index of the problem* κ_0 is that integer such that there is an allowed solution with uniqueness index $\kappa = \kappa_0$ but no allowed solution with $\kappa < \kappa_0$.

The result to be established is the following theorem.

Theorem 10: For $0 < \beta < 2$ and $A = 0$, one always has $\kappa_0 \leq 0$. If $\kappa_0 = 0$, there is one allowed solution with $\kappa = 0$ and a κ parameter infinity of allowed solutions with each $\kappa > 0$. If $\kappa_0 < 0$, there is one allowed solution with $\kappa = \kappa_0$, no allowed solution with $\kappa_0 < \kappa \leq |\kappa_0|$, and a κ parameter infinity of allowed solutions with each $\kappa > |\kappa_0|$. The second alternative is "exceptional" in the following senses: Except for $\kappa_0 = -1$, it requires the second Fredholm alternative to hold for Eq. (96), and for $\kappa_0 = -1$ it requires the (unique) D function obtained from the $\chi = 0$ equations to have asymptotic class $p > 0$.

Proof: The proof will proceed by considering in turn each of the various possibilities with regard to the Fredholm alternatives for the integral equation. Extensive use will be made of Appendix F, which parametrizes explicitly the most general possible form for the N/D decomposition corresponding to a given allowed solution. Also, Lemma 3 will be heavily used; of course, this already gives part of the theorem directly.

(a) First Fredholm alternative (F.a.) valid. By Theorem 7, this situation will happen for "most" choices of L and ρ. It will be shown that it corresponds to $\kappa_0 = 0$ or -1, the second case being exceptional in that then the D function obtained from the $\chi = 0$ equations has $p > 0$.

First look at the $\chi = 0$ Eqs. (73)–(76). The homogeneous Eq. (76) has no solution [here and in the following the restriction to functions $n = (\rho/u)^{\frac{1}{2}}N$ in L_2 is understood]. The inhomogeneous equation (74) has a unique solution, which gives the *only* pair of N and D functions yielding a solution with $\kappa \leq 0$; hence, from Appendix F, they actually yield a solution with $\kappa = 0$ or -1 (otherwise, further N and D functions could be constructed). The cases 0 or -1 will arise according to whether D has asymptotic class $p = 0$ or 1. Finally, since $D(0) \neq 0$ the solution can have no pole of R at $x = 0$. In summarizing, this becomes the following lemma.

Lemma 12: If the first F.a. holds, $\kappa_0 = 0$ or -1. The solution with $\kappa = \kappa_0$ may be obtained from the $\chi = 0$ equations (73) and (74) [valid for $D(0) \neq 0$], and the cases $\kappa_0 = 0$ or -1 arise according to whether the resulting D function has asymptotic class 0 or 1, respectively. This solution has no pole at $x = 0$.

Now consider the $\chi = 1$ equations (75) and (76), valid when $D(0) = 0$. They have a unique solution for each choice of the pair of parameters $\{d_1 k_1\}$ (here and in the following, the points b_i are to be fixed at arbitrary values, subject to the conditions in Theorem 6). The D function depends linearly on the parameters:

$$D(x) = d_1 P(x) + k_1 Q(x). \qquad (163)$$

The functions P and Q are linearly independent [since D is never identically zero, from (77)]. Because one can choose in particular $D = P$ or $D = Q$, it must be possible to assign an asymptotic class p to P and Q, separately. It will be shown that, in fact, they both have $p \leq 1$ and that at least one of them has $p = 0$.

For any choice of the pair $\{d_1 k_1\}$, (161) implies that, if $p \geq 2$,

$$\kappa \leq -2q \leq 0. \qquad (164)$$

Hence, from Lemma 3, $\kappa = \kappa_0$ ($= 0$ or -1), hence $q = 0$. On the other hand, since we are dealing with the equations valid for $D(0) = 0$ and the $\kappa = \kappa_0$ solution has no pole at $x = 0$, Φ must have a zero at $x = 0$, i.e., $q \geq 1$ which is a contradiction. This proves that $p \leq 1$ for all choices of $\{d_1 k_1\}$, hence, in particular for P and Q. Finally, if both P and Q had $p = 1$, it would be possible to choose the ratio d_1/k_1 so that the leading term of D vanished, i.e., so that D had $p > 1$; hence, at least one of P and Q has $p = 0$.

From this, it follows that D can actually only have $p = 1$ (i.e., the leading term if D can only vanish) if $\{d_1 k_1\}$ satisfy some equation

$$a\, d_1 + b\, k_1 = 0, \qquad (165)$$

where at least one of a and b is nonzero, in other words, if the ratio d_1/k_1 takes on some exceptional value (perhaps infinity).

From (161), $p = 0$ implies that $\kappa = 2 - 2q = 2, 0, -2 \cdots$. Then, Lemma 3 (Sec. 3) shows that, if $\kappa_0 = -1$, $\kappa = 2$; if $k_0 = 0$, $\kappa = 2$ or 0. If $\kappa = 0$, $q = 1$; from Appendix F this will only happen for one special choice of the ratio d_1/k_1. If $\kappa = 2$, $q = 0$; this implies that (i) every choice of the ratio d_1/k_1 gives a different solution F (Appendix F) and (ii) these solutions all have a pole of R at $x = 0$ [since Φ has no zero but $D(0) = 0$].

Finally, consider the exceptional case when $p = 1$. This must give a solution with $\kappa = 1 - 2q = 1, -1, -3, \cdots$. Lemma 3 shows that, in fact, if $\kappa_0 = 1$, $\kappa = -1$; if $\kappa_0 = 0$, $\kappa = 1$.

The above results may be summarized in the following lemma.

Lemma 13: If the first F.a. holds, there is a 1-parameter infinity of solutions with $\kappa = 2$ and a pole at $x = 0$, obtained from the $\chi = 1$ equations (70) and (71) by varying the ratio k_1/d_1. If $\kappa_0 = 0$, there is also a solution with $\kappa = 1$ and a pole at $x = 0$, obtained for some special value of this ratio.

Next look at the $\chi = 1$ equations (66) and (67), valid for $D(0) \neq 0$. The general form of D is

$$D(x) = G(x) + [d_1\, P(x) + k_1\, Q(x)], \qquad (166)$$

with G not identically zero. D will have $p = 0$ except when $\{d_1\, k_1\}$ satisfy some equation

$$c + a\, d_1 + b\, k_1 = 0. \qquad (167)$$

At least one of a and b is nonzero, so that this equation has a 1-parameter infinity of solutions $\{d_1\, k_1\}$.

The use of Lemma 3 and Appendix F, as above, shows that when $p = 0$, if $\kappa_0 = -1$, $\kappa = 2$; if $\kappa_0 = 0$, $\kappa = 2$ or 0; the case $\kappa = 0$ will only occur for a 1-parameter infinity subset of choices of $\{d_1\, k_1\}$; when $\kappa = 2$, every choice of $\{d_1\, k_1\}$ will give a different solution F.

Again, consider the exceptional cases when D has $p > 0$. The use of Appendix F, as above, shows that one never has $p > 2$ and that the case $p = 2$ can only occur if $\kappa_0 = 0$, and then only for one choice of $\{d_1\, k_1\}$. Thus, $p = 1$ except possibly for this one choice of $\{d_1\, k_1\}$. Then, Appendix F shows that, if $\kappa_0 = -1$, $\kappa = \kappa_0$; if $\kappa_0 = 0$, $\kappa = 1$.

These results may be summarized in the following lemma.

Lemma 14: If the first F.a. holds, there is a 2-parameter infinity of solutions with $\kappa = 2$ (and no pole at $x = 0$) obtainable from the $\chi = 1$ equations by varying the pair $\{d_1\, k_1\}$. If $\kappa_0 = 0$, there is also a 1-parameter infinity of solutions with $\kappa = 1$ (and no pole at $x = 0$) obtainable by varying the pair $\{d_1\, k_1\}$ subject to some linear constraint.

It may be shown by exactly similar arguments that similar results are obtained for $\chi > 1$. We summarize all this in the following important statement.

Theorem 11: Let $0 < \beta < 2$ and $A = 0$, and let the first Fredholm alternative hold for Eq. (96). Then $\kappa_0 = 0$ or -1. There is a unique solution with $\kappa = \kappa_0$, obtainable from the $\chi = 0$ equations (73) and (74); the cases $\kappa_0 = 0$ or -1 occur if D has asymptotic class

0 or 1, respectively. There is a κ-parameter infinity of solutions with each $\kappa > |\kappa_0|$. All the solutions with a given even value of κ may be obtained by solving the N/D equations with $\chi = \frac{1}{2}\kappa$, varying the κ parameters $\{d_1 \cdots d_\chi k_1 \cdots k_\chi\}$ over all possible choices excluding some $(\kappa - 1)$ parameter subset (consisting of those choices for which $q > 0$ and/or $p > 0$, i.e., for which there are coincident zeros and/or a vanishing of the leading term of D at infinity). All the solutions with the next lowest value of κ, i.e., $\kappa = 2\chi - 1$, may be obtained by varying $\{d_1 \cdots k_\chi\}$ over this subset, excluding some still smaller $(2\kappa - 2)$ parameter subset (those choices for which $q > 0$ and/or $p > 1$). This last subset yields solutions which are redundant in that they have $\kappa \leq 2\chi - 2$ and hence can be obtained by the above procedure with some smaller choice of χ.

Discussion: The results contained in the above theorem are no surprise; they constitute an advance on previous work only in their generality, and in the fact that an explicit proof has been spelled out. Their essential content is obviously the following: If a solution of the N/D equations has no coincident zeros and is no more well behaved at infinity than it has to be for the integrals to converge, then it will yield a solution with $\kappa = 2\chi$; this will "usually" be the case, so that there is a κ-parameter infinity of such solutions, obtainable by varying the 2χ parameters appearing in the N/D equations; if a condition is imposed on the parameters to ensure that the leading term of D at infinity vanishes, one will obtain a κ-parameter infinity of solutions with $\kappa = 2\chi - 1$.

When we move on to the case where the first Fredholm alternative ceases to be valid, the situation is quite different, and no previous treatment has tackled this case. It is obvious that the κ-parameter infinity of solutions (if it exists) cannot be found simply by varying the set $\{d_1 \cdots k_\chi\}$ because there is no longer a unique solution for each choice of the set.

(b) Second F.a. holds, one linearly independent solution, not orthogonal to $(\rho/u)^{\frac{1}{2}}L$. This is the case where the homogeneous $\chi = 0$ equation (76) has a unique solution (up to an over-all constant), but the inhomogeneous equation (74) has no solution. By the same arguments as for case (a), the former equation gives a unique solution F with $\kappa = \kappa_0 = 0$ or -1, according to whether D has $p = 0$ or 1. The only difference from the previous case is that, since it is the homogeneous equation which is involved, $D(0) = 0$; hence (because $q = 0$ by Appendix F) R has a pole at $x = 0$. Thus, we have the following.

Lemma 15: In case (b), $\kappa_0 = 0$ or -1. There is just one solution with $\kappa = \kappa_0$, obtainable from the $\chi = 0$ equation (76), and it has a pole at $x = 0$.

If we look again at the derivation of the N/D equations (Sec. 4), it will be seen that the choices of $x = 0$ as the normalization point for D was not essential; any point $x = x_0$ would have done, with $-\infty < x_0 < 1$. It is therefore always possible to avoid case (b) by choosing some other value for x_0, such that R no longer has a pole at x_0. Then [since, as will be seen, $\kappa_0 = 0$ or -1 only for cases (a) and (b)] one will recover case (a).

Because of this, no further investigation of case (b) is necessary.

(c) Second F.a. holds, one linearly independent solution, orthogonal to $(\rho/u)^{\frac{1}{2}}L$. This is the case where the $\chi = 0$ homogeneous equation (76) has one solution (up to an over-all constant) and the $\chi = 0$ inhomogeneous equation (74) has a 1-parameter infinity of solutions. These give the *only* N and D functions yielding solutions with $\kappa \leq 0$; hence, by Lemma 3 and Appendix F they must all yield a single solution with $\kappa = -2$ or -3 (the two cases arising according to whether the D function obtained from the homogeneous equations has $p = 0$ or 1) and no pole of R at $x = 0$. We summarize in the following lemma.

Lemma 16: In case (c), $\kappa_0 = -2$ or -3. There is a single solution with $\kappa = \kappa_0$, which is obtainable from the $\chi = 0$ equations and has no pole of R at $x = 0$, the cases $\kappa_0 = -2$ or -3 occurring when the solution D_0 of the homogeneous $\chi = 0$ equations has $p = 0$ or 1, respectively.

Let us note in passing that the result just deduced, namely, that every solution N of the inhomogeneous equation (74) is just equal to a (first-order) polynomial times the solution of the homogeneous equation, is far from obvious if one just looks at these integral equations themselves. One needs to know that, first, $F = N/D$ is an allowed solution of the problem with $\kappa \leq 0$ and, then, that the uniqueness Lemma 3 applies. It is here that previous treatments are not adequate, in that no assurance is given that F will be an allowed solution.

Now look at the case $\chi > 0$, treating first Eqs. (70) and (71), valid for $D(0) = 0$. The orthogonality condition for the inhomogeneous term imposes a linear condition

$$\sum_{i=1}^{\chi} (A_i d_i + B_i k_i) = 0, \tag{168}$$

and for each choice of $\{d_1 \cdots k_\chi\}$ satisfying this condition there is the 1-parameter infinity of D functions

$$D(x) = \sum_{i=1}^{\chi} [d_i P_i(x) + k_i Q_i(x)] + \lambda D_1(x), \quad (169)$$

where D_1 is the solution of the homogeneous equations.

It can easily be seen that at least one of the A_i or B_i is nonzero, indeed that one can never have $A_i = B_i = 0$ for *any* value of i. For suppose this were possible for $i = 1$ (this clearly involves no loss of generality); then for $\chi = 1$ there would be no restriction on $\{d_1 k_1\}$, and hence there would be a 3-parameter infinity of D functions all yielding solutions F with $\kappa = 2 - 2q - p \leq 2$ and hence, from Lemma 3, all yielding the solution with $\kappa = \kappa_0$ $(= -2$ or $-3)$. But, from Appendix F, this solution can only yield a 2-parameter infinity of D functions [with $D(0) = 0$], which is a contradiction.

Since not all the A_i and B_i are zero, the orthogonality condition removes one (but only one) of the 2χ independent parameters. In (169), this parameter may obviously be replaced by λ, so that we deduce, in fact, that there is still *just a* 2χ*-parameter infinity* of D functions, of the form

$$D(x) = \sum_{i=1}^{2\chi} a_i D_i(x). \quad (170)$$

Now, one can use arguments similar to those for case (a) above. From Lemma 3, the only solution F with $\kappa < |\kappa_0|$ $(= 2$ or $3)$ is that with $\kappa = \kappa_0$, so that all solutions of the N/D equations with $\chi = 1$ must yield this solution. Let us consider $\chi = 2$.

It is easily seen that at least one of the functions $D_0 \cdots D_3$ must have $p = 0$, for, if they all had $p \geq 1$, there would be four linearly independent D functions (i.e., all of them) with $K = 2\chi - 2q - p \leq 3$; on the other hand, Lemma 3 and Appendix F tell us that, if $\kappa_0 = -2$, there are only two such functions, and, if $\kappa_0 = -3$, only three functions.

Because one of the D_i has $p = 0$, it follows that D will have $p = 0$ for every choice of $\{d_1 \cdots k_2\}$ except when these parameters satisfy a linear homogeneous equation with at least one coefficient nonzero (expressing the condition that the leading term of D vanishes). This condition will be satisfied by just a $[2\chi - 1$ $(= 3)]$-parameter subset of choices of $\{d_1 \cdots k_2\}$.

When $p = 0$, $\kappa = 4 - 2q = 4, 2, 0 \cdots$. If $q > 0$, $\kappa = |\kappa_0|$, and so Appendix F tells us that $q > 0$ only for a $(2\chi - 1)$-parameter subset of choices of $\{d_1 \cdots k_2\}$. If $q = 0$, $\kappa = 4$, and Appendix F tells us that we obtain a different solution F for every choice of $\{d_1 \cdots k_2\}$ (except that F is independent of the over-all normalization). Since $q = 0$, these solutions have a pole at $x = 0$.

Finally, consider the exceptional cases when $p > 0$. If $p = 1$, $\kappa = 3 - 2q = 3, 1, -1 \cdots$. Hence, if $\kappa_0 = -3$, $\kappa = \kappa_0$; if $\kappa_0 = -2$, $\kappa = 3$. If $p \geq 2$, $\kappa \leq 2$, and hence $\kappa = \kappa_0$; but Appendix F tells us that this can only happen for a $(2\chi - 2)$-parameter subset of choices of $\{d_1 \cdots k_2\}$.

So we have deduced the next lemma.

Lemma 17: In case (c) there is a $(\kappa - 1)$-parameter infinity of solutions with $\kappa = 4$ and a pole at $\chi = 0$, obtainable from the $\chi = 2$ equations (70) and (71) [valid for $D(0) = 0$] by varying the set $\{d_1 \cdots k_2\}$ subject to the orthogonality condition. If $\kappa_0 = -2$, there is also a $(\kappa - 1)$-parameter infinity with $\kappa = 3$ (and a pole of R at $x = 0$), obtainable by varying the set $\{d_1 \cdots k_2\}$ subject also to a further linear condition.

The extension of this result to arbitrary $\chi \geq 2$ goes through in the same way.

Now, let us look at Eqs. (66) and (67) valid for $D(0) \neq 0$, for the case $\chi > 0$. The orthogonality condition will again remove one of the parameters $\{d_i \cdots k_\chi\}$, which may be replaced by λ, so that there is still a 2χ-parameter infinity of D functions of the form

$$D(x) = G(x) + \sum_{i=0}^{2\chi-1} a_i D_i(x). \quad (171)$$

Exactly similar arguments to those above then give us the following.

Lemma 18: In case (c), there is a κ-parameter infinity of solutions with $\kappa = 4$ (and no pole of R at $x = 0$), obtainable from the $\chi = 2$ equations (66) and (67) by varying the set $\{d_1 \cdots k_\chi\}$ subject to the orthogonality condition. If $\kappa_0 = -2$, there is also a κ-parameter infinity with $\kappa = 3$ (and no pole of R at $x = 0$), obtainable by varying the set subject to a further linear condition.

Again, the extension to $\chi > 2$ needs nothing new.

(d) Second Fredholm alternative holds, arbitrary number of linearly independent solutions, all orthogonal to $(\rho/u)^{\frac{1}{2}}L$. This is the case where the homogeneous $\chi = 0$ equation has (say) r linearly independent solutions, and the inhomogeneous $\chi = 0$ equation has correspondingly an r-parameter infinity of solutions. Exactly the same arguments as for (c) above give the following result.

Lemma 19: In case (d), $\kappa_0 = -2r$ or $-(2r + 1)$. The solution with $\kappa = \kappa_0$ may be obtained from the $\chi = 0$ equations and has no pole of R at $x = 0$.

For $\chi = 0$, there will be a solution only when orthogonality conditions of the form

$$\sum_{i=1}^{\chi} (A_{ni} d_i + B_{ni} k_i) = 0, \quad n = 1 \cdots r, \quad (172)$$

are satisfied, and, when this is the case, the most general solution D is

$$D(x) = G(x) + \sum_{i=1}^{\chi} [d_i P_i(x) + k_i Q_i(x)] + \sum_{n=1}^{r} \lambda_n D_n(x), \quad (173)$$

with $G = 0$ for Eqs. (70) and (71).

From Lemma 3 there is no solution with $\kappa \leq 2r$ except the one with $\kappa = \kappa_0$; hence, the smallest value of χ which is of interest is $\chi = r + 1$ (remember that $\kappa \leq 2\chi$). It will be shown that, for $\chi \geq r + 1$, (172) removes just r of the 2χ parameters $\{d_i, k_i\}$ and, hence, that the most general solution D is still of the form

$$D(x) = G(x) + \sum_{i=1}^{2\chi} a_i D_i(x). \quad (174)$$

From Appendix F the solution with $\kappa = \kappa_0$ generates a $2r$-parameter infinity of D functions satisfying Eq. (70) with $\chi = r$. As these are the only possible ones, just r of the parameters $\{d_i k_i\}$ can be independent [from (173)], i.e., when $\chi = r$, the orthogonality conditions remove just r of the parameters $\{d_i k_i\}$. Since there are r equations, this is the greatest number possible, and so we conclude that just r parameters are also removed when $\chi > r$ as required.

Knowing that the most general solution D is given by (174), we can easily show, by similar arguments to those used for the previous cases, that there is a κ-parameter infinity of solutions with each $\kappa > \kappa_0$, which can be obtained as in Lemmas 16–18.

(e) Second F.a. holds, arbitrary number of linearly independent solutions, not all orthogonal to $(\rho/u)^{\frac{1}{2}}L$. If there are $(r + 1)$ linearly independent solutions of the homogeneous $\chi = 0$ equations, the unique solution F which they yield must have $\kappa = \kappa_0 = -2r$ or $-(2r + 1)$ and must have *a pole of R at $x = 0$* (since there is no solution of the inhomogeneous $\chi = 0$ equations). Hence case (e) may always reduce to case (d) by a change in the normalization point for D. [See case (b) above.]

8. THE CASE OF ARBITRARY β

With the results of the last section, the treatment for $0 < \beta < 2$ is complete. It remains to consider the case of arbitrary $\beta > 0$—or rather to learn how to avoid considering it.

The point is that, if $F = F_0$ is a solution of the problem when $L = L_0$ and $\rho = \rho_0$, then $F = \Phi F_0$ is a solution of the problem when $L = \Phi L_0$ and $\rho = \Phi^{-1}\rho_0$, if Φ is any real rational function. This is obvious since (i) the replacement $R \to \Phi R$ does not alter the validity of the conditions under (I) if the problem and (ii) since Φ is real, (II) is not affected either.

For this reason, it is always possible to reduce the case of arbitrary β to the case $0 < \beta \leq 1$. There is, however, a difficulty in doing this: namely, that the connection between the uniqueness indices for the two cases is not very simple unless restrictive assumptions are made. This is due to the fact that (1) although the replacement $F \to \phi F$ will normally introduce new poles of R, it will not do so if the original R possesses zeros coincident with poles of ϕ and (2) the replacement may also *remove* poles of R.

To discuss this in more detail, it will be convenient to consider the case where Φ has just a single pole or zero. First, we have to tie up a loose end, namely, to ensure that the "transformed" problem still has F, L, and ρ satisfying the conditions presupposed in this paper (Sec. 1). The following is obviously true.

Lemma 20: Let $F = F_1$ be an allowed solution of the problem when $L = L_1$ and $\rho = \rho_1$ and let L_1 and ρ_1 satisfy the conditions of Sec. 1. Then, if $-\infty < x_0 < 1$ and k is an integer with $0 < k \leq \beta_1$, $F = F_2 \equiv (x - x_0)^k F_1$ is an allowed solution of the problem when $\rho = \rho_2 \equiv (x - x_0)^{-k}\rho_1$ and

$$L = L_2 \equiv (x - x_0)^k \left(L_1 - \sum_{i=1}^{k} c_i (x - x_0)^{-i} \right); \quad (175)$$

also, ρ_2 and L_2 satisfy the conditions of Sec. 1, provided that constants c_i are chosen so that the square bracket is $o(x^{-k})$. The converse of this statement is also true.

Now, let us look at the relation between κ_1 and κ_2. It will be convenient to assume that F, L, and N are analytic (and therefore well defined) at least out to the point x_0.

From Definition 1, the general relation is clearly

$$\kappa_1 = \kappa_2 - k + 2\xi_1, \quad (176)$$

where ξ_1 is zero unless F_1 has a pole at $x = x_0$, in which case it is the order of the pole or k, whichever is less.

Thus, there are two extreme cases: (a) if F_1 has no pole at $x = x_0$, $K_1 = K_2 - k$; (b) if F_1 has a kth order pole at $x = x_0$, $K_1 = K_2 + k$.

If we are considering F_1 as a solution of the problem, case (a) will be the "usual" one in that there is no particular reason to expect a pole at $x = x_0$.

However, if we are solving for F_2, it becomes case (b) which is "usual," because case (a) (or any intermediate one) can only occur if F_2 happens to have a zero at $x = x_0$.

Now, let us return to the task which motivated this discussion, namely,to solve the problem for $\beta \geq 2$ by reducing it to the case where β is in the canonical range $0 < \beta < 2$. If the original amplitude is F_1, then F_2 can always be made to lie in the canonical range by a suitable choice of k: Indeed, unless β = integer, there are two possible choices for k, as one may have F_2 with either $0 < \beta < 1$ or $1 < \beta < 2$. After the problem has been solved for F_2 by the method described in the preceding sections, one can then transform back to F_1 to obtain the solution of the original problem. The only difficulty is to keep track of the uniqueness index.

For this purpose it will be simplest, and quite adequate for physical applications, to impose a condition that F_1 has *no pole at the point* x_0. This then requires that F_2 has a kth-order zero here, i.e., that the parameters are chosen in such a way that the N function for F_2 has its first k derivatives vanishing, which will give k linear conditions:

$$H_n + \sum_{i=1}^{\chi} (F_{ni} d_i + G_{ni} k_i) = 0, \quad n = 1 \cdots k. \quad (177)$$

In general, these will remove k of the 2χ parameters $\{d_1 \cdots k_\chi\}$. Since the index κ_2 will "usually" be given by $\kappa_2 = 2\chi$ and since there is usually just one solution F_2 for "almost" every choice of $\{d_1 \cdots k_\chi\}$ (Theorem 11), we deduce that there is "usually" a $[(\kappa_2 - \kappa) = \kappa_1]$-parameter infinity of solutions satisfying the condition with each even $\kappa_2 \geq k$. Finally, the restriction to even κ_2 does not restrict the parity of $(\kappa_2 - k)$ since one can always choose k to have either of two adjacent values, unless β = integer, as mentioned above. So we deduce that *there will "usually" be a κ_1-parameter infinity of solutions F_1 with each* $\kappa_1 \geq 0$.

Thus the uniqueness index retains, at any rate for the "usual" case, the significance it had for $0 < \beta < 2$. A detailed discussion of all the exceptional cases along the lines of Sec. 7 would presumably allow a completely general statement similar to Theorem 10.

Finally, it is of interest to get some insight into the problems occurring if one were to try a direct N/D solution for $\beta > 2$. From Sec. 4 one can still get an N/D representation, with $(\rho/u)^{\frac{1}{2}}N$ being an L_2 solution of the integral equation (96). However, because of the term β in the definition of κ, it is easy to verify that, instead of the inequality $\kappa \leq 2\chi$, one has now $\kappa \leq 2\chi - b_e$, when b_e is the smallest even integer not greater than β (i.e., $b_e = 0, 2, \cdots$ for $0 < \beta < 2, 2 \leq \beta < 4, \cdots$).

Thus, if there is only to be a κ-parameter infinity of solutions, we expect that the N/D equations *cannot* have solutions for every choice of the 2χ-parameters $\{d_1 \cdots k_\chi\}$; instead we expect solutions only if these satisfy b_e conditions. In other words, we expect that for $\beta \geq 2$ the integral equation for N no longer has a solution for every (L_2) choice of the inhomogeneous term.

APPENDIX A: DISPERSION INTEGRALS

Here, we have collected a number of results concerning dispersion integrals over Hölder continuous functions. One is valid when there is essentially Hölder continuity even near and at the end points, and is taken more or less verbatum from Ref. 17. The remainder concern the case when the behavior is logarithmic in the sense of (24) of the text, or simple extensions of this case, and the results here are new, although the technique of proof is taken from Ref. 3.

First, we give some simple facts used in the text.

Lemma A1: Let $\phi(u)$ be H.c. in $a < u < b$, and let $F(\phi[u])$ have a bounded derivative with respect to ϕ for all the values taken by ϕ in this range. Then F is H.c. with respect to u in $a < u < b$.

Proof: By the mean value theorem, since F has a bounded derivative,

$$|F(\phi[u']) - F(\phi[u])| < \text{const}\, |\phi(u') - \phi(u)|. \quad (\text{A1})$$

Lemma A2: The function u^c is H.c. in $a < u < b$ with Hölder index c.

Lemma A3: The product of two H.c. functions is H.c.

Lemma A4: The inverse of a nonvanishing H.c. function is H.c.

Proofs: In Ref. 17, Chap. 1.

The result, valid when one essentially has Hölder continuity near and at the end points, is the following (Ref. 17, Chaps. 2, 3, and 4).

Lemma A5: Let ϕ be H.c. in any subinterval of $0 < u < 1$ not including an end point, and near an end point (call it $u = c$) let

$$\phi(u) = \phi^*(u)/|u - c|^\gamma, \quad (\text{A2})$$

with ϕ^* H.c. near and at $u = c$ and $0 \leq \gamma < 1$. Then

$$\Phi(u) = \frac{1}{\pi}\int_0^1 du' \frac{\phi(u')}{u' - u} \tag{A3}$$

is real analytic in the u plane cut along $0 < u < 1$ and on the upper side of the cut approaches the limits

$$\operatorname{Im}\Phi(u) = \phi(u), \tag{A4}$$

$$\operatorname{Re}\Phi(u) = \frac{P}{\pi}\int_0^1 du' \frac{\phi(u')}{u' - u}. \tag{A5}$$

The function Re Φ is H.c. except near and at the end points. Near $u = c$, one has the following:

(i) If $\gamma = 0$,

$$|\Phi(u)| < \text{const}/\log|u - c|; \tag{A6}$$

(ii) if $0 < \gamma < 1$,

$$|\Phi(u)| < \text{const}\,|u - c|^{\gamma}. \tag{A7}$$

On the interval $0 < u < 1$, Re Φ, in addition, satisfies the following:

(i) If $\gamma = 0$,

$$\operatorname{Re}\Phi(u) = \pm\frac{\phi(c)}{\pi}\log\left(\frac{1}{|u - c|}\right) + \Phi_0(u); \tag{A8}$$

(ii) if $0 < \gamma < 1$,

$$\operatorname{Re}\Phi(u) = \pm\cot\pi\gamma\frac{\phi^*(c)}{|u - c|^{\gamma}} + \frac{\Phi_0(u)}{|u - c|^{\gamma-\epsilon}}, \tag{A9}$$

some $\epsilon > 0$. Here the + and − signs hold for $u = 0$ and $u = 1$, respectively, and in both cases Φ_0 is H.c. near and at $u = c$.

For the logarithmic behavior, we first need the following results.

Lemma A6: Let

$$\tan\delta = \operatorname{Im}X/\operatorname{Re}X,$$

with

$$\operatorname{Re}X = \log x + f(x), \tag{A10}$$

Im X and f being H.c. in $0 < u < a$. Then, for $0 < u < u' < a$,

$$|\delta(u') - \delta(u)| < \text{const}\left(|u' - u|^{\mu} + \frac{1}{u\log^2 u}|u' - u|\right). \tag{A11}$$

Proof: Since $\log u$ diverges near $u = 0$,

$$|\tan\delta(u') - \tan\delta(u)| < \text{const}\left(|u' - u|^{\mu} + \left|\frac{1}{\log u'} - \frac{1}{\log u}\right|\right). \tag{A12}$$

From the mean value theorem, one has, for $u < u'$,

$$|\delta(u') - \delta(u)| < |\tan\delta(u') - \tan\delta(u)|, \tag{A13}$$

$$\left|\frac{1}{\log u'} - \frac{1}{\log u}\right| < \frac{1}{u\log^2 u}|u' - u|, \tag{A14}$$

from which the lemma follows.

Lemma A7: If δ satisfies the conditions of Lemma A6, then for $0 < u < u' < b$

$$|\sin^2(u') - \sin^2(u)| < \text{const}\,[|u' - u|^{\mu} + (u\log^3 u)^{-1}|u' - u|], \tag{A15}$$

and also $u^{\gamma}\sin^2 u$ satisfies a similar condition for any $\gamma \geq 0$.

Proof: Using the mean value theorem and remembering that $\log u$ diverges, we have

$$\begin{aligned}|\sin^2\delta(u') - \sin^2\delta(u)| &< |\tan^2\delta(u') - \tan^2\delta(u)| \\ &< \text{const}\left(|u' - u|^{\mu} + \left|\frac{1}{\log^2 u'} - \frac{1}{\log^2 u}\right|\right) \\ &< \text{const}\left(|u' - u|^{\mu} + \frac{1}{u\log^3 u}|u' - u|\right).\end{aligned} \tag{A16}$$

For $u^{\gamma}\sin^2 u$, one writes

$$u'^{\gamma}\sin^2 u' - u^{\gamma}\sin^2 u = u'^{\gamma}(\sin^2 u' - \sin^2 u) + \sin^2 u(u'^{\gamma} - u^{\gamma}) \tag{A17}$$

and uses Lemma A6.

Now, we state the analog of Lemma A5 for logarithmic behavior near the end point. For simplicity of notation, we take the point $u = 0$.

Lemma A8: Let ϕ defined on $0 < u < a$ be H.c. except near and at $u = 0$, and near this point let

$$\phi(u) = \phi^*(u)u^{-\gamma} \tag{A18}$$

(some γ, $0 \leq \gamma < 1$), with

$$\phi^*(u) < \text{const}\log^{-k} u \tag{A19}$$

(some k, $k \geq 1$) and

$$|\phi^*(u') - \phi^*(u)| < \text{const}\,[|u' - u|^{\mu} + u^{-1}(\log^{-(k+1)} u)|u' - u|] \tag{A20}$$

(all u, u', $u < u'$). Then

$$\Phi(u) = \frac{1}{\pi}\int_0^a du' \frac{\phi(u')}{u' - u} \tag{A21}$$

is real analytic in the plane cut along $0 < u < a$ and just above the cut

$$\operatorname{Im}\Phi(u) = \phi(u), \tag{A22}$$

$$\operatorname{Re}\Phi(u) = \frac{P}{\pi}\int_0^a du' \frac{\phi(u')}{u'-u}, \tag{A23}$$

with Re Φ H.c. except near and at $u = 0$. Near $u = 0$,

$$\Phi(u) \sim \int_{|u|}^{a} \frac{du'}{u'} \phi(u') + O[|u|^{-\gamma}\log^{-k}|u|]. \tag{A24}$$

It is important to take notice of the fact that no statement is made about the continuity properties of Re Φ near $u = 0$, as none is needed in the text. It would in any case be very difficult to obtain such a statement in concise form.

Proof: We follow Ref. 3. Define $x = 1/u$, and write

$$\begin{aligned}\operatorname{Re}\Phi &= \int_{a^{-1}}^{x(1+\epsilon)} dx' \frac{\phi(x')}{x'} \\ &+ \int_{a^{-1}}^{x(1-\epsilon)} dx' \frac{\phi(x')}{x'-x} + P\int_{x(1-\epsilon)}^{x(1+\epsilon)} dx' \frac{\phi(x')}{x'-x} \\ &+ \int_{x(1+\epsilon)}^{\infty} dx' \frac{x}{x'} \frac{\phi(x')}{x'-x} \\ &\equiv I_1 + I_2 + I_3 + I_4,\end{aligned} \tag{A25}$$

with $\epsilon > 0$. The first term gives the first term of (A24). (It will be seen in Lemma A9 below that the ϵ can be ignored here.) For I_2 and I_4, one has

$$\begin{aligned}|I_2| &< \text{const } x^{-1}\int_0^{x(1-\epsilon)} dx' x'^{\gamma} \log^{-k} x' \\ &< \text{const } x^{\gamma}\log^{-k} x,\end{aligned} \tag{A26}$$

$$\begin{aligned}|I_4| &< \text{const} \int_{x(1+\epsilon)}^{\infty} dx' \frac{x}{x'} \frac{\log^{-k} x'}{x'-x} x'^{\gamma} \\ &= \text{const } x^{\gamma} \int_{1+\epsilon}^{\infty} dy \frac{1}{y} \frac{\log^{-k}(xy)}{y-1} y^{\gamma} \\ &< \text{const } x^{\gamma}\log^{-k} x.\end{aligned} \tag{A27}$$

For I_3, use

$$\begin{aligned}\phi(x') - \phi(x) &= x'^{\gamma}[\phi^*(x') - \phi^*(x)] \\ &+ \phi^*(x)(x'^{\gamma} - x^{\gamma}).\end{aligned} \tag{A28}$$

Assuming that μ has been chosen so that $\mu < \frac{1}{2}\gamma$ (this may always be done for the applications in the text), we conclude via (20) and Lemma 2 that

$$\begin{aligned}&|\phi(x') - \phi(x)| \\ &\quad< \text{const } x'^{\gamma}\log^{-k} x \\ &\quad\times [x^{-2\mu}|x'-x|^{\mu} + x^{-1}(\log^{-(k+1)} x)|x'-x|]\end{aligned} \tag{A29}$$

for $x' < x$. Since I_3 may be written

$$I_3(x) = \int_{x(1-\epsilon)}^{x(1+\epsilon)} dx' \frac{\phi(x') - \phi(x)}{x'-x}, \tag{A30}$$

this is easily found to give

$$\begin{aligned}|I_3| &< \text{const } x^{\gamma}(\log^{-k} x)(x^{-\mu} + \log^{-(k+1)} x) \\ &< \text{const } x^{\gamma}\log^{-(2k+1)} x.\end{aligned} \tag{A31}$$

For x not on the real axis, write $x = |x|\,e^{-i\theta}$, and

$$\begin{aligned}\Phi &= \int_1^{|x|} dx' \frac{\phi(x')}{x'} \\ &+ \int_1^{|x|} dx' \frac{\phi(x')}{x'-x} + \int_{|x|}^{\infty} dx' \frac{x}{x'} \frac{\phi(x')}{x'-x}.\end{aligned} \tag{A32}$$

The first term gives the first term of (24), and it is easy to verify that, for fixed $\theta \neq 0$, the assumed bound on ϕ implies that the other terms are bounded by const $|x|^{\gamma}\log^{-k}|x|$. This completes the proof of Lemma A8. Finally, the leading term of Φ will be computed explicitly.

Lemma A9: Under the conditions of Lemma A8, as $u \to 0$,

(i) if $\gamma = 0$ and $k > 1$,

$$\Phi(u) = \Phi(0) + \log^{(1-k)} u + O(\log^{-k} u); \tag{A33}$$

(ii) if $\gamma = 0$ and $k = 1$,

$$\Phi = \log(\log u) + O(\log^{-1} u); \tag{A34}$$

(iii) if $0 < \gamma < 1$ and $k \geq 1$,

$$\Phi(u)/(u^{-\gamma}\log^{(1-k)} u) \to 0. \tag{A35}$$

Proof: Cases (i) and (ii) involve standard integrals. For case (iii), write

$$\begin{aligned}r(x) \equiv \left(\int_{u^{-1}}^{x} dx'(\log^{-k} x') \cdot x'^{\gamma-1}\right) \\ \times (x^{\gamma}\log^{(1-k)} x)^{-1}\end{aligned} \tag{A36}$$

and let $y \equiv \log x$, $y' \equiv \log x'$, and $z = y'/y$. Then, as $y \to \infty$,

$$r = \int_{(ay)^{-1}}^{1} dz z^{-k} e^{zy(\gamma-1)} \to 0. \tag{A37}$$

APPENDIX B: PROOF OF THEOREM 1, PART (e)

The proof of Theorem 1, part (e), of the text is given here. First we need the following.

Lemma B1: Let f be H.c. near and at $u = 0$ with Hölder index μ, and let $f(0) = 0$. Then the same is true of

$$g(u) \equiv u^{-\mu} f^2(u). \tag{B1}$$

Proof: Use

$$g(u') - g(u) = u^{-\mu}(f^2(u') - f^2(u)) + f^2(u')(u'^{-\mu} - u^{-\mu}) \tag{B2}$$

and take $u' < u$. Then, by the mean value theorem, the first term is bounded by

$$2u^{-\mu} f(u)\,|u' - u|^\mu, \quad u' < u_1 < u. \tag{B3}$$

Since $f <$ const u^μ, the required result follows.

Now consider

$$I(x) \equiv \frac{P}{\pi}\int_1^\infty dx' \left(\frac{x'-1}{x'}\right)^m \frac{\operatorname{Im} F(x')}{x' - x}, \tag{B4}$$

with $m = a$, a being defined as in Eq. (25). One has

$$I(x) = \int_1^\infty dx' \frac{\phi(x')}{x' - x}, \tag{B5}$$

with

$$\phi(x) = (x-1)^{m-\alpha} \phi^*(x), \tag{B6}$$

$$\phi^*(x) \equiv \left(x^{-m}\frac{(x-1)^\alpha}{\rho(x)}\right) \sin^2 \delta(x). \tag{B7}$$

The square bracket is H.c. near and at $x = 1$; so, by Lemma A5,

$$I \to C \sin^2 \delta(x) \cot \pi(\alpha - m)(x-1)^{m-\alpha}, \tag{B8}$$

where C is the limit if the square bracket as $x \to 1$.

Using Lemma 2, one has

$$\frac{\sin \delta(x) \cos \delta(x)}{\rho(x)} = L(x) + \left(\frac{x}{x-1}\right)^m I(x) + \sum_{i=1}^m \frac{C_i}{(x-1)^i} + \sum_i \frac{R_i}{x_i - x}. \tag{B9}$$

Examining the leading terms of both sides, one deduces that either $\delta(1) = \pi\alpha$ or $\delta(1) = 0 \pmod{\pi}$.

In the latter case, write next

$$\phi(x) = (x-1)^{m-\alpha+\mu}\phi^*(x), \tag{B10}$$

$$\phi^*(x) \equiv \left(x^{-m}\frac{(x-1)^\alpha}{\rho(x)}\right)(x-1)^{-\mu} \sin^2 \delta(x). \tag{B11}$$

Assuming for the moment that $m - \alpha + \mu < 0$, we see from Lemma A5 that

$$(x-1)^{-(m-\alpha+\mu)} I(x)$$

is H.c. near and at $x = 1$ and vanishes there. Then balancing both sides of (9) tells us that the same is true of $(x-1)^{-\mu} \sin \delta(x)$.

Now write

$$\phi(x) = (x-1)^{m-\alpha+3\mu}\phi^*(x), \tag{B12}$$

$$\phi^*(x) \equiv \left(x^{-m}\frac{(x-1)^\alpha}{\rho(x)}\right)(x-1)^{-3\mu} \sin^2 \delta(x). \tag{B13}$$

Assuming that also $m - \alpha + 3\mu < 0$, we deduce again that $(x-1)^{-(m-\alpha+3\mu)} I(x)$ is H.c. near and at $u = 1$ and vanishes there and that the same is true of $(x-1)^{-3\mu} \sin \delta(x)$.

We can continue in this fashion until, for some n, $[m - \alpha + (2n+1)\mu] > 0$. Then, Lemma A5 tells us that I itself is H.c. near and at $x = 1$, and hence from (9) the same is true of

$$(x-1)^{m-\alpha} \sin \delta(x).$$

If this quantity does not vanish at $x = 1$, the proof is complete. If it does, then, from (9), the same must be true of I. Then we can take $m = a - 1$ and repeat the above arguments, if necessary, until $m = 0$.

APPENDIX C: BOUNDS ON AN INTEGRAL

Here we prove the statement in the text that the right-hand side of (106) vanishes as $\epsilon \to 0$. Write

$$\begin{aligned} h(uu') &\equiv k(uu')\left(\frac{u}{u'}\right)^{\frac{1}{2}|\gamma|+\delta} \\ &= \frac{u'^{-\gamma} l(u') - u^{-\gamma} l(u)}{u' - u}(u \text{ or } u')^\gamma \left(\frac{u}{u'}\right)^\delta (\sigma(u')\sigma(u))^{\frac{1}{2}} \\ &= \frac{l(u') - l(u)}{u' - u}\left(\frac{u}{u'}\right)^\delta [\sigma(u')\sigma(u)]^{\frac{1}{2}} \\ &\quad + [l(u') \text{ or } l(u)]\left(\frac{u'^{|\gamma|} - u^{|\gamma|}}{u' - u}\right) \\ &\quad \times u'^{-|\gamma|}\left(\frac{u}{u'}\right)^\delta [\sigma(u')\sigma(u)]^{\frac{1}{2}} \\ &\equiv h_1 + h_2. \end{aligned} \tag{C1}$$

Here, $\gamma \equiv 1 - \beta$, $l \equiv u^\beta L$, $\sigma \equiv u^{-\beta}\rho$, and the two alternatives hold for $0 < \beta \le 1$ and $1 \le \beta < 2$ (i.e., $\gamma \ge 0$ and $\gamma \le 0$), respectively. It is clear that both l and σ are H.c. near and at $u = 0$, that $\sigma(0) = 1$, and that, if $A = 0$ (as is being assumed), then $l(0) = 0$.

First, it will be shown that, as $u \to 0$,

$$\int_0^1 du' h(uu') \to 0 \tag{C2}$$

if $0 < \delta < \mu$. One has indeed, by writing $z' = u'/u$,

$$\int_0^1 du' \,|h_1| < \text{const} \int_0^1 du' \left(\frac{u}{u'}\right)^\delta |u' - u|^{\mu - 1} \sim u^\delta, \tag{C3}$$

$$\int_0^1 du' \,|h_2| < \text{const} \int_0^1 du' u'^{\mu - |\gamma|} \frac{u'^{|\gamma|} - u^{|\gamma|}}{u' - u}\left(\frac{u}{u'}\right)^\delta < \text{const}\, u^\delta. \tag{C4}$$

In an exactly similar way, it follows that

$$\int_0^\epsilon du' \, |h(uu')| < \Delta(\epsilon), \tag{C5}$$

where Δ is *independent of* u and tends to zero as $\epsilon \to 0$. This proves the required statement.

APPENDIX D: THE SINGULAR OPERATOR k_s

The procedure outlined in Sec. 5 [Eq. (128) *et seq.*] is given here in detail, the essential idea being due to Cottingham.[25] A brief discussion is also given of none-L_2 solutions of the integral equation.

Consider the variable

$$\theta = \log u. \tag{D1}$$

The interval $0 < u < 1$ maps onto $0 < \theta < \infty$. Since

$$\int_0^1 du \, |f(u)|^2 = \int_0^\infty d\theta \, |\hat{f}(\theta)|^2, \tag{D2}$$

where

$$\hat{f}(\theta) = e^{-\frac{1}{2}\theta} f(u[\theta]), \tag{D3}$$

the relations (1) and (3) map the set of functions f belonging to $L_2(0, 1)$ onto the set $\hat{f}$ belonging to $L_2(0, \infty)$. Obviously, if $k(\theta, \theta')$ is the kernel of an operator in $L_2(0\ 1)$, then

$$\hat{k}(\theta, \theta') = e^{-\frac{1}{2}\theta} e^{-\frac{1}{2}\theta'} k(u[\theta], u[\theta'])$$

is the kernel of the corresponding operator in $L_2(0, \infty)$.

In view of this, consider

$$\hat{k}_s(\theta, \theta') = e^{-\frac{1}{2}\theta} e^{-\frac{1}{2}\theta'} k_2(u[\theta], u[\theta']). \tag{D4}$$

From (114) or (115), $\hat{k}_s$ is actually only a function of $y \equiv \theta - \theta'$, and, in fact, if $\gamma = 0$ ($\beta = 1$),

$$\hat{k}_s(y) = \frac{A}{\pi} \frac{y}{\sinh \frac{1}{2} y}, \tag{D5}$$

or, if $\gamma \neq 0$,

$$\hat{k}_s(y) = \frac{A}{\pi} \frac{\sinh(\frac{1}{2}\gamma y)}{\sinh(\frac{1}{2} y)}. \tag{D6}$$

Now, consider the decomposition

$$\hat{k}_s = \hat{g} + \bar{k}_s, \tag{D7}$$

where

$$\hat{g}(\theta, \theta') = \hat{k}_s(\theta + \theta'), \tag{D8}$$

$$\bar{k}_s(\theta, \theta') = \hat{k}_s(\theta - \theta') - \hat{k}_s(\theta + \theta'). \tag{D9}$$

This is the decomposition referred to in Eq. (128) of the text.

The operator $\hat{g}$ is Hilbert–Schmidt (and is therefore compact) since

$$\int_0^\infty d\theta \int_0^\infty d\theta' \, |\hat{k}_s(\theta + \theta')|^2 = \int_0^\infty d\theta \int_\theta^\infty dy \, |\hat{k}_s(y)|^2, \tag{D10}$$

which is clearly finite for both (5) and (6).

Regarding $\bar{k}_s$, it will now be shown that the inverse $(1 - \lambda\bar{k}_s)^{-1}$ is bounded in $L_2(0, \infty)$ except on the exceptional line of Lemma 4, in accordance with the statement of the text. First, we need the following standard results.

Lemma D1: If $f(\theta)$ is in $L_2(0, \infty)$, so is

$$\tilde{f}(p) \equiv (2/\pi)^{\frac{1}{2}} \int_0^\infty d\theta \sin p\theta f(\theta), \tag{D11}$$

and, in addition,

$$f(\theta) = (2/\pi)^{\frac{1}{2}} \int_0^\infty dp \sin p\theta \tilde{f}(p). \tag{D12}$$

Lemma D2: Let $k(y)$ be an even function and be in $L_1(-\infty, \infty)$ and let $f(y)$ be in $L_2(0, \infty)$; define $\tilde{f}$ by (11) and define $\tilde{k}$ by

$$\tilde{k}(p) = (2/\pi)^{\frac{1}{2}} \int_0^\infty dy \cos py \, k(y). \tag{D13}$$

Then, $\tilde{k}$ is bounded, and the following hold:

$$(2/\pi)^{\frac{1}{2}} \int_0^\infty d\theta \sin p\theta \times \left(\int_0^\infty d\theta' [k(\theta - \theta') - k(\theta + \theta')] f(\theta') \right) = (2\pi)^{\frac{1}{2}} \tilde{k}(p) \tilde{f}(p), \tag{D14}$$

$$(2/\pi)^{\frac{1}{2}} \int_0^\infty dp \sin p\theta [(2\pi)^{\frac{1}{2}} \tilde{k}(p) \tilde{f}(p)] = \int_0^\infty d\theta' [k(\theta - \theta') - k(\theta + \theta')] f(\theta'). \tag{D15}$$

Lemma D1 is on p. 70 of Ref. 26, and Lemma D2 follows from p. 90 of this reference if we observe that

$$\int_0^\infty d\theta' [k(\theta - \theta') - k(\theta + \theta')] f(\theta') = \int_{-\infty}^\infty k(\theta - \theta') f(\theta') \tag{D16}$$

if f is defined for $\theta < 0$ by requiring it to be odd.

These results imply that the problem of finding solutions f, in $L_2(0, \infty)$, of the integral equation

$$f(\theta) = c(\theta) + \frac{\lambda}{\pi} \int_0^\infty (k(\theta - \theta') - k(\theta + \theta')) f(\theta') \tag{D17}$$

[with c in $L_2(0, \infty)$] is completely equivalent to that of finding solutions $\tilde{f}$, in $L_2(0, \infty)$, of

$$\tilde{f}(p) = \tilde{c}(p) + \lambda (2/\pi)^{\frac{1}{2}} \tilde{k}(p) \tilde{f}(p), \tag{D18}$$

where $\tilde{c}$ is related to c by (11) [and hence is in $L_2(0, \infty)$].

Since $\tilde{k}$ is bounded, there will be an L_2 solution for every $\tilde{c}$ in L_2 if and only if the equation

$$1 - \lambda(2/\pi)^{\frac{1}{2}}\tilde{k}(p) = 0 \tag{D19}$$

has no solution for $0 < p < \infty$.

Now, returning to the special problem at hand, it is seen that $k \equiv \bar{k}_s$ is indeed even and is in $L_1(-\infty, \infty)$. One has,[27] if $\gamma = 0$,

$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\tilde{k} = A\frac{1}{\cosh^2 \pi p} \tag{D20}$$

or, if $\gamma \neq 0$,

$$\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\tilde{k} = A\frac{2\sin\pi\gamma}{\cos\pi\gamma + \cosh\pi p}, \tag{D21}$$

and, using these expressions, we can easily verify that (19) has a solution if and only if λ lies on the exceptional line of Lemma 4 of the text. This completes the proof of the statements made in the text.

Next, let us make a heuristic examination of the possibility that the equation $n = b + kn$ has solutions not in L_2. This may be done by looking for solutions (not in L_2) of

$$n_0(\theta) = \int_0^\infty d\theta' \bar{k}_s(\theta, \theta') n_0(\theta'). \tag{D22}$$

If Eq. (D19) has a solution $p = P(\lambda)$, then, when P is real (i.e., when λ is on the exceptional line), it is clear that there is a solution

$$n_0(\theta) = \text{const} \sin P\theta. \tag{D23}$$

[This follows from the orthogonality condition

$$\tfrac{1}{2}\pi\int_0^\infty \sin p\theta \sin P\theta' \, d\theta = \delta(p - P).] \tag{D24}$$

If λ is now moved off the exceptional line so that P becomes complex, one expects $\sin P\theta$ to remain a solution of (D22) as long as the integral

$$\int_0^\infty d\theta' \bar{k}_s(\theta, \theta') \sin P\theta' \tag{D25}$$

converges. This requires that

$$|\text{Im}\, P| < \tfrac{1}{2}(1 - \gamma), \tag{D26}$$

and it may be shown that this is satisfied throughout the λ place except on the cut of the function $P(\lambda)$; this is the line $-\infty \leq \lambda \leq 0$ or $0 \leq \lambda \leq \infty$ for $\gamma \geq 0$ or $\gamma < 0$, respectively (Figs. 2 and 3).

Thus, we expect that the general solution of the equation $n = b + \lambda kn$ is, for fixed b in L_2,

$$n = n_1(\lambda) + cn_2(\lambda), \tag{D27}$$

where n_1 is in L_2 (and has a cut along the exceptional line of Lemma 4), n_2 is not in L_2 (and has a cut along the line just mentioned), and c is an arbitrary parameter.

These results agree with those of Refs. 5 and 7. Of course, nothing has been proved rigorously about none-L_2 solutions (here or in Refs. 5 and 7), and, of course, these solutions may not be the only ones.

However, in the special case that λ lies between the two lines mentioned above, an N/D argument also suggests that there is indeed a 1-parameter infinity of none-L_2 solutions of (96) of the text and that these are the only ones (apart from the unique L_2 solution) which give rise to allowed solutions F of the problem.

To see this, consider the N/D equations. In the text (Secs. 4 and 6) it has been shown that any D function has, for $A \neq 0$, the asymptotic behavior

$$D \sim x^{c - \frac{1}{2}p}, \tag{D28}$$

with p an integer and c in the range $\frac{1}{2}(\beta - 1) < c < \frac{1}{2}\beta$. It was shown that, if one requires N and D to satisfy the N/D equations *with* $(\rho/u)^{\frac{1}{2}}N$ *in* L_2, then one must have $p \geq 0$. However, an examination of the asymptotic behavior of D and χ (Sec. 4) reveal that, for $A > 0$ or < 0, for $0 < \beta \leq 1$ or $1 < \beta < 2$, respectively, one can have the equations satisfied in addition for $p = -1$, provided the L_2 restriction is dropped. [Instead of being in L_2, $(\rho/u)^{\frac{1}{2}}N$ is not in L_q for some q in the range $1 < q < 2$.]

Looking now at Eq. (161) of the text, we find that, if there are no coincident zeros, $p = -1$ implies that

$$\kappa = 2\chi + 1. \tag{D29}$$

In particular, if $\chi = 0$, $\kappa = 1$; hence, we deduce that the $\chi = 0$ equation

$$n = (\rho/u)^{\frac{1}{2}}L + \int kn \tag{D30}$$

has a 1-parameter infinity of none-L_2 solutions (corresponding to the 1-parameter infinity of allowed solutions with $\kappa = 1$—we are assuming here that $\kappa_0 = 0$ in Theorem 10 of the text).

Since values of λ between the lines mentioned above can all be attained by redefining $L \to \lambda L$, with the new values of A in the range $0 < A < A_{\max}$ or $A_{\min} < A < 0$ [for $0 < \beta \leq 1$ or $1 < \beta < 2$, respectively], this substantiates the statement just made. We notice that the other half of the allowed range ($A_{\min} < A < 0$ and $0 < A < A_{\max}$, respectively), where the N/D argument yields no none-L_2 solution, corresponds always to λ on the line where $\sin P\theta$ is no longer a solution of (22). This gives us confidence that the 1-parameter infinity of none-L_2 solutions found by the N/D approach is indeed the same as those found from the Fourier transform method just given.

APPENDIX E: GOOD BEHAVIOR OF N AND D FUNCTIONS

Here it is shown how to prove Lemma 6, parts (b) and (c) of the text, and also the statement in part (d) that $\bar{N}$ is in L_2.

First, some results are needed about the general function spaces $L_p(0, 1)$, i.e., sets of functions such that

$$\int_0^1 du\,|f(u)|^p$$

exists, with $1 \leq p < \infty$. We shall just write L_p for (0, 1).

Lemma E1: If $f \in L_q$, then $f \in L_n$ for all n in the interval $1 \leq n \leq q$.

Proof:

$$\int |f|^q = \int (|f|^n)^{q/n}, \tag{E1}$$

so $|f|^n$ is $L_{(q/n)}$. Obviously, the function $g(u) = 1$ is in L_α for all α, hence, in particular for the α such that

$$\frac{1}{\alpha} + \frac{1}{q/n} = 1. \tag{E2}$$

Hence, by a standard result (Ref. 22, p. 42) the integral

$$\int 1 \cdot |f|^n$$

exists,which shows that $f \in L_n$.

Lemma E2: Let

$$\int_0^1 du'\,|K(uu')|^m < \text{const} \tag{E3}$$

for $0 < u < 1$ and some m such that $1 < m \leq 2$, and define n by

$$1/m + 1/n = 1. \tag{E4}$$

Then, for any $q \geq n\ (\geq 2)$, $f \in L_q$ implies

$$\left| \int_0^1 du' K(uu')f(u') \right| < \text{const}. \tag{E5}$$

Proof: Use Lemma E1 and the standard result

$$\left| \int Kf \right| \leq \left[\int |K|^m \right]^{1/m} \left[\int |f|^n \right]^{1/n}. \tag{E6}$$

Lemma E3: Let S be some positive function, and let K satisfy

$$\int_0^1 du' \left(\frac{S(u)}{S(u')} \right)^p |K(u, u')|^{p/r} < \text{const}, \tag{E7}$$

$$\int_0^1 du \left(\frac{S(u')}{S(u)} \right)^s |K(u, u')| < \text{const}, \tag{E8}$$

for some p, r, and s such that

$$1 < r < p \leq 2, \tag{E9}$$

$$1/r + 1/s = 1 \tag{E10}$$

(hence, $s \geq 2$ and $1 < p/r < 2$). Define q by

$$1/p + 1/q = 1 \tag{E11}$$

(hence $s \geq q \geq 2$). Then $f \in L_q$ implies that

$$\int_0^1 du' K(uu')f(u') \tag{E12}$$

is in L_s.

Proof: We follow the idea of Ref. 23, making use of the inequality (E6) at every step. The theorem is proved if it can be shown that $f \in L_q$ implies that

$$\sup_{\|g\|_r = 1} \int_0^1 du \int_0^1 du' g(u) K(uu') f(u') < \infty, \tag{E13}$$

where

$$\|g\|_r \equiv \int_0^1 du\,|g|^r. \tag{E14}$$

Write the integrand as

$$\left(g^{r/p}(u) \frac{S(u)}{S(u')} K^{1/r}(uu') \right) \times \left(g^{1-r/p}(u) \frac{S(u')}{S(u)} K^{1/s}(uu') f(u') \right),$$

so that the integral is bounded by

$$\left[\int du \int du'\,|g(u)|^r \left(\frac{S(u)}{S(u')} \right)^p K^{p/r}(uu') \right]^{1/p} \times \left[\int du \int du'\,|g(u)|^{q(1-r/p)} \left(\frac{S(u')}{S(u)} \right)^q \times K^{q/s}(uu')\,|f(u')|^q \right]^{1/q}.$$

The first term is bounded (since $\|g\|_n = 1$) by

$$\text{const} \sup_{0<u<1} \int du' \left(\frac{S(u)}{S(u')} \right)^p K^{p/r}(uu').$$

For the second, define u and t by

$$u = s/q, \tag{E15}$$

$$1/u + 1/t = 1, \tag{E16}$$

so that

$$tq(1 - r/p) = r. \quad \text{(E17)}$$

Then, since $\|f\|_q$ and $\|g\|_r$ are finite, the second term is bounded by

$$\text{const} \sup_{0<u'<1} \left[\int du\, |g(u)|^{q(1-r/p)} \left(\frac{S(u')}{S(u)}\right)^q K^{q/s}(uu') \right]$$
$$\leq \text{const} \sup_{0<u'<1} \int du \left(\frac{S(u')}{S(u)}\right)^s |K(uu')|. \quad \text{(E18)}$$

This completes the proof.

Using the preceding results, we can prove the following.

Lemma E4: Let m be some fixed number, $1 < m < 2$. Let δ take on the values $\delta_{\min} \leq \delta \leq \delta_{\max}$ (with $\delta_{\min} \geq 0$), and suppose there exists a continuous function $a(\delta)$, with

$$a(\delta_{\min}) = (2/m) - 1, \quad \text{(E19)}$$

$$a(\delta_{\max}) = 0, \quad \text{(E20)}$$

such that the kernel k satisfies

$$\int_0^1 du' \left| \left(\frac{u}{u'}\right)^{ac} \left(\frac{u}{u'}\right)^{\delta} k(uu') \right|^m < \text{const}, \quad \text{(E21)}$$

$$\int_0^1 du \left| \left(\frac{u'}{u}\right)^{c} \left(\frac{u}{u'}\right)^{\delta} k(uu') \right| < \text{const}, \quad \text{(E22)}$$

for some $c(\delta)$. For any function $f(u)$ and any integer $n > 0$, define

$$(k^n f)(u) = \int_0^1 du_{n-1} k(u, u_{n-1}) \int_0^1 du_{n-2}$$
$$\times\, k(u_{n-1}, u_{n-2}) \cdots \int_0^1 du_1 k(u_2 u_1) f(u_1). \quad \text{(E23)}$$

Then, for every f in L_2 and some n,

$$(k^n f)(u) < \text{const}\, u^{-\delta_{\max}}. \quad \text{(E24)}$$

Proof: First, observe that

$$u^{\delta_{\min}}(kf)(u) = \int_0^1 du' \left(\frac{u}{u'}\right)^{\delta_{\min}} k(uu') u'^{\delta_{\min}} f(u') \quad \text{(E25)}$$

and that $u^{\delta_{\min}} f$ is certainly in L_2 if f is. In Lemma 3, take $K = (u/u')^{\delta_{\min}} k$, $p = 2$, $r = p/m$, and $S(u) = u^{c/s}$. Then, if we choose $a = (2/m) - 1$ $(= r/s)$, conditions (21) and (22) are identical with (7) and (8); then, Lemma 3 tells us that $u^{\delta_{\min}}(kf)(u)$ is in L_s where $s = (1 - \frac{1}{2}m)^{-1}$.

Next, choose $\delta = \delta_2$ $(> \delta_{\min})$ such that it is possible to have $a = (2/m^2) - 1$, write

$$u^{\delta_2}(k^2 f)(u) = \int_0^1 du' \left(\frac{u}{u'}\right)^{\delta_2} k(uu') [u'^{\delta_2}(kf)(u')], \quad \text{(E26)}$$

and observe that the square bracket is certainly in L_s. Now, use Lemma E3 again, with $K = (u/u')^{\delta_2} k$, $p = r_1 \equiv (2/m)$, $r = (r_1/m)$ $[= (2/m^2)]$, and $S(u) = u^{c/s}$. Then, with $a = (2/m^2) - 1$, conditions (21) and (22) are again identical with (7) and (8); hence, the lemma tells us that $u^{\delta_2}(k^2 f)(u)$ is in L_s where $s = (1 - \frac{1}{2}m^2)^{-1}$.

We may continue in this way until we deduce that $u^{\delta_{\max}}(k^{n-1} f)(u)$ is in L_s, where $s = (1 - \frac{1}{2}m^{(n-1)})^{-1}$ is sufficiently large that r $(= 2/m^{(n+1)})$ is less than m. Then, finally, Lemma 3 [with $K = (u/u')^{\delta_{\max}} k$ and choosing $a = 0$ so that condition (21) is identical with (7)] implies that $u^{\delta_{\max}}(k^n f)(u)$ is bounded. This proves the lemma. Finally, this result can be used to make a statement about integral equations.

Lemma E5: Let k satisfy the conditions of Lemma E4, let $b(u)$ be in L_2, and let $|b| <$ const $u^{-\delta_{\max}}$. Then, any L_2 solution of

$$f = b + \int_0^1 du' k(uu') f(u') \quad \text{(E27)}$$

satisfies $|f| <$ const $u^{-\delta_{\max}}$.

Proof: Eq. (E27) implies that

$$u^{\delta_{\max}} f(u) = u^{\delta_{\max}}[(kb)(u) + (k^2 b)(u) + \cdots (k^{n-1} b)(u)]$$
$$+\, u^{\delta_{\max}}(k^n f)(u). \quad \text{(E28)}$$

Lemma E4 says that the last term is bounded, so consider the terms involving b. Since $[u^{\delta_{\max}} b(u)]$ is bounded, it is in L_p for all p; hence, we may use Lemma E2 to deduce that

$$u^{\delta_{\max}}(kb)(u) = \int_0^1 du' \left(\frac{u}{u'}\right)^{\delta_{\max}} k(uu') [u'^{\delta_{\max}} b(u')] \quad \text{(E29)}$$

is bounded [with $a = 0$, condition (E21) is identical with condition (E7)]. Then, we may use the lemma again to show that $u^{\delta_{\max}}(k^2 b)(u)$ is bounded, and so on, for all the terms involving b.

Now, we apply this general result to the problem at hand. The following will be proved.

Lemma E6: For $0 < \beta < 2$ and $A = 0$, any L_2 solution of Eq. (96) of the text satisfies

$$|u^{\frac{1}{2}|\gamma|} n(u)| < \text{const}, \quad \text{(E30)}$$

where $\gamma = 1 - \beta$.

Proof: It is clear that $u^{\frac{1}{2}|\gamma|} b$ is bounded. It will be shown that k satisfies the conditions of Lemma E4 with $\delta_{\min} = 0$ and $\delta_{\max} = \frac{1}{2}|\gamma|$.

In the notation of Appendix C, write

$$k = k_1 + k_2, \tag{E31}$$

$$k_1(u, u') \equiv \frac{l(u') - l(u)}{u' - u}\left(\frac{u}{u'}\right)^{-\frac{1}{2}\gamma}[\sigma(u')\sigma(u)]^{\frac{1}{2}}, \tag{E32}$$

$$k_2(u, u') \equiv \frac{u'^{\gamma} - u^{\gamma}}{u' - u}\, l(u')(u'u)^{-\frac{1}{2}\gamma}[\sigma(u')\sigma(u)]^{\frac{1}{2}}. \tag{E33}$$

Because of the standard inequality

$$\left(\int |f + g|^m\right)^{1/m} \leq \left(\int |f|^m\right)^{1/m} + \left(\int |g|^m\right)^{1/m}, \tag{E34}$$

it will be enough to verify that k_1 and k_2 separately satisfy the conditions of Lemma E4, with the same choice of m, $a(\delta)$, and $c(\delta)$.

For k_1, we have

$$\begin{aligned}\int_0^1 du \left|\left(\frac{u'}{u}\right)^c \left(\frac{u}{u'}\right)^1 k_1(uu')\right| &< \text{const} \int_0^1 du \left(\frac{u}{u'}\right)^{\delta - c - \frac{1}{2}\gamma} |u' - u|^{\mu - 1} \\ &= \text{const}\, u'^{\mu} \int_0^{1/u'} dy\, y^{\delta - c - \frac{1}{2}\gamma} |y - 1|^{\mu - 1},\end{aligned} \tag{E35}$$

which is bounded if

$$\delta - \tfrac{1}{2}\gamma < c < 1 + \delta - \tfrac{1}{2}\gamma. \tag{E36}$$

Similarly, choosing $1/m = 1 - (\mu - \epsilon)$ with

$$0 < (\mu - \epsilon) < \tfrac{1}{2},$$

we see that

$$\begin{aligned}\int_0^1 du' \left|\left(\frac{u}{u'}\right)^{ac} \left(\frac{u}{u'}\right)^{\delta} k_1(uu')\right|^m &< \text{const}\, u^{1-(1-\mu)m} \int_0^{1/u} dy\, y^{-m(\delta + ac - \frac{1}{2}\gamma)} |y - 1|^{m(\mu - 1)}\end{aligned} \tag{E37}$$

is bounded if

$$-(\delta - \tfrac{1}{2}\gamma) \leq ac < 1 - \mu + \epsilon - (\delta - \tfrac{1}{2}\gamma). \tag{E38}$$

These inequalities are compatible if

$$-\frac{(\delta - \frac{1}{2}\gamma)}{1 + (\delta - \frac{1}{2}\gamma)} \leq a < \frac{1 - \mu + \epsilon - (\delta - \frac{1}{2}\gamma)}{(\delta - \frac{1}{2}\gamma)} \text{ or } \infty, \tag{E39}$$

for $\delta > \frac{1}{2}\gamma$ and $\delta \leq \frac{1}{2}\gamma$, respectively (recall that $-1 < \gamma < 1$ and $\delta \geq 0$).

For k_2, we may similarly use

$$|k_2| < \frac{u'^{\gamma} - u^{\gamma}}{u' - u}\, u'^{\mu}(u'u)^{-\frac{1}{2}\gamma} \tag{E40}$$

to deduce that

$$\int_0^1 du \left|\left(\frac{u'}{u}\right)^c \left(\frac{u}{u'}\right)^{\delta} k_2(u, u')\right| \tag{E41}$$

is bounded if

$$(\delta + \tfrac{1}{2}\gamma) - \mu < c < 1 + (\delta - \tfrac{1}{2}\gamma) \tag{E42}$$

or

$$(\delta - \tfrac{1}{2}\gamma) - \mu < c < 1 + (\delta + \tfrac{1}{2}\gamma), \tag{E43}$$

for $\gamma \geq 0$ and $\gamma \leq 0$, respectively, and that

$$\int_0^1 du' \left|\left(\frac{u}{u'}\right)^{d} \left(\frac{u}{u'}\right)^{\delta} k(uu')\right|^m \tag{E44}$$

is bounded if

$$-(\delta - \tfrac{1}{2}\gamma) \leq ac < 1 + \epsilon - (\delta + \tfrac{1}{2}\gamma) \tag{E45}$$

or

$$-(\delta + \tfrac{1}{2}\gamma) \leq ac < 1 + \epsilon - (\delta - \tfrac{1}{2}\gamma), \tag{E46}$$

for the two cases. These inequalities are compatible for the two cases if

$$-\frac{(\delta - \frac{1}{2}\gamma)}{1 + (\delta - \frac{1}{2}\gamma)} \leq a < \frac{1 + \epsilon - (\delta + \frac{1}{2}\gamma)}{(\delta + \frac{1}{2}\gamma) - \mu} \text{ or } \infty \tag{E47}$$

(for $\delta + \frac{1}{2}\gamma > \mu$ or $\leq \mu$) and

$$-\frac{(\delta + \frac{1}{2}\gamma)}{1 + (\delta + \frac{1}{2}\gamma)} < a < \frac{1 + \epsilon - (\delta - \frac{1}{2}\gamma)}{(\delta - \frac{1}{2}\gamma) - \mu} \text{ or } \infty \tag{E48}$$

(for $\delta - \frac{1}{2}\gamma > \mu$ or $\leq \mu$).

It is now straightforward (though very tedious) to verify that, provided we choose $\mu - \epsilon < (1 - |\gamma|)/(2 - |\gamma|)$, the bounds for c and a are compatible for k_1 and k_2 and that one may choose $a\,(\delta = 0) = (2/m) - 1$ and $a\,(\delta = \frac{1}{2}|\gamma|) = 0$. This ends the proof of Lemma E6. From the above result, it is easy to deduce the following.

Lemma E7: For $0 < \beta \leq 1$ and $A = 0$, $u^{-\beta}N$ is bounded. For $1 \leq \beta < 2$ and $A = 0$, $u^{-1}N$ is bounded.

Proof: Since $n = (\rho/u)^{\frac{1}{2}}N$, one has

$$(u^{-\beta} \text{ or } u^{-1})N = u^{\frac{1}{2}|\gamma|}n/\sigma^{\frac{1}{2}}, \tag{E49}$$

for $0 < \beta \leq 1$ and $1 \leq \beta < 2$, respectively. It has to be shown that the vanishing of σ as $u \to 1$ causes no infinity in this expression. One has

$$\begin{aligned}&(u^{-\beta} \text{ or } u^{-1})N(u) \\ &\quad = (u^{-\beta} \text{ or } u^{-1})B(u) + \int_0^1 du' \frac{H(u, u')}{[\sigma(u')]^{\frac{1}{2}}}[u'^{\frac{1}{2}|\gamma|}n(u')],\end{aligned} \tag{E50}$$

where

$$\begin{aligned}H(u, u') &= \left(\frac{u}{u'}\right)^{\frac{1}{2}|\gamma|}\left(\frac{\sigma(u')}{\sigma(u)}\right)^{\frac{1}{2}} k(uu') \\ &= \frac{u'^{-\gamma}l(u') - u^{-\gamma}l(u)}{u' - u}(u \text{ or } u')^{\gamma}\sigma(u').\end{aligned} \tag{E51}$$

In the course of proving Lemma E6, it was shown that the kernel $[(u/u')^{\frac{1}{2}|\gamma|}k(uu')]$ satisfies the condition of Lemma E2; $H/[\sigma(u')]^{\frac{1}{2}}$ differs from this only in that the bounded function $[\sigma(u')\sigma(u)]^{\frac{1}{2}}$ is replaced by the bounded function $[\sigma(u')]^{\frac{1}{2}}$; hence, $H/[\sigma(u')]^{\frac{1}{2}}$ also satisfies the condition of Lemma E2, and so we deduce that the integral in (E50) is bounded. This completes the proof.

Next we want to prove that $\mu^{-\beta}N$ or $u^{-1}N$ is H.c. in $0 < u < 1$. This is a direct consequence of the following [using (E50)].

Lemma E8: Let $X(u)$ be bounded and defined $H(u, u')$ by (E51). Then, for $0 < \beta < 2$ and $A = 0$,

$$\int_0^1 du' H(uu')X(u')$$

is H.c. in $0 < u < 1$, with Hölder index λ equal to $\mu + \epsilon$ if $\gamma = 0$ or equal to $\epsilon + \min(\mu, |\gamma|)$ otherwise, with $\epsilon > 0$.

Proof: It is obviously enough to establish the inequality

$$\int_0^1 du' \, |H(u_1u') - H(u_2u')| < \text{const}\, |u_1 - u_2|^\lambda. \quad \text{(E52)}$$

Let us write

$$H = A + B, \quad \text{(E53)}$$

$$A \equiv \frac{l(u') - l(u)}{u' - u}\, \sigma(u'), \quad \text{(E54)}$$

$$B \equiv \frac{u'^{|\gamma|} - u^{|\gamma|}}{u' - u}\, u'^{-|\gamma|}[l(u') \text{ or } l(u)]\sigma(u'), \quad \text{(E55)}$$

where the first and second alternatives will always refer to $\beta \le 1$ and $\beta \ge 1$, respectively. The inequality (E52) is proved for A in Ref. 17, p. 47. For B we follow p. 79 of that reference. We define

$$I \equiv |B(u_1u') - B(u_2u')|, \quad \text{(E56)}$$

$$\Delta \equiv u_2 - u_1, \quad \text{(E57)}$$

$$u \equiv u_1, \quad \text{(E58)}$$

$$a \equiv |\gamma|, \quad \text{(E59)}$$

and write

$$\int_0^1 = \int_0^{u-2\Delta} + \int_{u+2\Delta}^1 + \int_{u-2\Delta}^{u+2\Delta}. \quad \text{(E60)}$$

(If $u - 2\Delta < 0$ or $u + 2\Delta > 0$, the first or second terms will be absent, and there will be some obvious modifications of the argument.)

Consider the last term. It has the bound

$$\int_{u-2\Delta}^{u+2\Delta} du' I(u_1, u_2, u') < \text{const} \int_{u-2\Delta}^{u+2\Delta} du' \times u'^{-a}(|u' - u|^{a-1} + |u' - (u + \Delta)|^{a-1})(u'^\mu \text{ or } u^\mu), \quad \text{(E61)}$$

which may be seen on making the substitution $y = u'/u$ to be bounded by Δ^λ as required.

For the other two terms, use the identity

$$\frac{u'^a - u^a}{u' - u} - \frac{u'^a - (u+\Delta)^a}{u' - (u+\Delta)} \equiv -\frac{u^a - (u+\Delta)^a}{u' - u} + \frac{[(u+\Delta)^a - u'^a]\Delta}{[u' - u][u' - (u+\Delta)]}. \quad \text{(E62)}$$

The first term gives a contribution to $\int_{u+2\Delta}^1 du' I$ bounded by

$$\Delta^a \int_{u+2\Delta}^1 du' \frac{u'^{-a}}{u' - u} (u' \text{ or } u)^\mu,$$

which may be shown to be bounded by $\Delta^a \log \Delta$ or Δ^μ (for $a \le \mu$ and $a > \mu$, respectively). The contribution to $\int_0^{u-2\Delta}$ may be treated similarly. The second term gives a contribution to $\int_{u+2\Delta}^1 du' I$ bounded by

$$\text{const}\, \Delta \int_{u+2\Delta}^1 du' \frac{u'^{-2}}{u' - u} |u' - (u+\Delta)|^{a-1} (u' \text{ or } u)^\mu < \text{const}\, \Delta^a \int_{u+2\Delta}^1 du' \frac{u'^{-a}}{u' - a} (u' \text{ or } u)^\mu, \quad \text{(E63)}$$

which is the same as the above expression, and similarly for its contribution to $\int_0^{u-2\Delta}$. This completes the proof.

Knowing that $u^{-\beta}N$ or $u^{-1}N$ is H.c., it is now possible to investigate the corresponding D function. The procedure is slightly different for $\beta \le 1$ and $\beta > 1$, and we just give the argument for the first case.

For $\beta \le 1$, we have

$$u^{-\beta}N = u^{-\beta}B + \int_0^1 du\, H(uu')[u'^{-\beta}N(u')]. \quad \text{(E64)}$$

The first term vanishes at $u = 0$, and, by decomposing the integral into two principal-valued integrals and using Lemma A5, it may be seen that the same is true of the integral (remember that $A = 0$ throughout). Hence the quantity $u^{-\beta}N$ is zero at $u = 0$.

Now, consider

$$D(u) = 1 - \frac{1}{\pi} \int_0^1 du' \frac{[u'^{-\beta}N(u')]\sigma(u')}{u' - u} + \text{(pole terms)}. \quad \text{(E65)}$$

From Lemma A5, D has the properties listed in parts (a) and (b) of Lemma 5 of the text. If D does not vanish at the end points $u = 0$ and $u = 1$, then part (c) of this lemma will also be satisfied. If D does vanish, further work needs to be done. The procedure will be given in detail only for the more difficult case $u = 0$. We need the following result, which may be proved by using the decomposition (E53) and a similar one for $\hat{H}$.

Lemma E9: Let h be either equal to H as defined by (E51) or to

$$\hat{H} = \frac{u'^{\beta}l(u') - u^{\beta}l(u)}{u' - u}\, u'^{-\beta}\sigma(u'). \tag{E66}$$

Then, for $0 < \beta \leq 1$ and $A = 0$, any positive $\delta < 1$ and any bounded function X,

$$\int_0^1 du' h(u, u')\left(\frac{u}{u'}\right)^{\delta} X(u') < \text{const}\, u^{\lambda+\epsilon} \tag{E67}$$

(all $\epsilon > 0$), where $\lambda = \min(\mu, \delta)$.

We also need the extension of Lemma E8 to $\hat{H}$, which may be proved in the same way as that lemma.

Lemma E10: Let X be bounded and define $\hat{H}$ by (E66). Then, for $0 < \beta < 1$ and $A = 0$,

$$\int_0^1 du' \hat{H}(u, u') X(u') \tag{E68}$$

is H.c. in $0 < u < 1$ with Hölder index λ equal to $\epsilon + \min(\mu, \beta)$, with $\epsilon > 0$.

We now proceed as follows. When D vanishes at $u = 0$, then, by substituting into the integral equation for N the identity

$$-\frac{u'}{u}\frac{1}{u' - u} = \frac{1}{u' - u} - \frac{1}{u}, \tag{E69}$$

we deduce that

$$N_1(u) = B_1(u) + \int \hat{H}(u, u') N_1(u'), \tag{E70}$$

where $N_1 \equiv N/u$, and

$$B_1 = \sum_{i=1}^{\chi}\left(c_i \frac{d_i}{u - c_i} + c_i \frac{k_i}{u - c_i}\right), \tag{E71}$$

where $c_i \equiv b_i^{-1}$.

Then, by repeated use of Lemma E9 (with $\delta = 1 - \beta - \mu,\ 1 - \beta - 2\mu, \cdots$), we deduce that N_1 is bounded. Then Lemma E10 tells us that N_1 is H.c. in $0 < u < 1$.

If $N_1(0) \neq 0$, then, by using

$$D_1(u) \equiv \frac{D(u)}{u} = \frac{1}{\pi}\int_0^1 du' \frac{u'^{-\beta}N_1(u')\sigma(u')}{u' - u} + (\text{pole terms}) \tag{E72}$$

together with Lemma A5, we again have a behavior of D near $u = 0$ which is in accordance with Lemma 5, part (c), of the text.

If $N_1(0) = 0$, then, by using identities like (E69), one can show that

$$\bar{N}_1 \equiv (u - c_{\chi})N_1, \tag{E73}$$

$$\bar{D}_1 = (u - c_{\chi})D_1 \tag{E74}$$

satisfy N/D equations (E50) and (E65) with only $(\chi - 1)$ pole terms. Now, the whole cycle from Lemma E7 onwards may be repeated, until either $D_1(0)$ or $N_1(0)$ is nonzero, or else we arrive at the $\chi = 0$ equations (no pole terms). In the latter case, each new cycle produces a new linearly independent L_2 solution of the homogeneous form of Eq. (96); hence, from Theorem 7, the cycle will eventually terminate [by $D_1(0)$ or $N_1(0)$ failing to vanish]. Thus, in all cases we end up with the behavior of Lemma 5 of the text, as required.

The other cases [i.e., the end point $u = 1$, the case $1 < \beta < 2$, and the proof that $(\rho/u)^{\frac{1}{2}}\bar{N}$ is in L_2 in part (d) of Lemma 5] may all be dealt with by similar arguments.

APPENDIX F: GENERAL N/D DECOMPOSITION

In this appendix the most general N/D decomposition corresponding to a given amplitude is exhibited explicitly, so that the number of parameters appearing for any particular case can be seen by inspection. This is necessary for Sec. 7.

(*a*) *Case where* $\kappa \geq -1$. Since $\kappa = 2\chi - 2q - p$ [Eq. (161) of the text], it is clear that there is a decomposition N_0/D_0 with $q = 0$, $p = p_0 = 0$ or 1, and $\chi = \chi_0 = \frac{1}{2}\kappa$ or $\frac{1}{2}(\kappa + 1)$ (according to whether κ is even or odd). Since $q = 0$, the rational function Φ of Sec. 4 can have no zeros, and it must therefore be completely determined by the pole positions $b_1 \cdots b_{\chi}$. Hence N_0/D_0 is the *only* decomposition with $q = 0$ and $p = 0$ or 1 [apart from the normalization factor if $D_0(0) = 0$]. Remembering that one always has $q \geq 0$ and $p \geq 0$, we see that the most general N/D decomposition for any $\chi > \chi_0$ is therefore given by $N = \Psi N_0$, $D = \Psi D_0$ with Ψ as follows.

(i) If $D_0(0) = 1$ and $D(0) = 1$,

$$\Psi = \frac{\prod_{i=1}^{q} (1 - x/a_i)}{\prod_{i=\chi_0+1}^{\chi} (1 - x/b_i)}, \quad q \leq \chi - \chi_0. \qquad \text{(F1)}$$

(ii) If $D_0(0) = 1$ and $D(0) = 0$,

$$\Psi = \lambda x \frac{\prod_{i=1}^{q} (a_i - x)}{\prod_{i=\chi_0+1}^{\chi} (b_i - x)}, \quad q \leq \chi - \chi_0 - 1. \qquad \text{(F2)}$$

(iii) If $D_0(0) = 0$ and $D(0) = 0$,

$$\Psi = \lambda \frac{\prod_{i=1}^{q} (a_i - x)}{\prod_{i=\chi_0+1}^{\chi} (1 - x/b_i)}, \quad q \leq \chi - \chi_0. \qquad \text{(F3)}$$

(iv) It is impossible to have $D_0(0) = 0$ and $D(0) = 1$ without choosing one of the b_i to be zero, which we agreed not to do (in Theorem 6).

(*b*) *Case where* $\kappa < -1$. In this case there is a unique decomposition N_0/D_0 with $\chi = \chi_0 \equiv 0, q = 0$, and $p = p_0 \equiv |\kappa|$. The most general decomposition with $\chi > 0$ is $N = \Psi N_0$ and $D = \Psi D_0$, with Ψ given by exactly the same expressions as above for the cases (i), (ii), and (iii) [(iv) not being allowed]; but the inequalities on q are now

(i) $q \leq \chi + \frac{1}{2}|\kappa|$ or $q \leq \chi + \frac{1}{2}(|\kappa| - 1)$,
(ii) $q \leq \chi + \frac{1}{2}|\kappa| - 1$ or $q \leq \chi + \frac{1}{2}(|\kappa| - 1) - 1$,
(iii) $q \leq \chi + \frac{1}{2}|\kappa|$ or $q \leq \chi + \frac{1}{2}(|\kappa| - 1)$,

where the alternatives for κ are even and odd, respectively.

The number of parameters appearing for any values of κ, χ, and q can now be read off from the appropriate expression.

[1] D. H. Lyth, "Use of the *N*/*D* Method in Approximate Calculations (Arbitrary $J > 0$)," Lancaster University (UK) preprint, 1968.

[2] G. Chew and S. Mandelstam, Phys. Rev. **119**, 467 (1960); Nuovo Cimento **19**, 752 (1961).

[3] G. Frye and R. L. Warnock, Phys. Rev. **130**, 478 (1963); R. L. Warnock, *ibid.* **131**, 1320 (1963).

[4] D. Atkinson and D. Morgan, Nuovo Cimento **41**, A559 (1966).

[5] D. Atkinson and A. P. Contogouris, Nuovo Cimento **39**, 1082, 1102 (1965).

[6] D. Morgan, Nuovo Cimento **36**, 813 (1965).

[7] D. Atkinson, J. Math. Phys. **7**, 1607 (1966).

[8] A. P. Balachandran and F. Von Hippel, Ann. Phys. (N.Y.) **30**, 446 (1964); A. P. Balachandran, *ibid.* **35**, 209 (1965).

[9] D. Atkinson, J. Math. Phys. **8**, 2281 (1967).

[10] C. Lovelace, Commun. Math. Phys. **4**, 261 (1967).

[11] R. L. Warnock, *Lectures in Theoretical High Energy Physics*, H. Aly, Ed. (Interscience, New York, 1968), Chap. 10.

[12] G. Rasche and W. S. Woolcock, Phys. Rev. **157**, 1473 (1967).

[13] M. B. Halpern, Phys. Rev. **140**, B1570 (1965); J. Math. Phys. **7**, 1226 (1966).

[14] R. J. Eden, *Particle Interactions at High Energies* (Oliver and Boyd, Edinburgh, 1966), p. 59.

[15] D. H. Lyth, Phys. Rev. Letters **17**, 830 (1966).

[16] Recently two alternatives to the *N*/*D* method have been invented: S. Ciulli, Gr. Ghika, M. Stihi, and M. Visinescu, Phys. Rev. **154**, 345 (1967), and G. Wanders and F. Reuse (Lausanne preprint); the second of these preserves the symmetry between the left- and right-hand cuts.

[17] N. I. Muskhelishvili, *Singular Integral Equations*, transl. by J. R. Radok (S-H Service Agency, New York, 1961), 2nd ed.

[18] In this paper the notation $f(x) \to g(x)$ is used to mean that $f(x)/g(x) \to 1$, in contrast to $f(x) \sim g(x)$, which merely means that $f(x)/g(x) \to \text{const} \neq 0$.

[19] Ref. 17, p. 36.

[20] L. Castillejo, D. Dalitz, and F. Dyson, Phys. Rev. **101**, 453 (1956).

[21] N. Dunford and J. T. Schwarz, *Linear Operators* (Interscience, New York, 1958, 1963), Parts I and II.

[22] F. Riesz and B. Sz.-Nagy, *Functional Analysis* (Ungar, New York, 1969).

[23] G. Tiktopoulos, Phys. Rev. **133**, B1231 (1964), Appendix.

[24] F. G. Tricomi, *Integral Equations* (Interscience, New York, 1957), p. 153.

[25] N. Cottingham (private communication).

[26] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford U.P., London, 1969), 2nd ed.

[27] H. Bateman, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954).

Quantum Theory of the Electromagnetic Field in a Variable-Length One-Dimensional Cavity*†

GERALD T. MOORE
Department of Physics, Brandeis University, Waltham, Massachusetts 02154

(Received 13 November 1969; Revised Manuscript Received 30 January 1970)

The quantum theory of linearly polarized light propagating in a 1-dimensional cavity bounded by moving mirrors is formulated by utilizing the symplectic structure of the space of solutions of the wave equation satisfied by the Coulomb-gauge vector potential. The theory possesses no Hamiltonian and no Schrödinger picture. Photons can be created by the exciting effect of the moving mirrors on the zero-point field energy. A calculation indicates that the number of photons created is immeasurably small for nonrelativistic mirror trajectories and continuous mirror velocities. Automorphic transformations of the wave equation are used to calculate mode functions for the cavity, and adiabatic expansions for these transformations are derived. The electromagnetic field may be coupled to matter by means of a transformation from the interaction picture to the Heisenberg picture; this transformation is generated by an interaction Hamiltonian.

I. INTRODUCTION

We describe in this paper the quantum mechanics of light propagating in a 1-dimensional cavity formed by two ideal, infinite, parallel, plane mirrors which move with arbitrary externally prescribed, timelike trajectories $x = q_1(t)$ and $x = q_2(t)$. By "ideal mirrors" we mean mirrors which are perfectly conducting and whose effects may, therefore, be described by means of appropriate boundary conditions on the electromagnetic field at the ends of the cavity.

Although the quantization procedure we use can be applied to other types of Bose fields with time-dependent boundary conditions, in particular the scalar field, only the case of the electromagnetic field has any obvious practical importance. The theory we give is, for example, relevant to the understanding of the operation of lasers with moving mirrors, a subject which has received some attention in the literature.[1-3] This paper, for the sake of simplicity, treats only the case of linearly polarized light, so that the vector nature of the field does not play a very prominent role.

Most of this paper is concerned with the free field. The term "free field," as used here, means that the only interaction of the light is with the cavity mirrors, not that the field is actually free.

As we shall show, even the free quantum field has some rather remarkable properties:

(1) There exists no Hamiltonian to describe the time evolution of the field and, consequently, there exists no Schrödinger picture.

(2) Photons can be created from the vacuum by the exciting effect which the moving mirrors have on the zero-point energy of the field. Subject to the restriction of continuous mirror velocities, the number of photons created is not divergent. It is, in fact, ordinarily very small.

(3) We also show how the usual standing-wave modes characteristic of a fixed cavity may be generalized in a natural way to a cavity with moving mirrors by means of a group of automorphic transformations of the wave equation.

II. THE BOUNDARY CONDITION

We begin by formulating the theory classically. Assuming that the electric field $\mathbf{E}(x, t)$ is polarized in the z direction, we may write

$$\mathbf{A} = A\hat{\mathbf{k}},$$
$$\mathbf{E} = E\hat{\mathbf{k}} = -\frac{\partial A}{\partial t}\hat{\mathbf{k}}, \tag{1}$$
$$\mathbf{B} = \operatorname{curl}\mathbf{A} = -\frac{\partial A}{\partial x}\hat{\mathbf{j}} = B\hat{\mathbf{j}},$$

where $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ are the usual triad of unit vectors, $\mathbf{B}$ is the magnetic field, and $\mathbf{A}$ is the vector potential in the Coulomb gauge. According to Maxwell's equations, A, E, and B all satisfy the simple wave equation

$$\frac{\partial^2 A}{\partial t^2} = \frac{\partial^2 A}{\partial x^2}, \tag{2}$$

where we have chosen units in which the velocity of light (c) equals one. For fixed mirrors, the usual continuity requirement on $\mathbf{E}$ is that $\mathbf{E} = 0$ at the mirrors. For moving mirrors, the natural generalization is that the electric field vanishes at the mirrors in the Lorentz frames in which the mirrors are instantaneously at rest. We now show that this condition is satisfied if the condition that A vanishes on the boundaries is satisfied. We require, therefore, that

$$A(q_1(t), t) = A(q_2(t), t) = 0. \tag{3}$$

Consider now a Lorentz transformation parametrized by $\mathbf{v} = v\hat{\mathbf{i}}$. The transformed electric field is

$$\mathbf{E}' = \gamma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = -\gamma\left(\frac{\partial A}{\partial t} + v\frac{\partial A}{\partial x}\right)\hat{\mathbf{k}}. \quad (4)$$

When v is put equal to $\dot{q}_i(t)$, $i = 1, 2$, and A is evaluated for $x = q_i(t)$, then $\partial A/\partial t + v\partial A/\partial x$ is just the directional derivative of A along the mirror trajectory and, hence, is identically zero.

Since the vector nature of $\mathbf{A}$ plays no role in the subsequent formulation of our field theory, we generally use the symbol ϕ instead of A in discussing the abstract properties of the theory, and we reserve the symbol A for particular applications to electromagnetism. Also, we generally consider ϕ to be dimensionless.

After we perform the quantization of the field, $\phi(x, t)$ will become the field operator of the theory (really an operator-valued distribution). Because of the boundary conditions (3), all quantum fluctuations of the field are assumed to be suppressed at the mirrors. This, of course, cannot be expected to occur with ordinary real mirrors, which do not behave at all ideally at very high frequencies (for instance, in the X-ray range). Nevertheless, it is reasonable to apply the results of our quantum theory to cavities which are approximately ideal at optical frequencies and below, provided that the physical observables one calculates do not depend appreciably (particularly in a divergent way) on contributions from the very high frequency modes. It is fortunate and not at all *a priori* obvious that it is possible to formulate a well-defined quantum theory which satisfies these conditions within the boundary-condition idealization and the additional idealizations that the mirrors have precisely defined trajectories and that there be no radiation reaction on the mirrors.

It is an immediate consequence of the vanishing boundary conditions on $\phi(x, t)$ that, except for the case of fixed mirrors, the quantum theory we are constructing does not possess a Hamiltonian. The proof does not depend on whether $\phi(x, t)$ is a free or an interacting field. Suppose, by way of contradiction, that there were a Hamiltonian. This would imply the existence of a unitary time-translation operator $U(t, t_0)$ with the property

$$\phi(x, t) = U^{+}(t, t_0)\phi(x, t_0)U(t, t_0).$$

However, if we require the point (x, t_0) to lie on a mirror trajectory, then $\phi(x, t_0) = 0$ by the boundary conditions, and it is impossible to get a nonzero value of $\phi(x, t)$ by a unitary transformation.

It follows as a corollary that no Schrödinger picture exists for the moving mirror system.

The same sort of argument can be used to show that no momentum operator exists. This is true even for the case of fixed mirrors.

III. SYMPLECTIC STRUCTURE OF THE SPACE OF CLASSICAL SOLUTIONS

Since the wave equation (2) is linear, it will be satisfied by expectation values of the field operator in any given state. Therefore, it is to be expected that the structure of the space of classical solutions of Eq. (2) will be important in constructing the quantum theory. These solutions are known to have the general form

$$g(x, t) = g_1(t - x) + g_2(t + x). \quad (5)$$

The wave equation is second order in time and, viewed as an initial-value problem, its solutions are specified by giving values for both the field and its first time derivative along any one spacelike curve in the (x, t) space. For simplicity, we generally specify initial conditions along a constant-time line.

We now define the vector space S to consist of all real solutions $f(x, t)$ of Eqs. (2) and (3) for which f and $\partial f/\partial t$ are square-integrable over x for arbitrary fixed time. If $f_1(x, t)$ and $f_2(x, t)$ are any two elements of S, we define their bracket product to be

$$\{f_1 \mid f_2\} = \int_{q_1(t)}^{q_2(t)} f_2(x, t) \overleftrightarrow{\frac{\partial}{\partial t}} f_1(x, t)\, dx,$$

$$f_2 \overleftrightarrow{\frac{\partial}{\partial t}} f_1 \equiv f_2 \frac{\partial}{\partial t} f_1 - f_1 \frac{\partial}{\partial t} f_2. \quad (6)$$

The value of the bracket product is independent of time; its time derivative may easily be shown to vanish. Because of the existence of the antisymmetric form $\{f_1 \mid f_2\} = -\{f_2 \mid f_1\}$, S forms a symplectic space in a natural way. The form $\{f_1 \mid f_2\}$ is nondegenerate; that is, for every $f_1(x, t) \not\equiv 0$, there exists an $f_2(x, t)$ such that $\{f_1 \mid f_2\} \neq 0$. To prove this, define f_2 by its initial conditions at time t_0 in such a way that

$$f_2(x, t_0) = \frac{\partial}{\partial t} f_1(x, t_0),$$

except possibly very close to the mirrors, and

$$\frac{\partial}{\partial t} f_2(x, t_0) = -f_1(x, t_0).$$

Then $\{f_1 \mid f_2\}$ is positive definite.

The expression

$$\int_{q_1}^{q_2} \left[\left(\frac{\partial f}{\partial t}\right)^2 + \left(\frac{\partial f}{\partial x}\right)^2\right] dx,$$

corresponding to the classical field energy, is not conserved when one has moving mirrors. Physically, this occurs because of the Doppler shift undergone by light while being reflected.

The elements of S obey the equation

$$f(x, t) = \int_{q_1(t')}^{q_2(t')} D(x, t; x', t') \overleftrightarrow{\frac{\partial}{\partial t'}} f(x', t')\, dx', \quad (7)$$

where $D(x, t; x', t')$ is called the propagator or the commutator function and possesses the following properties:

(a) $D(x, t; x', t')$ satisfies the wave equation and the boundary conditions in both the unprimed and primed variables.

(b) $D(x, t; x', t') = -D(x', t'; x, t)$. (8)

(c) $D(x, t; x', t) = 0$. (9)

(d) $$\frac{\partial^2}{\partial t^2} D(x, t; x', t')\Big|_{t=t'} = \frac{\partial^2}{\partial t'^2} D(x, t; x', t')\Big|_{t=t'} = \frac{\partial^2}{\partial t \partial t'} D(x, t; x', t')\Big|_{t=t'} = 0. \quad (10)$$

(e) $$\frac{\partial}{\partial t} D(x, t; x', t')\Big|_{t=t'} = -\frac{\partial}{\partial t'} D(x, t; x', t')\Big|_{t=t'} = \delta(x - x'). \quad (11)$$

If we choose to specify f by giving initial conditions at a particular value of t', we see that Eq. (7) provides the general solution of the initial-value problem.

The construction of the commutator function proceeds by the method of images as follows. For 1-dimensional light propagation with no mirrors present, the commutator function is easily shown to be

$$D_F(x, t; x', t') = \tfrac{1}{4}[\epsilon(t - t' - x + x') + \epsilon(t - t' + x - x')], \quad (12)$$

where ϵ is the usual sgn function

$$\epsilon(u) = 1, \quad u > 0,$$
$$= -1, \quad u < 0. \quad (13)$$

D_F is equal to $+\frac{1}{2}$ or $-\frac{1}{2}$, depending on whether (x, t) is in the forward or backward light cone of (x', t'), and is zero for spacelike separations. The construction of the propagator for arbitrarily moving

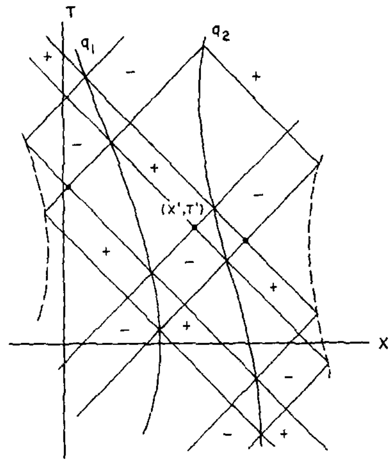


FIG. 1. The propagator $D(x, t; x', t')$ for light contained in a cavity bounded by moving mirrors. The cavity is located between the mirror trajectories $q_1(t)$ and $q_2(t)$. The point (x', t') and its image points are indicated by heavy dots. The propagator is equal to $+\frac{1}{2}$ or $-\frac{1}{2}$ in the rectangles marked (+) or (−) and is zero elsewhere. The image mirrors are indicated by dashed lines.

mirrors is of the form

$$D(x, t; x', t') = \sum_i (-1)^i D_F(x, t; x_i', t_i'), \quad (14)$$

where (x_i', t_i') includes (x', t') and also an infinite set of image points chosen so as to preserve the boundary conditions at the mirrors. A diagram showing $D(x, t; x', t')$ for a particular fixed value of (x', t') is given in Fig. 1. Both (x', t') and its image points are indicated by heavy dots. Inside the rectangles marked (+) or (−), the propagator has the value $+\frac{1}{2}$ or $-\frac{1}{2}$. Elsewhere, the propagator is equal to zero. The image mirrors, along which vanishing boundary conditions are also satisfied, are indicated with dashed lines. The construction of the D function can always be made, provided only that the mirror trajectories are everywhere timelike. The reader should try visualizing the propagator in terms of square pulses moving back and forth in the cavity and suffering reflections at the walls.

Consider now a set of vectors $|A_n\}$ and $|B_n\}$, $n = 1, 2, \cdots, \infty$, in S with the following properties:

$$\{A_m | A_n\} = 0,$$
$$\{B_m | B_n\} = 0, \quad (15)$$
$$\{A_m | B_n\} = \delta_{mn}.$$

If every member of S can be expressed as a linear combination of the $|A_n\}$ and the $|B_n\}$, then these vectors are said to form a canonical basis for S. We now show by constructing an explicit example that it is indeed possible to span S with a canonical basis. Let $f_n(x)$ be a complete set of real orthonormal functions on the interval $q_1(0)$ to $q_2(0)$ relative to the inner product

$$\int_{q_1(0)}^{q_2(0)} dx f(x) g(x)$$

and obeying suitable boundary conditions, for instance, $f_n(q_1(0)) = f_n(q_2(0)) = 0$. Now define

$$A_n(x, t) = \int_{q_1(0)}^{q_2(0)} D(x, t; x', 0) f_n(x')\, dx',$$
$$B_n(x, t) = -\int_{q_1(0)}^{q_2(0)} \frac{\partial D(x, t; x', t')}{\partial t'}\bigg|_{t'=0} f_n(x')\, dx'. \quad (16)$$

Since the functions $f_n(x)$ are complete, arbitrary initial conditions at $t = 0$ can be expanded in terms of them. Hence, by Eq. (7), the functions $A_n(x, t)$ and $B_n(x, t)$ do form a basis for S. The fact that this is a canonical basis is easily established by evaluating the bracket products at $t = 0$.

Given a canonical basis, we may now write an arbitrary vector in the space in the form

$$|f\} = \sum_{n=1}^{\infty} y_n\, |A_n\} + z_n\, |B_n\}, \quad (17)$$

where the y_n and z_n are real constants. Taking matrix elements of Eq. (17) with the various basis vectors gives

$$y_n = -\{B_n \,|\, f\},$$
$$z_n = \{A_n \,|\, f\}, \quad (18)$$

so that we can write

$$|f\} = \sum_n |B_n\}\{A_n \,|\, f\} - |A_n\}\{B_n \,|\, f\}. \quad (19)$$

We can now recognize the important operator relation

$$\sum_n |B_n\}\{A_n| - |A_n\}\{B_n| = 1, \quad (20)$$

where 1 is the unit operator in the symplectic space. By writing Eq. (19) in the function language and comparing with Eq. (7), we can see that Eq. (20) is equivalent to the relation

$$\sum_n B_n(x, t)A_n(x', t') - A_n(x, t)B_n(x', t') = -D(x, t; x', t'). \quad (21)$$

Let us now discuss more generally the subject of linear operators in S. If V is any such operator, we define the adjoint of V by the equation

$$\{f|\, V\, |g\} = -\{g|\, V^{\psi}\, |f\}, \quad (22)$$

where $|f\}$ and $|g\}$ are arbitrary elements of S. We have used the notation V^{ψ} instead of V^{+} to emphasize that this adjoint is different from the usual Hilbert-space one. For instance, VV^{ψ} is not necessarily a positive operator; in fact, $\{f|\, VV^{\psi}\, |f\} = 0$ for all $|f\}$. The identity $(VW)^{\psi} = W^{\psi}V^{\psi}$ may be readily verified by using a basis expansion of the operators V and W. An important class of operators are those obeying the relation $UU^{\psi} = U^{\psi}U = 1$. These operators are called symplectic operators and form a group called the infinite symplectic group. Symplectic operators are used for transforming from one canonical basis to another. It is clear that if the set $|A_n\}$, $|B_n\}$, $n = 1, \cdots, \infty$, forms a canonical basis, so does the set $U\,|A_n\}$, $U\,|B_n\}$.

IV. QUANTIZATION OF THE FIELD

We now perform the algebraic part of the program of quantizing the field, using a method developed by Segal.[4] We define a mapping $\mathfrak{R}$ from vectors in S onto Hermitian operators in the Hilbert space K of physical states. The space K, of course, is yet to be defined. The mapping is defined to have the property

$$[\mathfrak{R}(f_1), \mathfrak{R}(f_2)] = -i\{f_1 \,|\, f_2\}. \quad (23)$$

Suppose we pick a canonical basis $A_n(x, t)$, $B_n(x, t)$ in S. Then we define

$$\mathfrak{R}(A_n) = p_n,$$
$$\mathfrak{R}(B_n) = q_n, \quad (24)$$

from which follow the canonical commutation relations

$$[p_n, p_m] = 0,$$
$$[q_n, q_m] = 0, \quad (25)$$
$$[p_n, q_m] = -i\delta_{mn}.$$

For arbitrary f in S, we now have

$$\mathfrak{R}(f) = \mathfrak{R}\Big(\sum_n |B_n\}\{A_n \,|\, f\} - |A_n\}\{B_n \,|\, f\}\Big)$$
$$= \sum_n [-\{f \,|\, A_n\}q_n + \{f \,|\, B_n\}p_n]$$
$$= \Big\{f \,\Big|\, \sum_n B_n p_n - A_n q_n\Big\}$$
$$= \{f \,|\, \phi\}, \quad (26)$$

where we have introduced the field operator $\phi(x, t)$, defined as

$$\phi(x, t) = \sum_n B_n(x, t)p_n - A_n(x, t)q_n. \quad (27)$$

Note that, with this definition, by the use of Eq. (21),

$$[\phi(x, t), \phi(x', t')] = -iD(x, t; x', t'). \quad (28)$$

In particular, the commutator is zero for spacelike separations, a fact which is an important requirement for relativistic causality. Moreover, because of Eq. (11), ϕ and $\pi = \partial\phi/\partial t$ satisfy a δ-function commutation relation at equal times.

The inverse to the mapping $\mathfrak{R}$ may be expressed by the formula

$$f(x, t) = -i[\mathfrak{R}(f), \phi(x, t)]. \quad (29)$$

The field operator ϕ is required to be independent of the particular canonical basis in which one chooses to expand it. This means that the basis vectors A_n, B_n and the operators p_n, q_n must transform contragrediently.

V. THE NUMBER OPERATOR

To complete the quantization program, we want to define the Hilbert space K and to give the theory a particle interpretation. We can do this by defining a number operator for the theory. Once the number operator N is defined, K can be taken to be the Fock space engendered by N.

Proceeding first in a general way, we pick an arbitrary canonical basis $|A_n\}$, $|B_n\}$ in S with associated canonical operators p_n, q_n. One can pass over to creation and annihilation operators a_n^+ and a_n by the relations

$$q_n = (a_n + a_n^+)/\sqrt{2},$$
$$p_n = -i(a_n - a_n^+)/\sqrt{2}, \qquad (30)$$

with the inverse relation

$$a_n = (q_n + ip_n)/\sqrt{2}. \qquad (31)$$

More complicated relations than (30) and (31) are possible, but these are sufficiently general, since the more complicated relations can always be regarded as compositions of (30) and (31) with appropriate symplectic transformations. The number operator associated with the $|A_n\}$, $|B_n\}$ basis is then

$$N = \sum_n a_n^+ a_n. \qquad (32)$$

The way in which N depends on the choice of basis may be expressed explicitly by means of the following relation:

$$\begin{aligned} N &= \sum_n \tfrac{1}{2}(q_n - ip_n)(q_n + ip_n) \\ &= \sum_n \tfrac{1}{2}(\{B_n \mid \phi\} - i\{A_n \mid \phi\})(\{B_n \mid \phi\} + i\{A_n \mid \phi\}) \\ &= -\tfrac{1}{2}\sum_n \{\phi| [|B_n\}\{B_n| + |A_n\}\{A_n| \\ &\quad + i\,|B_n\}\{A_n| - i\,|A_n\}\{B_n|]\,|\phi\} \\ &= -\tfrac{1}{2}\{\phi|\,\mathfrak{J} + i\mathsf{I}\,|\phi\}. \end{aligned} \qquad (33)$$

The operator

$$\mathfrak{J} \equiv \sum_n |B_n\}\{B_n| + |A_n\}\{A_n| \qquad (34)$$

determines the basis dependence of N. The symplectic operators U which commute with $\mathfrak{J}$ determine the changes of basis which leave N invariant. They form a subgroup $G(\mathfrak{J})$ of the symplectic group. One easily verifies that $\mathfrak{J}^\psi = -\mathfrak{J}$ and $\mathfrak{J}^2 = -\mathsf{I}$, so that $\mathfrak{J}$ is itself a symplectic operator. It turns out that not only is $\mathfrak{J}$ basis dependent, but so is the group $G(\mathfrak{J})$. This amounts to the statement that $G(\mathfrak{J})$ is not a normal subgroup. Note in Eq. (33) that, because ϕ is an operator, $\{\phi \mid \phi\}$ is equal to infinity, not zero.

In function language, $\mathfrak{J}$ is written as

$$\begin{aligned} \mathfrak{J} \to\ & D^{(i)}(x, t; x', t') \\ &= \sum_n B_n(x, t)B_n(x', t') + A_n(x, t)A_n(x', t') \end{aligned} \qquad (35)$$

and is called the anticommutator function because it is the vacuum expectation value of the field operator anticommutator:

$$\begin{aligned} & D^{(i)}(x, t; x', t') \\ &\quad = \langle 0|\, \phi(x, t)\phi(x', t') + \phi(x', t')\phi(x, t)\, |0\rangle. \end{aligned} \qquad (36)$$

Equation (36) may be verified by substituting Eqs. (30) into Eq. (27) and normal-ordering the field-operator anticommutator.

We have established, then, that there is a 1-to-1 correspondence between the choice of a number operator and the choice of an anticommutator function. What this choice should be, however, cannot be determined from the formalism of the theory, but must be made on physical grounds. Also, the choice involves more than just a matter of convenience because, since we are dealing with a dynamical system with an infinite number of degrees of freedom, the Fock spaces K engendered by different number operators need not necessarily be identical. We need, therefore, to define the number operator in order to be able to determine what are the allowed physical states of the field.

VI. THE FIXED-MIRROR PROBLEM

As a means toward this end, consider the case of a cavity with fixed mirrors located at $x = 0$ and $x = L$. This problem is amenable to traditional field-quantization techniques, and, when treated by these means, the choice of an anticommutator functionemerges in a very natural and satisfying way. The wave equation can be solved by means of Fourier transformation, and the general form of the classical solution is

$$\begin{aligned} f(x, t) &= \sum_{n=1}^{\infty}(n\pi)^{-\frac{1}{2}} c_n e^{-ik_n t}\sin(k_n x) + \text{c.c.}, \\ k_n &= n\pi/L, \end{aligned} \qquad (37)$$

which separates into a sum of positive- and negative-frequency pieces. The set of sequences $\{c_n\}$ form a complex Hilbert space J in a natural way. The passage to quantum theory is made by second-quantizing this Hilbert space, introducing creation and annihilation operators a_n^+ and a_n which create and destroy photons in the mode n. The states of the field upon which a_n and a_n^+ act lie in a Fock space K. The number of photons in the mode n is measured by the operator $a_n^+ a_n$. The total number of photons $N = \sum_n a_n^+ a_n$ is invariant under changes of basis in J. The field operator is written

$$\phi(x, t) = \sum_n (n\pi)^{-\frac{1}{2}} a_n e^{-ik_n t} \sin(k_n x) + \text{H.c.} \quad (38)$$

The presence of the factor $(n\pi)^{-\frac{1}{2}}$ causes ϕ and π to obey canonical equal-time commutation relations and, more generally, causes Eq. (28) to be satisfied. The time development given in Eq. (38) can be regarded as being generated by the Hamiltonian

$$H = \sum_n k_n a_n^+ a_n ,$$

using the Heisenberg equation of motion. Then a_n is interpreted as an operator in the Schrödinger picture and $a_n(t) = a_n e^{-ik_n t}$ as an operator in the Heisenberg picture. This Hamiltonian also leaves N independent of time.

A canonical basis in the space S, which will give the same N as the traditional theory, is

$$A_n = (\tfrac{1}{2} n\pi)^{-\frac{1}{2}} \sin(k_n t) \sin(k_n x),$$
$$B_n = (\tfrac{1}{2} n\pi)^{-\frac{1}{2}} \cos(k_n t) \sin(k_n x). \quad (39)$$

Then the function

$$D^{(i)}(x, t; x', t') = \sum_n (\tfrac{1}{2} n\pi)^{-1} \sin(k_n x) \sin(k_n x') \cos[k_n(t - t')] \quad (40)$$

is time-translationally invariant, corresponding to the time-translational invariance of N and, in particular, the vacuum of N. This invariance is the real physical justification for the traditional definition of N. The function $D^{(i)}$ is the imaginary part of the positive- and negative-frequency propagators

$$D^{(\pm)} = \tfrac{1}{2}(D \pm iD^{(i)}),$$

which explains the rationale of the use of the superscript $^{(i)}$.

Equation (40) can be summed to give $D^{(i)}$ in closed form. For instance, its equal-time value can be calculated to be

$$D^{(i)}(x, t; x', t) = -\frac{1}{\pi} \log \left| \frac{\sin[\frac{1}{2}\pi(x - x')/L]}{\sin[\frac{1}{2}\pi(x + x')/L]} \right|. \quad (41)$$

This exhibits a logarithmic divergence when $x = x'$. It is interesting that, in one dimension, the anticommutator function in the absence of mirrors $D_F^{(i)}$ exhibits an infrared divergence and is infinite everywhere. If, however, it is renormalized by subtracting an infinite constant, its equal-time behavior for x near x' is similar to that of Eq. (41).

VII. SCATTERING

For a system with moving mirrors, one cannot find a $D^{(i)}$ function which is time-translationally invariant for all times. However, let us now pose a sort of "scattering" problem. We consider a system in which the mirrors are fixed with separation L_1 for $-\infty < t < t_1$ (region I), after which they move arbitrarily (region II) until time t_2 and then again become fixed (region III), this time with separation L_2. Let $C_{\rm I}$ denote the forward light cone passing through the points $(q_1(t_1), t_1)$ and $(q_2(t_1), t_1)$ and $C_{\rm III}$ denote the backward light cone passing through the points $(q_1(t_2), t_2)$ and $(q_2(t_2), t_2)$. Since the mirror motion is determined in a manner causally independent of the field dynamics, the description of the system for $t < t_1$ must be independent of whether the mirrors move later on. This means that the $D^{(i)}$ function (40) appropriate for a fixed cavity of length L_1 should be used to define N in this region. Similarly, in region III the $D^{(i)}$ function for a fixed cavity of length L_2 should be used. But notice now that the function $D_{\rm I}^{(i)}$ propagates into region III by means of the wave equation. Also, $D_{\rm III}^{(i)}$ propagates backward in time into region I. The point is that these two functions do not necessarily have to be identical. This can be seen by considering an example in which region II is so small that the light cones $C_{\rm I}$ and $C_{\rm III}$ intersect. Inside the rectangle formed by their intersection both $D_{\rm I}^{(i)}$ and $D_{\rm III}^{(i)}$ take on their fixed-mirror values. However, from the form of Eq. (41) it is clear that $D_{\rm I}^{(i)}$ and $D_{\rm III}^{(i)}$ are not generally going to agree in the overlap region.

Thus, the N's used for regions I and III are, in general, different. The state $|0_{\rm I}\rangle$ which has the properties of a vacuum in region I will not be vacuum-like in region III, but will have some probability of containing real energy-carrying photons. We may think of this effect as being a physical manifestation of the zero-point energy of the field (for the fixed-mirror problem an infinite, but L-dependent constant) being partially converted into a particle form by the exciting effect of the moving mirrors. In the moving-mirror region II, the very concept of photons becomes muddy, just because the absence of photons (namely, a time-translationally invariant vacuum state) cannot be defined. There are many ways of defining N in region II as a function of time so as to connect continuously to $N_{\rm I}$ and $N_{\rm III}$, but all of them are rather artificial. Another possible physical interpretation of the creation of photons is that they are radiated by currents necessarily associated with the moving mirrors. We have here a suggestive link between the concept of radiation and that of zero-point field energy.

It is interesting to compare our results for the cavity with moving mirrors to the situation that occurs in a simple one-mode problem, namely, the harmonic oscillator with a time-dependent coupling constant:

$$\ddot{x} + \omega^2(t)x = 0. \tag{42}$$

All of the quantization machinery we have been using can be applied here in the same way; in this case, the bracket product is just the Wronskian

$$\{x_1 \mid x_2\} = \dot{x}_1 x_2 - x_1 \dot{x}_2 . \tag{43}$$

The analogous "scattering" problem here is the creation, because of the changing potential, of excitations in a state which was originally the ground state of the oscillator. Again, as in the moving-mirror problem, this is a purely quantum phenomenon and does not occur in the classical theory.

Unlike the oscillator problem, for the moving-mirror problem, we have to show that the changes in the number operator caused by the motion of the mirrors leave the number operator defined in the same Hilbert space throughout. This is not easily proven, but later we give an argument, though not a proof, that it is true, provided that the mirror velocities q_1 and q_2 are continuous.

VIII. MODE FUNCTIONS

First, however, we need to discuss how one obtains analytic expressions for the functions in S. Consider the transformation of variables

$$t - x = f(u - s),$$
$$t + x = g(u + s). \tag{44}$$

For arbitrary functions f and g, this transformation has the property of automorphically mapping the wave equation into itself:

$$\frac{\partial^2\phi}{\partial t^2} = \frac{\partial^2\phi}{\partial x^2} \rightleftarrows \frac{\partial^2\phi}{\partial u^2} = \frac{\partial^2\phi}{\partial s^2} . \tag{45}$$

The very extensive class of moving-mirror problems for which the boundary conditions (3) are separable by means of such transformations, and which are thereby rendered exactly soluble, we call type-A problems. As a simple example, let

$$t - x = e^{u-s},$$
$$t + x = e^{u+s} \tag{46}$$

within the forward light cone of $(x, t) = (0, 0)$. The ratio x/t depends only on s and not on u, so that (46) is appropriate for solving the problem of two uniformly moving mirrors which intersect at $(x, t) = (0, 0)$. For instance, the canonical mode functions

$$A_n(s, u) = (\tfrac{1}{2}n\pi)^{-\frac{1}{2}} \sin(2\delta n\pi u) \sin(2\delta n\pi s),$$
$$B_n(s, u) = (\tfrac{1}{2}n\pi)^{-\frac{1}{2}} \cos(2\delta n\pi u) \sin(2\delta n\pi s) \tag{47}$$

describe a cavity with one fixed mirror and one mirror with velocity $v = \tanh[1/(2\delta)]$. Note that these form a particularly natural basis for the definition of the number operator in the quantum theory and that the vacuum so defined will possess translational invariance in u. Physically, this is a scaling invariance, the wave equation being unchanged by $x \to bx$, $t \to bt$ for any constant b. Clearly, any type-A problem is going to exhibit some such translational invariance and is going to possess a natural choice of a number operator, the vacuum state of which will share this invariance. Furthermore, there will exist a quasi-Hamiltonian which generates these translations. In the case described by Eqs. (47), this is of the form

$$H = \sum_n 2\delta n\pi a_n^+ a_n . \tag{48}$$

The mode functions (47) also may be separated into positive- and negative-frequency parts, "frequency" here referring to Fourier transformation with respect to u. A typical negative-frequency solution, expressed in terms of x and t, is

$$f_n(x, t) = (t - x)^{2\delta n\pi i} - (t + x)^{2\delta n\pi i}. \tag{49}$$

These mode functions were obtained by a different means by Solimene[5] and, more recently, by Baranov and Shirokov.[3] In the adiabatic limit $v \ll 1$, one obtains

$$f_n \simeq \frac{2}{i} \exp\left(\frac{in\pi}{v}\log t\right) \sin\left(\frac{n\pi x}{vt}\right). \tag{50}$$

These resemble the mode functions for a fixed cavity, except that the cavity length $L = vt$ is now a slowly varying function of time. The derivative of the phase angle defines the frequency and is seen to be $n\pi/vt = n\pi/L$, which is the limit one expects. Notice that, if we interpret f_n as the vector potential A_n for the electromagnetic field, then A_n consists of oscillations at constant amplitude, but that the electric field $E_n = -\partial A_n/\partial t$ and the magnetic field $B_n = -\partial A_n/\partial x$ have amplitudes which are roughly proportional to $1/L$. These decreasing amplitudes result from the spreading of the field energy throughout the expanding cavity and the loss of energy at the moving mirror because of the Doppler effect.

We now want to extend our discussion of the uniformly moving mirror to more general situations. For simplicity, we restrict ourselves henceforth to type-*A* problems where only one mirror [with trajectory $q(t)$] is allowed to move. The other mirror is required to lie at $x = 0$ and also at $s = 0$. We see from Eqs. (44) that now only one transformation function $f = g$ is needed.

We introduce a real function $R(t)$, which we require to satisfy the equation

$$R(t - q(t)) = R(t + q(t)) - 2. \qquad (51)$$

Mode functions for the moving-mirror problem may then be written down in the form

$$f_n(x, t) = e^{-in\pi R(t+x)} - e^{-in\pi R(t-x)}, \quad n = 1, 2, \cdots, \infty. \qquad (52)$$

The complex conjugate functions $f_n^*(x, t)$ are also solutions. The mode functions vanish at the moving mirror because of Eq. (51), In fact, one may identify R with the inverse of the function f of the transformation Eqs. (44).

The construction of mode functions amounts then to solving Eq. (51). If $R(t)$ is a given function, then finding the corresponding trajectory $q(t)$ is a purely algebraic problem. We list a few examples below:

(a) $R(t) = 2\delta \log t;$

$$q = \{\tanh [1/(2\delta)]\}t = vt. \qquad (53)$$

(b) $R(t) = t^2/d^2;$

$$q = d^2/t. \qquad (54)$$

(c) $R(t) = \alpha_1(t - \tau)/\tau, \quad t < \tau,$

$$\begin{aligned} &= \alpha_2(t-\tau)/\tau, \quad t > \tau; \\ q(t) &= \tau/\alpha_1, \quad t < \tau(1 - 1/\alpha_1), \\ &= [(\alpha_1 - \alpha_2)(t - \tau) + 2\tau]/(\alpha_1 + \alpha_2), \\ &\qquad \tau(1 - 1/\alpha_1) < t < \tau(1 + 1/\alpha_2), \\ &= \tau/\alpha_2, \quad t > \tau(1 + 1/\alpha_2). \end{aligned} \qquad (55)$$

Example (b) shows that it is possible to obtain, in certain cases, mirror trajectories which have the unphysical property of being in part spacelike. Example (c) yields a trajectory which, at first sight, looks like an example of a "scattering" problem such as we were considering in connection with the number operator. However, since the mode functions go through the scattering region without becoming intermixed or changed in amplitude, there is, in fact, no scattering. This is true of any type-A problem which asymptotically approaches a fixed-mirror problem as $t \to \pm\infty$.

The construction of the R function when $q(t)$ is a given function can be made by means of a perturbation technique in the adiabatic limit $\dot{q} \ll 1$. We rewrite Eq. (51) as a more conventional boundary-value problem by putting

$$\begin{aligned} R(t \mp x) &\equiv R(x, t), \\ \frac{\partial R}{\partial t} &= \mp \frac{\partial R}{\partial x}, \\ R(\pm q(t), t) &= R(\mp q(t), t) - 2. \end{aligned} \qquad (56)$$

Now we introduce a scaled coordinate $\xi = x/q(t)$ and write the differential Eq. (56) in terms of ξ and t. The result is

$$\begin{aligned} (1 \mp \xi\dot{q}) \frac{\partial R}{\partial \xi} \pm q \frac{\partial R}{\partial t} &= 0, \\ R(\xi = \pm 1, t) = R(\xi &= \mp 1, t) - 2. \end{aligned} \qquad (57)$$

Let us now look for solutions which in the limit $\dot{q} \ll 1$ reduce to the positive-frequency solutions for the fixed-mirror problem. Generalizing from our adiabatic result [Eq. (50)] for the uniformly-moving-mirror problem, we guess that the major portion of the time dependence of the R function is $\int^t q^{-1}\, dt$. Putting

$$R = g + \int^t (1/q)\, dt, \qquad (58)$$

we get

$$(1 \mp \xi\dot{q}) \frac{\partial g}{\partial \xi} \pm q \frac{\partial g}{\partial t} \pm 1 = 0. \qquad (59)$$

We are assuming that g is a slowly varying function of time, and, to indicate this, we set $s = \epsilon t$, where ϵ is a small parameter. Then

$$\begin{aligned} (1 \mp \xi\epsilon q') \frac{\partial g}{\partial \xi} \pm \epsilon q \frac{\partial g}{\partial s} \pm 1 &= 0, \\ q' &\equiv \frac{dq}{ds}. \end{aligned} \qquad (60)$$

We next assume that $g(\xi, s)$ can be expanded in a Taylor's series in ϵ:

$$g(\xi, s) = \sum_{n=0}^{\infty} g^{(n)}(\xi, s)\epsilon^n. \qquad (61)$$

The function $g^{(0)}$ is easily calculated up to an additive function of s:

$$g^{(0)} = \mp\xi + \alpha_{10}(s). \qquad (62)$$

This satisfies the boundary condition given in Eqs. (57), so that the higher-order functions must obey

$$g^{(n)}(-1, s) = g^{(n)}(1, s). \qquad (63)$$

Substituting (61) into (60) and equating powers of ϵ, we get

$$\frac{\partial g^{(k)}}{\partial \xi} \mp \xi q' \frac{\partial g^{(k-1)}}{\partial \xi} \pm q \frac{\partial g^{(k-1)}}{\partial s} = 0, \quad k = 1, 2, 3, \cdots. \qquad (64)$$

We now expand the functions $g^{(k)}$ in a power series in ξ (only a polynomial is needed):

$$g^{(k-1)}(\xi, s) = \sum_{j=0}^{k} \alpha_{kj}(s)(\pm\xi)^j. \tag{65}$$

Substituting into (64) gives

$$(j+1)\alpha_{k,j+1} - q'j\alpha_{k-1,j} + q\alpha'_{k-1,j} = 0. \tag{66}$$

This recursion relation may be written in a form which allows an iterative solution for all the α_{kj} in terms of $\alpha_{k-j,0}$:

$$\begin{aligned}\alpha_{kj} &= -\frac{1}{j}\, q^j[q^{-(j-1)}\alpha_{k-1,j-1}]' \\ &= \frac{1}{j!}(-q)^j \frac{d^j}{ds^j}[\alpha_{k-j,0}].\end{aligned} \tag{67}$$

To make α_{kk} well-defined, we also need the definition of a_{00}, which is chosen so as to obtain the right result for $g^{(0)}$:

$$\alpha_{00} = \int^s q^{-1}\, ds. \tag{68}$$

Equations (64) may be used to solve for the $g^{(k)}$ functions one at a time. Solving the kth equation determines $g^{(k)}$ up to an additive function $\alpha_{k+1,0}(s)$ which is found by applying the boundary condition (63) to the solution for $g^{(k+1)}$. In particular, one finds that $\alpha'_{10}(s) = 0$, so that we can take $\alpha_{10} = 0$. The resulting solution for R, when substituted back into Eq. (52), yields the mode function

$$\begin{aligned} f_n(x,t) \propto \exp\Bigg\{ in\pi\Bigg[-\int^t \frac{1}{q}\,dt + \frac{\dot{q}x^2}{2q^2} \\ + \frac{1}{6}\left(2\int^t \frac{\dot{q}^2}{q}\,dt - \dot{q}\right) + O(\dot{q}^3)\Bigg]\Bigg\} \\ \times \sin\left\{ n\pi\left[\frac{x}{q} + \tfrac{1}{6}(2\dot{q}^2 - \ddot{q})\left(\frac{x^3}{q^3} - \frac{x}{q}\right) + O(\dot{q}^4)\right]\right\}. \end{aligned} \tag{69}$$

To any order, this mode function by construction satisfies the boundary conditions at the mirrors exactly, but it is only approximately a solution of the wave equation. We next derive a perturbation expansion which to any order is an exact solution of the wave equation, but to finite order only approximates the boundary condition on the moving mirror.

If we apply the boundary condition (63) to Eq. (67), we discover

$$\sum_{j=1,3,\cdots}^{k} \frac{1}{j!} q^j \frac{d^j}{ds^j}[\alpha_{k-j,0}] = 0, \quad k = 2, 3, \cdots. \tag{70}$$

We notice that each of these equations involves either odd or even subscripts for the α_{l0}, but never both. Since $\alpha_{10} = 0$, it is consistent to require $\alpha_{l0} = 0$ for l odd. Equation (70) is then satisfied for all even values of k. The remaining equations may be written in the form

$$\sum_{i=0}^{l} \frac{1}{(2i+1)!} q^{2i} \frac{d^{2i}\gamma_{l-i}}{ds^{2i}} = 0, \quad l = 1, 2, \cdots, \tag{71}$$

where we have put

$$\gamma_l \equiv \frac{d\alpha_{2l,0}}{ds}. \tag{72}$$

These equations can be solved one at a time in an unambiguous manner, starting from $\gamma_0 = 1/q$. The first few are listed below:

$$\begin{aligned} \gamma_0 &= \frac{1}{q}, \\ \gamma_1 &= -\frac{1}{3!}\left(-q'' + \frac{2q'^2}{q}\right), \\ \gamma_2 &= \frac{1}{(3!)^2} q^2 \frac{d^2}{ds^2}\left[q^2 \frac{d^2}{ds^2}\left(\frac{1}{q}\right)\right] - \frac{1}{5!} q^4 \frac{d^4}{ds^4}\left(\frac{1}{q}\right), \\ &\;\;\vdots \end{aligned} \tag{73}$$

The expressions become more and more cumbersome and involve higher and higher derivatives as one proceeds further.

Using Eq. (67) and putting now $\epsilon = 1$, we may write the R function in the form

$$R = \sum_{k=0}^{\infty} \sum_{j=0}^{k} \frac{1}{j!} (\mp q)^j \frac{d^j}{dt^j}[\alpha_{k-j,0}]\xi^j. \tag{74}$$

After interchanging the order of summation and using Taylor's theorem, this becomes simply

$$R = \sum_{k=0}^{\infty} \alpha_{k,0}(t - x). \tag{75}$$

Substituting into Eq. (52) gives the set of mode functions

$$\begin{aligned} f_n(x,t) = \exp\left[-in\pi\int_\tau^{t+x}\left(\sum_{l=0}^{\infty}\gamma_l(t')\right)dt'\right] \\ - \exp\left[-in\pi\int_\tau^{t-x}\left(\sum_{l=0}^{\infty}\gamma_l(t')\right)dt'\right], \end{aligned} \tag{76}$$

which is the desired expansion.

The convergence properties of the expansions (69) and (76) are not known, though it would seem a minimal assumption that $q(t)$ be analytic for real t, since the expansions require the use of arbitrarily high derivatives. In practical cases, one is usually only interested in the first few terms, and it is reasonable to suppose that the remainder which one neglects is small, whether or not all the higher derivatives exist.

IX. JOINING OF TWO TYPE-A PROBLEMS

We may write down a canonical basis which corresponds to the natural number operator of a type-A problem with one moving mirror as

$$\begin{aligned} A_n(x, t) &= (2n\pi)^{-\frac{1}{2}}\{\cos [n\pi R(t - x)] \\ &\qquad - \cos [n\pi R(t + x)]\}, \\ B_n(x, t) &= (2n\pi)^{-\frac{1}{2}}\{\sin [n\pi R(t + x)] \\ &\qquad - \sin [n\pi R(t - x)]\}. \end{aligned} \tag{77}$$

For $R(t) = t/L$ this reduces to the fixed-mirror basis of Eqs. (39). Now suppose we consider a mirror trajectory $q(t)$ which is a junction of two (or more) type-A motions:

$$\begin{aligned} q(t) &= q_1(t), \quad t < 0, \\ &= q_2(t), \quad t > 0, \end{aligned} \tag{78}$$

and

$$q_0 \equiv q(0) = q_1(0) = q_2(0).$$

We define the natural number operator for q_1 to be N_1 and that for q_2 to be N_2. We want to calculate $\langle 0_1| N_2 |0_1\rangle$. If q_1 and q_2 are asymptotically stationary as $t \to -\infty$ and $t \to +\infty$, respectively, then $\langle 0_1| N_2| 0_1\rangle$ is the average number of photons created by the motion of the mirror $q(t)$. If we find that this is finite, it will constitute evidence that the Hilbert space is preserved by the "scattering" and that our quantum theory has well-defined physical states.

The general formula for $\langle 0_1| N_2 |0_1\rangle$ for a number operator N_2 defined with respect to a basis $|A_{k2}\}$, $|B_{k2}\}$ and a vacuum state $|0_1\rangle$ defined with respect to a basis $|A_{n1}\}$, $|B_{n1}\}$ is found by expressing the creation and annihilation operators associated with the former basis as linear combinations of the creation and annihilation operators associated with the latter basis, using Eqs. (31), (26), (27), and (30), and then normal-ordering N_2. One gets

$$\begin{aligned} \langle 0_1| N_2 |0_1\rangle &= \tfrac{1}{4} \sum_{k,n} [(\{A_{k2} \mid B_{n1}\} + \{B_{k2} \mid A_{n1}\})^2 \\ &\qquad + (\{B_{k2} \mid B_{n1}\} - \{A_{k2} \mid A_{n1}\})^2] \\ &= \langle 0_2| N_1 |0_2\rangle, \end{aligned} \tag{79}$$

the last step following from the symmetry of the result with respect to interchange of the two bases. The condition that $\langle 0_1| N_2 |0_1\rangle$ be zero, namely,

$$\{A_{k2} \mid B_{n1}\} = -\{B_{k2} \mid A_{n1}\}$$

and

$$\{B_{k2} \mid B_{n1}\} = \{A_{k2} \mid A_{n1}\},$$

also implies that $N_2 = N_1$. This means that any symplectic transformation which leaves the vacuum invariant leaves the number operator invariant also.

The matrix elements needed for Eq. (79) may be calculated at $t = 0$, using Eqs. (77). The results are

$$\{A_{k2}| B_{n1}\} = (2n\pi)^{-\frac{1}{2}}(2k\pi)^{-\frac{1}{2}} \times \int_{-q_0}^{q_0} \cos [k\pi R_2(x)] \overleftrightarrow{\frac{\partial}{\partial x}} \sin [n\pi R_1(x)]\, dx, \tag{80}$$

$$\{B_{k2}| A_{n1}\} = (2n\pi)^{-\frac{1}{2}}(2k\pi)^{-\frac{1}{2}} \times \int_{-q_0}^{q_0} \sin [k\pi R_2(x)] \overleftrightarrow{\frac{\partial}{\partial x}} \cos [n\pi R_1(x)]\, dx, \tag{81}$$

$$\{B_{k2}| B_{n1}\} = (2n\pi)^{-\frac{1}{2}}(2k\pi)^{-\frac{1}{2}} \times \int_{-q_0}^{q_0} \sin [n\pi R_1(x)] \overleftrightarrow{\frac{\partial}{\partial x}} \sin [k\pi R_2(x)]\, dx, \tag{82}$$

$$\{A_{k2}| A_{n1}\} = (2n\pi)^{-\frac{1}{2}}(2k\pi)^{-\frac{1}{2}} \times \int_{-q_0}^{q_0} \cos [n\pi R_1(x)] \overleftrightarrow{\frac{\partial}{\partial x}} \cos [k\pi R_2(x)]\, dx. \tag{83}$$

Upon substituting Eqs. (80)–(83) into Eq. (79), we get

$$\langle 0_1| N_2 |0_1\rangle = \sum_{k,n} \frac{1}{16\pi^2 nk} \times \left| \int_{-q_0}^{q_0} e^{in\pi R_1(x)} i \overleftrightarrow{\frac{\partial}{\partial x}} e^{ik\pi R_2(x)}\, dx \right|^2. \tag{84}$$

Assuming that $R_1(x)$ and $R_2(x)$ are sufficiently smooth functions, we can calculate a series expansion for the integral of Eq. (84) in the adiabatic limit by doing repeated integrations by parts. Suppose, to begin with, that we assume that the series can be approximated adequately by just the first term. Then we get

$$I_{kn} \equiv \left| \int_{-q_0}^{q_0} (n\pi R_1' - k\pi R_2') e^{in\pi R_1 + ik\pi R_2}\, dx \right| \approx \left| \frac{n\pi R_1' - k\pi R_2'}{n\pi R_1' + k\pi R_2'} \Big|_{-q_0}^{q_0} \right|. \tag{85}$$

By repeatedly differentiating Eq. (5), we obtain the following equations:

$$R'(t-q) = \frac{1+\dot{q}}{1-\dot{q}} R'(t+q), \tag{86}$$

$$R''(t-q) = \frac{2\ddot{q}}{(1-\dot{q})^3} R'(t+q) + \left(\frac{1+\dot{q}}{1-\dot{q}}\right)^2 R''(t+q). \tag{87}$$

In the approximation of Eq. (85) we find that I_{kn} vanishes exactly if $\dot{q}_1(0) = \dot{q}_2(0)$. After substituting Eq. (86) into (85), pulling out a factor $[\dot{q}_2(0) - \dot{q}_1(0)]$, and otherwise neglecting $\dot{q}_1(0)$ and $\dot{q}_2(0)$ compared to one, we get

$$I_{kn} \approx \left| 4kn\pi^2(\dot{q}_2(0) - \dot{q}_1(0)) \frac{R_1'(q_0)R_2'(q_0)}{[n\pi R_1'(q_0) + k\pi R_2'(q_0)]^2} \right|. \tag{88}$$

Now, from our earlier perturbation calculation, we know that

$$R_1'(q_0) \approx 1/q_1(q_0),$$
$$R_2'(q_0) \approx 1/q_2(q_0). \tag{89}$$

However, q_0 is just the transit time for light across the cavity at $t = 0$, and in this time q_1 and q_2 change very little from their value at $t = 0$, which is q_0. Therefore, it is consistent with our approximation to put $R_1(q_0) = R_2(q_0)$ in Eq. (88). After doing this and substituting I_{kn} into Eq. (84), we obtain

$$\langle 0_1| N_2 |0_1\rangle \approx \frac{1}{\pi^2}(\dot{q}_2(0) - \dot{q}_1(0))^2 \sum_{k,n} \frac{kn}{(k+n)^4}. \tag{90}$$

The factor $[\dot{q}_2(0) - \dot{q}_1(0)]^2$ is very small, but $\sum_{k,n} kn/(k+n)^4$ diverges logarithmically. This may be seen by letting $m = k + n$ and writing

$$\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} \frac{kn}{(k+n)^4} = \sum_{m=1}^{\infty}\sum_{n=1}^{m-1}\left(\frac{n}{m^3} - \frac{n^2}{m^4}\right) = \sum_{m=1}^{\infty}\frac{1}{6}\left(\frac{1}{m} - \frac{1}{m^3}\right) = \infty. \tag{91}$$

Thus, a discontinuity in mirror velocity creates an infinite number of photons. This result is not too surprising if we recall that, in classical electromagnetism, it is accelerations which give rise to radiation. Here, we have an infinite acceleration, and it is not unreasonable that we get an infinite amount of radiation. This is, of course, a crude argument, since we are dealing with idealized mirrors and not charged particles.

For mirror motions which do not have velocity discontinuities, but for which the velocity changes at a rate fast compared to the mode frequencies of interest, we can still interpret Eq. (90) summed over n but for fixed k to be the number of photons created in the mode k by the "scattering." The effect is, of course, much too small to be measured experimentally.

We next show that for continuous mirror velocities $\langle 0_1| N_2 |0_1\rangle$ is finite. To do this, we have to integrate the integral in Eq. (84) by parts once more. We get

$$I_{kn} \approx \left|\left[\frac{1}{n\pi R_1' + k\pi R_2'}\frac{d}{dx}\left(\frac{n\pi R_1' - k\pi R_2'}{n\pi R_1' + k\pi R_2'}\right)\right]\Bigg|_{-q_0}^{q_0}\right|. \tag{92}$$

By exactly the same sort of calculation as before, but now using Eq. (87) as well as (86), we can show that

$$\langle 0_1| N_2 |0_1\rangle \approx q_0^2(\ddot{q}_2(0) - \ddot{q}_1(0))^2 \sum_{n,k}\frac{nk}{(n+k)^6}. \tag{93}$$

This time, however, the sum is finite:

$$\sum_{n,k}\frac{nk}{(n+k)^6} = \sum_{m=1}^{\infty}\frac{1}{6}\left(\frac{1}{m^3} - \frac{1}{m^5}\right) < \infty. \tag{94}$$

Equation (93) gives a result which is of order $1/c^4$, which is ordinarily very small indeed.

X. INTERACTIONS

So far, we have been concentrating exclusively on the properties of the free field. If we now allow the field to interact with another dynamical system (for instance, the active medium in a laser), we must again circumvent the difficulty that the interacting system does not possess a total Hamiltonian. On the other hand, we want to avoid having to come to grips with the complexities and divergences which usually characterize interacting field theories. We simply want to show how the standard methods of field theory can also be applied to interacting systems with moving mirrors.

As we demonstrate in this section, the unitary transformation S from interaction picture to Heisenberg picture provides a formal definition of the interacting field dynamics which does not depend on the existence of a total Hamiltonian or a free radiation Hamiltonian. We begin by deriving the properties of this unitary transformation for the case of a resonant cavity with fixed mirrors. In this exceptional case, there *is*, of course, a free radiation Hamiltonian. However, we shall discover, after finishing the derivation, how to generalize the definition of S to apply to systems with moving mirrors.

For the sake of definiteness, we work with a particular example of an interacting system. Physically, we can think of this example as describing the dipole interaction of light with a gas of two-level atoms of mass M. Let us define $\psi_a(x,t)$ and $\psi_b(x,t)$ to be two nonrelativistic boson fields describing matter contained in a fixed box B somewhere between the resonant-cavity mirrors. In the absence of interactions, $\psi_a = \psi_{a0}$ and $\psi_b = \psi_{b0}$ obey the field equations

$$i\frac{\partial \psi_{j0}}{\partial t} = \left(-\frac{1}{2M}\frac{\partial^2}{\partial x^2} + E_j\right)\psi_{j0}, \quad j = a \text{ or } b. \tag{95}$$

They also obey suitable boundary conditions (periodic boundary conditions in B will do) and have equal-time commutation relations characteristic of annihilation operators:

$$\begin{aligned} [\psi_{j0}(x,t), \psi_{k0}(x',t)] &= 0, \\ [\psi_{j0}(x,t), \psi_{k0}^{+}(x',t)] &= \delta_{jk}\delta(x-x'). \end{aligned} \tag{96}$$

In the case where the resonant-cavity mirrors are fixed, not only Eqs. (95) but also Eq. (2) for the vector potential A_0 may be derived from a Hamiltonian H_0, which is, of course, just a sum of Hamiltonians for the three fields involved. Now let us couple these three fields by adding to H_0 an interaction Hamiltonian

$$H_{\text{int}} = \gamma\int_B A(x,t)j(x,t)\,dx, \tag{97}$$

where

$$j(x,t) \equiv \psi_a^{+}(x,t)\psi_b(x,t) + \psi_b^{+}(x,t)\psi_a(x,t). \tag{98}$$

In the Schrödinger picture not only the total Hamiltonian H but also H_0 and H_{int} are independent of time. Given any operator in the Schrödinger picture, in particular the vector potential $A(x,\tau)$, the corresponding Heisenberg and interaction picture operators are then

$$\begin{aligned} A(x,t) &= e^{iH(t-\tau)}A(x,\tau)e^{-iH(t-\tau)}, \\ A_0(x,t) &= e^{iH_0(t-\tau)}A(x,\tau)e^{-iH_0(t-\tau)}. \end{aligned} \tag{99}$$

Therefore,

$$A(x,t) = S^{+}(t,\tau)A_0(x,t)S(t,\tau), \tag{100}$$

where

$$S(t,\tau) = e^{iH_0(t-\tau)}e^{-iH(t-\tau)}. \tag{101}$$

The unitary operator $S(t,\tau)$ transforms from the interaction picture to the Heisenberg picture. It obeys the equation

$$i\frac{\partial S(t,\tau)}{\partial t} = H_{\text{int}}(t)S(t,\tau), \tag{102}$$

where

$$H_{\text{int}}(t) = e^{iH_0(t-\tau)}H_{\text{int}}e^{-iH_0(t-\tau)} \tag{103}$$

is the interaction Hamiltonian in the interaction picture. In this particular case,

$$H_{\text{int}}(t) = \gamma\int_B A_0(x,t)j_0(x,t)\,dx. \tag{104}$$

We see now that Eqs. (100), (102), and (104) make sense even for a system with moving mirrors and that they provide a formal solution to the interacting problem. The field $A_0(x,t)$ is the free vector potential field operator which we have constructed in preceding sections of this paper. We now *define* $S(t,\tau)$ by Eqs. (102) and (104) as well as the initial condition $S(\tau,\tau) = 1$. Since the operator S is generated by a Hermitian Hamiltonian $H_{\text{int}}(t)$, S is formally unitary. Also, S clearly depends only on t and not on x. Hence, the equal-time commutation relations of the fields A, ψ_a, and ψ_b are the same as those of A_0, ψ_{a0}, and ψ_{b0} and are thus independent of time.

Equations (100), (102), and (104) and the three free-field equations (2) and (95) are sufficient to enable one to derive interacting field equations for A, ψ_a, and ψ_b. Inside B, these turn out to be

$$\begin{aligned} \frac{\partial^2 A}{\partial t^2} &= \frac{\partial^2 A}{\partial x^2} - \gamma j, \\ i\frac{\partial \psi_a}{\partial t} &= \left(-\frac{1}{2M}\frac{\partial^2}{\partial x^2} + E_a\right)\psi_a + \gamma A\psi_b, \\ i\frac{\partial \psi_b}{\partial t} &= \left(-\frac{1}{2M}\frac{\partial^2}{\partial x^2} + E_b\right)\psi_b + \gamma A\psi_a. \end{aligned} \tag{105}$$

It is easily verified that the function $\theta(t-t') \times D(x,t;x',t')$ is the causal Green's function associated with Eq. (2):

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)[\theta(t-t')D(x,t;x',t')] = \delta(x-x')\delta(t-t'). \tag{106}$$

Therefore, assuming the interaction is turned on at time τ, we may write the first of Eqs. (105) as the integral equation

$$A(x,t) = A_0(x,t) + \int_B dx\int_\tau^\infty dt'\theta(t-t')D(x,t;x',t')j(x',t'). \tag{107}$$

Similar equations for ψ_a and ψ_b may be derived in terms of the Green's functions of the free atoms. The results of iterating these three integral equations are equivalent to the results obtained by solving Eq. (102) using the Dyson perturbation theory.[6]

XI. CONCLUDING REMARKS

We have shown how the theory of linearly polarized light propagating in a 1-dimensional cavity may be quantized by utilizing the symplectic structure of the space S. There would appear to be no great difficulty in extending our theory to the case of nonzero mass particles and to three spatial dimensions.

In applications of our theory to practical experimental situations, the creation of photons from the zero-point energy is altogether negligible. Since, in practice, mirror trajectories are not known exactly because of experimental errors, it is reasonable and mathematically simplifying to assume that they are always type-A trajectories.

The Dyson expansion does not provide a rigorous solution to the problem of including interactions in the theory. It may not converge and its individual terms may even be infinite. However, in this respect, having stationary mirrors is no particular advantage over having moving mirrors.

ACKNOWLEDGMENT

The author wishes to express his appreciation for the guidance and encouragement given him by his research advisor, H. N. Pendleton III.

* The research reported in this paper was supported by the U.S. Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract F19628-68-C-0062.

† Adapted from the author's Ph.D. thesis, submitted to Brandeis University in June, 1969.

[1] W. C. Henneberger and H. J. Schulte, J. Appl. Phys. **37**, 2189 (1966).

[2] P. W. Smith, Appl. Phys. Letters **10**, 51 (1967).

[3] R. I. Baranov and Yu. M. Shirokov, Zh. Eksp. Teor. Fiz. **53**, 2123 (1967) [Sov. Phys. JETP **26**, 1199 (1968)].

[4] Irving Segal, in *Mathematical Problems of Relativistic Physics, Summer Seminar, Boulder, Colorado, 1960* (American Mathematical Society, Providence, R.I., 1963), Vol. II.

[5] N. Solimene, Technical Note No. 135, TRG, Inc., Melville, N.Y., 1966.

[6] F. J. Dyson, Phys. Rev. **75**, 486 (1949).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Intermediate Statistics*

SHIGETOSHI KATSURA, KATSUJI KAMINISHI, AND SAKARI INAWASHIRO
Department of Applied Physics, Tohoku University, Sendai, Japan

(Received 4 March 1969; Revised Manuscript Received 13 April 1970)

The distinctions between intermediate statistics, parastatistics, and Okayama statistics are discussed and it is pointed out that the distribution function of the intermediate statistics does not follow from the para-Fermi statistics. The partition function, the pressure, and the specific heat of free particles which obey intermediate statistics are calculated in one, two, and three dimensions.

1. INTRODUCTION

Quantum statistics is classified as either Bose or Fermi statistics. The usual reasoning is as follows.

When a state Ψ is operated upon by any permutation P, the result is physically the same state, apart from the case of accidental degeneracy; thus, $P\Psi = c\Psi$. Take a transposition (i, k) as P, then $P^2\Psi = c^2\Psi = \Psi$; hence, $c = \pm 1$. The generalization of P to any permutation also gives $c = \pm 1$, which permits only symmetric and antisymmetric states. In the former, the maximum occupation number is infinity, and, in the latter, it is one. The commutation relations of creation and annihilation operators are

$$[a_k, a_l^+]_\pm = \delta_{kl}, \quad [a_k, a_l]_\pm = 0, \tag{1.1}$$

where $[\ ,\ ]_\pm$ represents the commutator for Bose statistics and the anticommutator for Fermi statistics.

The generalization of the quantum statistics has been carried out in several ways. One is to replace the above statement $P\Psi = c\Psi$ by the following statement: The expectation value of any observable A is the same in the state $P\Psi$ as in the state Ψ, i.e., $(\Psi, A\Psi) = (P\Psi, AP\Psi)$. The statistics generalized in this way is called parastatistics and was studied by Green,[1] Volkov,[2] Kamefuchi and Takahashi,[3] and others.

Parastatistics is classified as para-Fermi statistics and para-Bose statistics. In para-Fermi statistics of order p, a 1-particle state can be occupied by up to p particles. The case $p = 1$ in the para-Fermi statistics is the ordinary Fermi statistics. In para-Bose statistics of order p, a wavefunction in a state vector can be antisymmetric with respect to at most p particles, though the maximum occupation number is infinite. The case $p = 1$ in the para-Bose statistics is the ordinary Bose statistics. The commutation relations for $p = 2$ and 3 are obtained in Ref. 3. For example, for $p = 2$,

$$a_k a_l^+ a_m \pm a_m a_l^+ a_k = \pm 2\delta_{lm} a_k + 2\delta_{kl} a_m, \tag{1.2}$$

$$a_m a_k a_l^+ \pm a_l^+ a_k a_m = 2\delta_{kl} a_m, \tag{1.3}$$

$$a_k a_l a_m \pm a_m a_l a_k = 0, \tag{1.4}$$

space S. There would appear to be no great difficulty in extending our theory to the case of nonzero mass particles and to three spatial dimensions.

In applications of our theory to practical experimental situations, the creation of photons from the zero-point energy is altogether negligible. Since, in practice, mirror trajectories are not known exactly because of experimental errors, it is reasonable and mathematically simplifying to assume that they are always type-A trajectories.

The Dyson expansion does not provide a rigorous solution to the problem of including interactions in the theory. It may not converge and its individual terms may even be infinite. However, in this respect, having stationary mirrors is no particular advantage over having moving mirrors.

ACKNOWLEDGMENT

The author wishes to express his appreciation for the guidance and encouragement given him by his research advisor, H. N. Pendleton III.

* The research reported in this paper was supported by the U.S. Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract F19628-68-C-0062.

† Adapted from the author's Ph.D. thesis, submitted to Brandeis University in June, 1969.

[1] W. C. Henneberger and H. J. Schulte, J. Appl. Phys. **37**, 2189 (1966).
[2] P. W. Smith, Appl. Phys. Letters **10**, 51 (1967).
[3] R. I. Baranov and Yu. M. Shirokov, Zh. Eksp. Teor. Fiz. **53**, 2123 (1967) [Sov. Phys. JETP **26**, 1199 (1968)].
[4] Irving Segal, in *Mathematical Problems of Relativistic Physics, Summer Seminar, Boulder, Colorado, 1960* (American Mathematical Society, Providence, R.I., 1963), Vol. II.
[5] N. Solimene, Technical Note No. 135, TRG, Inc., Melville, N.Y., 1966.
[6] F. J. Dyson, Phys. Rev. **75**, 486 (1949).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Intermediate Statistics*

SHIGETOSHI KATSURA, KATSUJI KAMINISHI, AND SAKARI INAWASHIRO
Department of Applied Physics, Tohoku University, Sendai, Japan

(Received 4 March 1969; Revised Manuscript Received 13 April 1970)

The distinctions between intermediate statistics, parastatistics, and Okayama statistics are discussed and it is pointed out that the distribution function of the intermediate statistics does not follow from the para-Fermi statistics. The partition function, the pressure, and the specific heat of free particles which obey intermediate statistics are calculated in one, two, and three dimensions.

1. INTRODUCTION

Quantum statistics is classified as either Bose or Fermi statistics. The usual reasoning is as follows.

When a state Ψ is operated upon by any permutation P, the result is physically the same state, apart from the case of accidental degeneracy; thus, $P\Psi = c\Psi$. Take a transposition (i, k) as P, then $P^2\Psi = c^2\Psi = \Psi$; hence, $c = \pm 1$. The generalization of P to any permutation also gives $c = \pm 1$, which permits only symmetric and antisymmetric states. In the former, the maximum occupation number is infinity, and, in the latter, it is one. The commutation relations of creation and annihilation operators are

$$[a_k, a_l^+]_\pm = \delta_{kl}, \quad [a_k, a_l]_\pm = 0, \tag{1.1}$$

where $[\ ,\ ]_\pm$ represents the commutator for Bose statistics and the anticommutator for Fermi statistics.

The generalization of the quantum statistics has been carried out in several ways. One is to replace the above statement $P\Psi = c\Psi$ by the following statement: The expectation value of any observable A is the same in the state $P\Psi$ as in the state Ψ, i.e., $(\Psi, A\Psi) = (P\Psi, AP\Psi)$. The statistics generalized in this way is called parastatistics and was studied by Green,[1] Volkov,[2] Kamefuchi and Takahashi,[3] and others.

Parastatistics is classified as para-Fermi statistics and para-Bose statistics. In para-Fermi statistics of order p, a 1-particle state can be occupied by up to p particles. The case $p = 1$ in the para-Fermi statistics is the ordinary Fermi statistics. In para-Bose statistics of order p, a wavefunction in a state vector can be antisymmetric with respect to at most p particles, though the maximum occupation number is infinite. The case $p = 1$ in the para-Bose statistics is the ordinary Bose statistics. The commutation relations for $p = 2$ and 3 are obtained in Ref. 3. For example, for $p = 2$,

$$a_k a_l^+ a_m \pm a_m a_l^+ a_k = \pm 2\delta_{lm} a_k + 2\delta_{kl} a_m, \tag{1.2}$$

$$a_m a_k a_l^+ \pm a_l^+ a_k a_m = 2\delta_{kl} a_m, \tag{1.3}$$

$$a_k a_l a_m \pm a_m a_l a_k = 0, \tag{1.4}$$

where upper and lower signs correspond to para-Fermi and para-Bose cases, respectively. From these commutation relations, $a_{k_1}^+ a_{k_2}^+ |0\rangle$ is shown to be orthogonal to $a_{k_2}^+ a_{k_1}^+ |0\rangle$ for $k_1 \neq k_2$. For $p = 2$, in the possible nonzero $n!$ (n is the number of particles) state vectors $a_{k_1}^+ a_{k_2}^+ \cdots a_{k_n}^+ |0\rangle \equiv |1\, 2 \cdots n\rangle$, which are generated by operating n creation operators to zero vector, and state vectors, which are not identical on permutation of only alternate operators in the product of creation operators operating to zero vector,[3] are independent of each other. For example, for three particles with different k_1, k_2, and k_3, there are three independent states out of $3! = 6$, i.e.,

$$|123\rangle (= \mp |321\rangle),$$

$$|231\rangle (= \mp |132\rangle),$$

$$|312\rangle (= \mp |213\rangle).$$

The number of independent states is $n!/[(\frac{1}{2}n)!]^2$ for even n and $n!/\{[\frac{1}{2}(n-1)]!\,[\frac{1}{2}(n+1)]!\}$ for odd n, for all different n states.

The second way of the generalization is to postulate simply that the maximum number of particles in a 1-particle state to be a finite number which is denoted by ν, and that a state of a system is characterized by specifying a number of particles in each 1-particle state $\Psi = |n_1, n_2, \cdots\rangle = (a_{k_1}^+)^{n_1}(a_{k_2}^+)^{n_2} \cdots |0\rangle$, $n_i \leq \nu$. The statistics defined in this way was introduced by Gentile.[4] It is an intermediate one between Bose statistics and Fermi statistics and is called intermediate statistics. The creation and annihilation operators in the intermediate statistics are *commutable* for different k and realized by

$$a_k^+ = \begin{bmatrix} 0 & & & & & \\ 1 & & & & & \\ & \sqrt{2} & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & \sqrt{\nu} & 0 \end{bmatrix}_k,$$

$$a_k = \begin{bmatrix} 0 & 1 & & & & \\ & & \sqrt{2} & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \cdot \\ & & & & & \sqrt{\nu} \\ & & & & & 0 \end{bmatrix}_k. \qquad (1.5)$$

The commutation relations for $\nu = 2$ can be shown to be

$$a_k a_k^+ a_k - a_k^+ a_k a_k = a_k, \qquad (1.6)$$

$$a_k a_k a_k = 0, \qquad (1.7)$$

$$a_k a_k a_k^+ + 2a_k^+ a_k a_k = 2a_k, \qquad (1.8)$$

$$a_k a_k a_k^+ a_k^+ + a_k a_k^+ a_k^+ a_k + a_k^+ a_k^+ a_k a_k = 2, \qquad (1.9)$$

$$[a_k, a_l^+]_- = 0, \quad k \neq l, \qquad (1.10)$$

$$[a_k, a_l]_- = 0. \qquad (1.11)$$

The number of different state vectors for a given set of numbers of particles in the specified 1-particle states is one for intermediate statistics and not one, in general, for para-Fermi statistics. Thus, *the para-Fermi statistics does not lead to the intermediate statistics*. Intermediate statistics has no mathematically natural basis in any symmetry properties of wave-functions or any generalized field quantization scheme. Gentile[4] and Schubert[5] obtained the distribution functions in intermediate statistics. ter Haar[6] discussed the nature of the statistics and regarded the case of N (total number) $= \nu$ (maximum number in one state) as a physical case.

Application of the intermediate statistics to the Heisenberg model of the ferromagnetism is briefly mentioned. Usually, the magnon in the Heisenberg model is regarded as a boson, neglecting the kinematical interaction. In the system of N spins of $S = \frac{1}{2}$, no more than N reversed spins can exist. Then what difference would arise in treating the system as an assembly of bosons or as an assembly of particles obeying intermediate statistics of $\nu = N$? The susceptibility at zero field below the critical temperature is infinity for the former and finite for the latter. The details will be discussed in the future.

Another generalization of statistics is Okayama's statistics.[7] The commutation relations derived by Okayama are

$$a_k a_l^+ a_m - a_l^+ a_m a_k = \delta_{kl} a_m, \qquad (1.12)$$

$$\sum_{(\text{perm})} a_k a_l a_m = 0, \qquad (1.13)$$

and the other two are the same as Eqs. (1.8) and (1.9). Kamefuchi and Takahashi[3,8] showed that the commutation relations (1.12), (1.13), (1.8), and (1.9) give only null vectors in Hilbert space when applied to a system of many degrees of freedom[3] and that Okayama statistics is to be regarded as a modified version of parastatistics which provides a consistent theory only when applied to a system of just one degree of freedom.[8] Okayama statistics is *neither para-Fermi statistics nor intermediate statistics*. The distribution function in the intermediate statistics follows *neither* *from para-Fermi statistics nor from Okayama statistics* (in the latter, the distribution has no meaning).

In this paper, the grand partition function of the ideal gas obeying intermediate statistics is obtained, and physical quantities in 1-, 2-, and 3-dimensional cases are calculated. The results show a gradual shift of thermal properties from Fermi statistics to Bose statistics as ν increases from 1 to ∞.

2. THERMAL PROPERTIES OF IDEAL GAS OBEYING INTERMEDIATE STATISTICS

The grand partition function $\Xi(\alpha, \beta)$ for the system obeying intermediate statistics, of which the number of particles in one state is at most ν, is expressed by

$$\begin{aligned}\Xi(\alpha,\beta) &= \sum_{N=0}^{\infty} \sum_{\substack{n_k \\ \Sigma n_k = N \\ \max(n_k)=\nu}} \exp\left[-\left(\beta \sum_k n_k E_k + \alpha N\right)\right] \\ &= \prod_k (1 + e^{-(\alpha+\beta E_k)} + e^{-2(\alpha+\beta E_k)} + \cdots \\ &\quad + e^{-\nu(\alpha+\beta E_k)}) \\ &= \prod_k \frac{1 - e^{-(\alpha+\beta E_k)(\nu+1)}}{1 - e^{-(\alpha+\beta E_k)}}, \end{aligned} \tag{2.1}$$

where $\beta = 1/k_B T$, $\alpha = -\beta\mu$, k_B is the Boltzmann constant, T the absolute temperature, μ the chemical potential, N the number operator, and k labels a 1-particle state. The number of ways of placing particles in n distinct boxes with no more than ν particles was given by ter Haar[6] and by Fisher.[9]

The differentiation of Ξ with respect to βE_k gives directly the average number of particles in the state k, that is,

$$\begin{aligned}\bar{n}_k &= -\frac{\partial \log \Xi}{\partial \beta E_k} \\ &= \frac{1}{e^{\alpha+\beta E_k} - 1} - \frac{\nu+1}{e^{(\alpha+\beta E_k)(\nu+1)} - 1}. \end{aligned} \tag{2.2}$$

The result agrees with that found by Gentile[4] and others.[5,6]

Consider N free particles (obeying intermediate statistics) contained in the n-dimensional box each of whose edge lengths is $L(= V^{1/n})$. Periodic boundary conditions are imposed.

The energy eigenvalue $E_{\mu_1\mu_2\cdots\mu_n}$ for the free particle in the periodic condition is

$$E_{\mu_1\mu_2\cdots\mu_n} = \sum_{i=1}^{n} \frac{\hbar^2}{2m} \frac{\pi^2}{L^2} \mu_i^2, \tag{2.3}$$

where $\mu_i = 0, \pm 1, \pm 2, \cdots$ and m is the mass of the particle. When the volume of the system $V(\equiv L^n)$ is sufficiently large, the density of energy levels $g(E)$ is obtained from Eq. (2.3) as

$$g(E) = \frac{V}{\Gamma(\frac{1}{2}n)} \left(\frac{2m}{4\pi\hbar^2}\right)^{\frac{1}{2}n} E^{\frac{1}{2}n-1}. \tag{2.4}$$

In the limit $V \to \infty$, $\log \Xi(\alpha, \beta)$ reduces to

$$\begin{aligned}&\lim_{V\to\infty} \frac{1}{V} \log \Xi(\alpha,\beta) \\ &\quad = \frac{1}{V} \int_0^\infty g(E) \log\left(\frac{1 - e^{-(\alpha+\beta E)(\nu+1)}}{1 - e^{-(\alpha+\beta E)}}\right) dE. \end{aligned} \tag{2.5}$$

Substituting Eq. (2.4) into Eq. (2.5) and integrating by parts, we have

$$\begin{aligned}\frac{1}{V} \log \Xi(\alpha,\beta) &= \frac{1}{\Gamma(\frac{1}{2}n + 1)} \frac{1}{\lambda^n} \int_0^\infty \left(\frac{e^{-\alpha}}{e^x - e^{-\alpha}}\right. \\ &\quad \left. - \frac{(\nu+1)e^{-\alpha(\nu+1)}}{e^{x(\nu+1)} - e^{-\alpha(\nu+1)}}\right) x^{\frac{1}{2}n}\, dx, \end{aligned} \tag{2.6}$$

where

$$\lambda \equiv h(2\pi m k_B T)^{-\frac{1}{2}}.$$

The integrals of the first term and the second term in the bracket in Eq. (2.6) converge when $\alpha > 0$. The integral of the difference, however, converges even when $\alpha < 0$. The integral of the difference for $\alpha < 0$ is equal to the difference of the analytic continuations (to the regions $\alpha < 0$) of the analytic functions defined by the integral of the first term for $\alpha > 0$ and that for the second term.

The function $\phi(z, s)$ is defined by[10–12]

$$\phi(z,s) \equiv \sum_{n=1}^{\infty} \frac{z^n}{n^s}$$

and its analytic continuations. It has integral representation

$$\phi(z,s) \equiv \frac{z}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}}{e^t - z}\, dt, \quad \text{when } z < 1. \tag{2.7}$$

We define

$$\kappa(\alpha,s,\nu) \equiv \phi(e^{-\alpha},s) - (\nu+1)^{1-s}\phi(e^{-\alpha(\nu+1)},s), \quad -\infty < \alpha < \infty. \tag{2.8}$$

From Eq. (2.6), we have

$$N = -\frac{\partial \log \Xi(\alpha,\beta)}{\partial \alpha} = \frac{V}{\lambda^n} \kappa(\alpha, \tfrac{1}{2}n, \nu), \tag{2.9}$$

$$E = -\frac{\partial \log \Xi(\alpha,\beta)}{\partial \beta} = \tfrac{1}{2}n \frac{V}{\beta\lambda^n} \kappa(\alpha, \tfrac{1}{2}n + 1, \nu), \tag{2.10}$$

$$p = \frac{1}{V} \frac{\log \Xi(\alpha,\beta)}{\beta} = \frac{1}{\beta\lambda^n} \kappa(\alpha, \tfrac{1}{2}n + 1, \nu), \tag{2.11}$$

for the mean number N, energy E, and the pressure p.

The specific heat at constant volume C_V is given by

$$C_V = \left(\frac{\partial E}{\partial T}\right)_{V,N} = \tfrac{1}{2}k_B n \frac{V}{\lambda^n}(\tfrac{1}{2}n + 1)\kappa(\alpha, \tfrac{1}{2}n + 1, \nu) - \tfrac{1}{2}n\frac{V}{\lambda^n \beta}\left(\frac{\partial \alpha}{\partial T}\right)_{T,N}\kappa(\alpha, \tfrac{1}{2}n, \nu). \quad (2.12)$$

Since the partial differentiation of Eq. (2.9) gives

$$\left(\frac{\partial \alpha}{\partial T}\right)_{V,N}\kappa(\alpha, \tfrac{1}{2}n - 1, \nu) = \frac{n}{2T}\kappa(\alpha, \tfrac{1}{2}n, \nu), \quad (2.13)$$

the specific heat becomes

$$C_V = \tfrac{1}{2}nNk_B\left((\tfrac{1}{2}n + 1)\frac{\kappa(\alpha, \tfrac{1}{2}n + 1, \nu)}{\kappa(\alpha, \tfrac{1}{2}n, \nu)} - \tfrac{1}{2}n\frac{\kappa(\alpha, \tfrac{1}{2}n, \nu)}{\kappa(\alpha, \tfrac{1}{2}n - 1, \nu)}\right). \quad (2.14)$$

For the purpose of introducing an explicit expression in terms of T, we define T_0 by

$$V/N = v = (\lambda^n)_{T=T_0}. \quad (2.15)$$

Then Eq. (2.9), where v is the specific volume, can be written in the form

$$T/T_0 = [\kappa(\alpha, \tfrac{1}{2}n, \nu)]^{-2/n}. \quad (2.9')$$

In the specific heat (2.14), α is regarded as a function of the temperature T/T_0, with use of (2.9′). In the pressure equation (2.11), α is regarded as a function of the specific volume $(V/N)\lambda^{-n}$ by Eq. (2.9).

In Eq. (2.8), the singular point $z = 1$ in $\phi(z, s)$ is canceled by subtraction, and $\alpha = 0$ is not a singular point in $\kappa(\alpha, s, \nu)$ for finite ν. Hence, the specific heat and the pressure in the intermediate statistics for finite ν have no singularity. Since $\kappa(\alpha, s, 1) = -\phi(-e^{-\alpha}, s)$ and $\kappa(\alpha, s, \infty) = \phi(e^{-\alpha}, s)$, the specific heat (2.14) for Fermi statistics and that for Bose statistics are the same function of T/T_0 when α is eliminated by Eq. (2.9′) in the case of two dimensions. This was pointed out by Toda and Takano[13] and by May.[14]

3. HIGH- AND LOW-TEMPERATURE LIMIT AND HIGH- AND LOW-DENSITY LIMIT

When $\alpha \to +\infty$, $\kappa(\alpha, s, \nu)$ tends to $e^{-\alpha}$. Then we have

$$p\beta = 1/v \quad (3.1)$$

and

$$C_V/k_B N = \tfrac{1}{2}n, \quad (3.2)$$

irrespective of ν and n. This means the results of classical statistics are obtained, when the specific volume is large or the temperature is high.

Consider the case $\alpha \to -\infty$. Using the asymptotic expansions (A4) and (A5), we have

$$\frac{\lambda^n}{v} \simeq \frac{\nu}{\Gamma(\tfrac{1}{2}n + 1)}(-\alpha)^{\frac{1}{2}n} \quad (3.3)$$

and

$$p\beta\lambda^n \simeq \frac{\nu}{\Gamma(\tfrac{1}{2}n + 2)}(-\alpha)^{\frac{1}{2}n+1}. \quad (3.4)$$

Eliminating α from Eqs. (3.3) and (3.4), we have

$$p\beta\lambda^n = \frac{[\Gamma(\tfrac{1}{2}n + 1)]^{(n+2)/n}}{\Gamma(\tfrac{1}{2}n + 2)\nu^{2/n}}\left(\frac{\lambda^n}{v}\right)^{(n+2)/n}. \quad (3.5)$$

Equation (3.5) holds as $\lambda^n/v \to \infty$ for fixed ν. It does not hold for fixed λ^n/v as $\nu \to \infty$.

Inserting the first three terms of (A4) or (A5) into Eq. (2.8) and using it in Eq. (2.14), we have

$$\frac{2C_V}{nNk_B} = \tfrac{2}{3}\pi^2\frac{1}{\nu + 1}(-\alpha)^{-1} \quad (3.6)$$

$$= \tfrac{2}{3}\pi^2\frac{1}{\nu + 1}\left(\frac{v}{\lambda^n}\frac{\nu}{\Gamma(\tfrac{1}{2}n + 1)}\right)^{2/n}$$

$$= \tfrac{2}{3}\pi^2\frac{1}{\nu + 1}\left(\frac{\nu}{\Gamma(\tfrac{1}{2}n + 1)}\right)^{2/n}\frac{T}{T_0}, \quad (3.7)$$

after the cancellation of the contribution of the first and the second terms. Again, Eq. (3.7) is not to be used as $\nu \to \infty$ for fixed T, but for fixed ν as $T \to 0$.

For the case $\nu = \infty$ (Bose statistics), the limit $\alpha \to 0$ corresponds to $v \to 0$ or $v \to v_c$. In this limit, we have

$$\lambda^n/v = \Gamma(-\tfrac{1}{2}n + 1)\alpha^{\frac{1}{2}n-1} + \zeta(\tfrac{1}{2}n) - \alpha\zeta(\tfrac{1}{2}n - 1), \quad n = 1, 3, \quad (3.8)$$

$$\lambda^n/v = -\log[1 - e^{-\alpha}], \quad n = 2,$$

$$p\beta\lambda^n = \Gamma(-\tfrac{1}{2}n)\alpha^{\frac{1}{2}n} + \zeta(\tfrac{1}{2}n + 1) - \alpha\zeta(\tfrac{1}{2}n), \quad n = 1, 2, 3, \quad (3.9)$$

using the Lindelöf expansion. Here, $\zeta(s)$ is the Riemann ζ function. Thus,

$$p\beta\lambda^3 \to \zeta(\tfrac{5}{2}) - \frac{\zeta(\tfrac{3}{2})}{4\pi}\left(\frac{\lambda^3}{v} - \frac{\lambda^3}{v_c}\right)^2,$$

$$\lambda^3/v \to \lambda^3/v_c \equiv \zeta(\tfrac{3}{2}), \quad n = 3,$$

$$p\beta\lambda^2 \to \zeta(2) - (\lambda^2/v)e^{-\lambda^2/v},$$

$$\lambda^2/v \to 0, \quad n = 2,$$

$$p\beta\lambda \to \zeta(\tfrac{3}{2}) - 2\pi(v/\lambda),$$

$$\lambda/v \to 0, \quad n = 1. \quad (3.10)$$

When we substitute the first two terms of the Lindelöf expansion to Eq. (2.14) and retain the leading

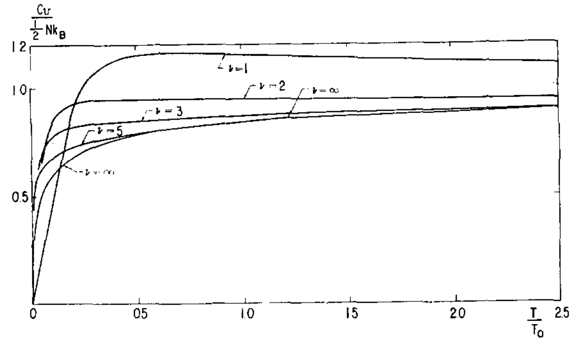


FIG. 1. Specific heat in one dimension.

term, we have

$$\frac{2C_V}{Nk_B} = \tfrac{3}{2}\zeta(\tfrac{3}{2})\frac{1}{\Gamma(\frac{1}{2})}\alpha^{\frac{1}{2}} \tag{3.11}$$

$$= \tfrac{3}{2}\zeta(\tfrac{3}{2})\left(\frac{T}{T_0}\right)^{\frac{1}{2}} \tag{3.12}$$

for one dimension,

$$\frac{C_V}{Nk_B} = -\frac{2\zeta(2)}{\log(1-e^{-\alpha})} \tag{3.13}$$

$$= 2\zeta(2)\frac{T}{T_0} \tag{3.14}$$

for two dimensions, and

$$\frac{2C_V}{3Nk_B} = \frac{5}{2}\frac{\zeta(\frac{5}{2})}{\zeta(\frac{3}{2})} - \frac{3\zeta(\frac{3}{2})}{2\sqrt{\pi}}\alpha^{\frac{1}{2}} \tag{3.15}$$

for three dimensions. Defining λ_c and T_c by

$$\frac{\lambda_c^3}{v} = \zeta(\tfrac{3}{2}) = \left(\frac{T_0}{T_c}\right)^{\frac{3}{2}} \tag{3.16}$$

and eliminating α from Eqs. (3.8) and (3.15), we have

$$\frac{2C_V}{3Nk_B} = \frac{5}{2}\frac{\zeta(\frac{5}{2})}{\zeta(\frac{3}{2})} - \frac{3\zeta(\frac{3}{2})}{(2\sqrt{2})\pi}\left[\left(\frac{T_0}{T_c}\right)^{\frac{3}{2}} - \left(\frac{T_0}{T}\right)^{\frac{3}{2}}\right] \tag{3.17}$$

for $T \geqslant T_c$. For $T < T_c$, let $\alpha = 0$ and $\lambda^3/v = \zeta(\frac{3}{2})$; then

$$\frac{2C_V}{3Nk_B} = \tfrac{5}{2}\zeta(\tfrac{5}{2})\frac{v}{\lambda^3}$$

$$= \tfrac{5}{2}\zeta(\tfrac{5}{2})\left(\frac{T}{T_0}\right)^{\frac{3}{2}}. \tag{3.18}$$

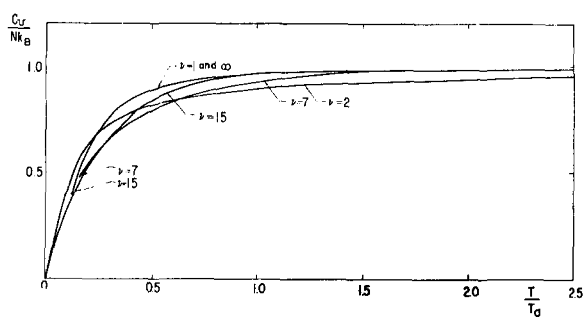


FIG. 2. Specific heat in two dimensions.

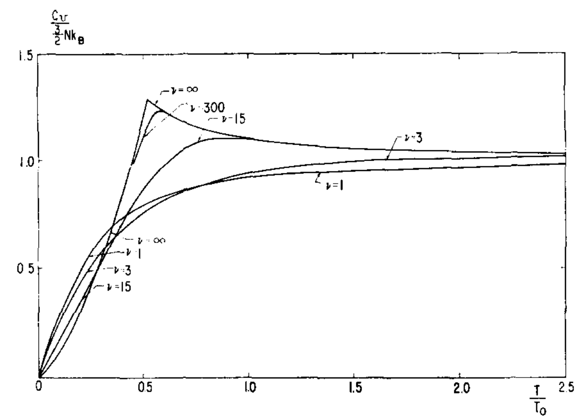


FIG. 3. Specific heat in three dimensions.

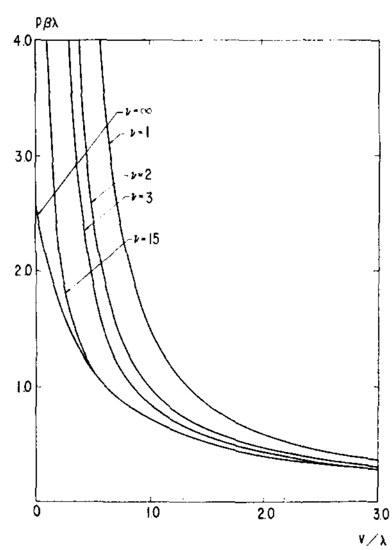


FIG. 4. Pressure in one dimension.

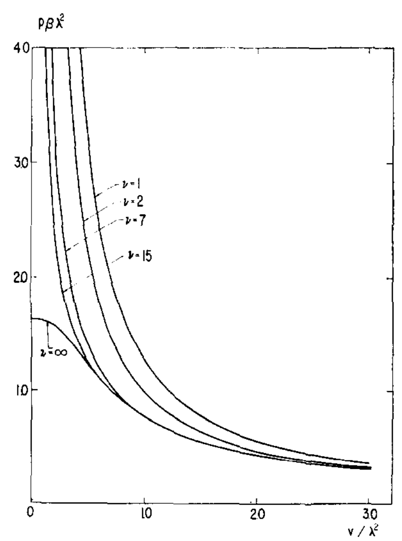


FIG. 5. Pressure in three dimensions.

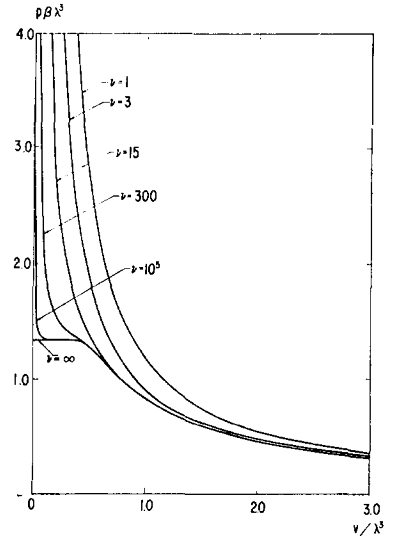


FIG. 6. Pressure in three dimensions.

These equations for the ideal Bose condensation have been derived for the sake of comparison.

The specific heat vs reduced temperature is shown in Figs. 1–3 and the pressure vs volume is shown in Figs. 4–6, in 1-, 2-, and 3-dimensional cases, respectively. These figures show gradual shifts from Fermi to Bose statistics as ν increases.

When ν is small, the increase of ν causes noticeable changes of physical quantities. The difference between $\nu = 1$ and $\nu = 2$ is especially large. When ν becomes large, the shift to Bose statistics becomes slower as the dimension increases.

The specific heat for finite ν is larger at low temperature and smaller at high temperature than that for $\nu = \infty$ (Bose statistics). The pressure decreases monotonically as ν increases and diverges as $V/N \to 0$, except in the case $\nu = \infty$. For large V/N or large T/T_0, the pressure and the specific heat tend to those of the classical ideal gas.

ACKNOWLEDGMENTS

The authors thank Professor K. Hiroike for his valuable discussions and Y. Yamazaki for his aid in programming.

APPENDIX: PROPERTIES OF $\phi(z, s)$

The function $\phi(z, s)$ (see Fig. 7) is defined by

$$\phi(z, s) = \sum_{n=1}^{\infty} \frac{z^n}{n^s}, \quad |z| < 1, \tag{A1}$$

and its analytic continuations.

For $s = 0, 1, 2$, Eq. (A1) reduces to

$$\phi(z, 0) = z/(1 - z),$$
$$\phi(z, 1) = -\log (1 - z),$$
$$\phi(z, 2) \equiv L_2(z) \quad \text{(dilogarithm)}.$$

The analytic continuation of (A1) is given by the Lindelöf expansion[10]

$$\phi(z, s) = \Gamma(1 - s)(-\log z)^{s-1} + \sum_{n=0}^{\infty} \zeta(s - n) \frac{(\log z)^n}{n!}, \quad |\log z| < 2\pi, \tag{A2}$$

for $s \neq 1, 2, \cdots$.

Taking the limit $s \to m$, $m = 2, 3, \cdots$, we have[11,15]

$$\phi(z, m) = \frac{(\log z)^{m-1}}{(m-1)!}\left[\psi(m) - \psi(1) - \log\log\left(\frac{1}{z}\right)\right] + \sum_{\substack{n=0 \\ n \neq m-1}}^{\infty}{}' \zeta(m - n) \frac{(\log z)^n}{n!}, \quad m = 2, 3, 4, \cdots, \tag{A3}$$

where $\psi(z) = \Gamma'(z)/\Gamma(z)$.

Asymptotic expansions of $\phi(z, s)$ and $\phi(\exp(\pi i)z, s)$ for nonintegral values of s are given by Katsura[12]:

$$\phi(z, s) = -\frac{(\log z)^s}{\Gamma(s+1)} + \Gamma(1 - s)(-\log z)^{s-1} + \sum_{n=1}^{m} \frac{(2\pi)^{2n} B_n}{\Gamma(s + 1 - 2n)(2n)!} (\log z)^{s-2n} + O((\log z)^{s-2m-2}), \tag{A4}$$

$$\phi(\exp(\pi i)z, s) = -\frac{(\log z)^s}{\Gamma(s+1)} - \sum_{n=1}^{m} (1 - 2^{1-2n}) \times \frac{(2\pi)^{2n} B_n}{\Gamma(s + 1 - 2n)(2n)!} (\log z)^{s-2n} + O((\log z)^{s-2m-2}), \tag{A5}$$

where B_n is the Bernoulli number.[16]

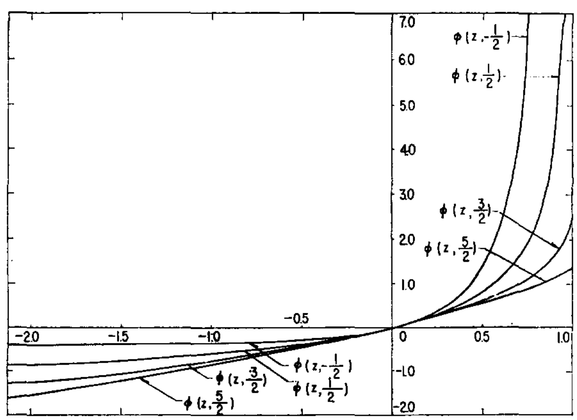


FIG. 7. The function $\phi(z, s)$.

Taking the limit $s \to m$, $m = 2, 3, \cdots$, we have

$$\phi(z,m) = -\frac{(\log z)^m}{m} + \frac{(\log z)^{m-1}}{(m-1)!}[\psi(m) - \psi(1) - \log\log(z^{-1})] + \sum_{n=1}^{m\geq 2n} \frac{(2\pi)^{2n}B_n}{(m-2n)!\,(2n)!}(\log z)^{m-2n} + O((\log z)^{-1}). \quad \text{(A6)}$$

For $m = 2$, the equation

$$L_2(z) + L_2(1-z) = \tfrac{1}{6}\pi^2 - \log z \log(1-z) \quad \text{(A7)}$$

holds.[17]

Tables of $\phi(z, \frac{5}{2})$, $\phi(z, 2)$, $\phi(z, \frac{3}{2})$, $\phi(z, \frac{1}{2})$, and $\phi(z, -\frac{1}{2})$ were calculated using these formulas.

Tables of Fermi–Dirac functions

$$F_{s-1}(\alpha) = \Gamma(s)\phi(-\exp(\alpha), s) \quad \text{(A8)}$$

by McDougall and Stoner[18] and by Chisnall,[19] together with the relation

$$\phi(-z, s) = \phi(z, s) - 2^{1-s}\phi(z^2, s), \quad \text{(A9)}$$

were also useful for numerical calculation for some part.

These properties of $\phi(z, s)$ functions can be used in several topics in the statistical mechanics.[20]

* Preliminary report of the 3-dimensional part of this work was given in the meeting of the Physical Society of Japan, April, 1954, Bull. Phys. Soc. Japan **3**, 2 (1954).

1 H. S. Green, Phys. Rev. **90**, 270 (1953).
2 D. V. Volkov, Zh. Eksp. Teor. Fiz. **36**, 1560 (1959); **38**, 518 (1960) [Sov. Phys. JETP **9**, 1107 (1959); **11**, 375 (1960)].
3 S. Kamefuchi and Y. Takahashi, Nucl. Phys. **36**, 177 (1962); **42**, 166 (1963).
4 G. Gentile, Nuovo Cimento **17**, 493 (1940).
5 G. Schubert, Z. Naturforsch. **1**, 113 (1946).
6 D. ter Haar, Physica **18**, 199 (1952).
7 T. Okayama, Progr. Theoret. Phys. (Kyoto) **7**, 597 (1952).
8 S. Kamefuchi and Y. Takahashi, Progr. Theoret. Phys. (Kyoto) Suppl. No. **37–38**, 244 (1966).
9 M. E. Fisher, Am. J. Phys. **30**, 49 (1962).
10 E. Lindelöf, *Le Calcul des résidus et ses applications à la théorie des fonctions* (Gauthier–Villars, Paris, 1905), p. 138.
11 C. Truesdell, Ann. Math. **46**, 144 (1945). See also J. E. Robinson, Phys. Rev. **83**, 687 (1951), and J. Clunie, Proc. Phys. Soc. **A67**, 632 (1954).
12 S. Katsura, Progr. Theoret. Phys. (Kyoto) **16**, 589 (1956).
13 M. Toda and F. Takano, J. Phys. Soc. Japan **9**, 14 (1954).
14 R. May. Phys. Rev. **135**, A1515 (1964).
15 *Higher Transcendental Functions*, A. Erdélyi, Ed. (McGraw-Hill, New York, 1953), Vol. 1, p. 30.
16 The notation of Bernoulli numbers is the same as in E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge U.P., London, 1927).
17 L. Lewin, *Dilogarithms and Associated Functions* (MacDonald, London, 1958).
18 J. McDougall and E. C. Stoner, Phil. Trans. Roy. Soc. **A237**, 67 (1938).
19 G. A. Chisnall, Jodrell Bank Ann. **1**, 126 (1956).
20 S. Katsura, *Statistical Mechanics* (in Japanese) (Hirokawa Shoten, Tokyo, 1969).

Finite and Infinitesimal Canonical Transformations*

FRANCIS J. TESTA
Drexel Institute of Technology, Philadelphia, Pennsylvania 19104

(Received 6 March 1970)

The general relation between the infinitesimal generator of a 1-parameter subgroup of canonical transformations and the usual finite generating functions is obtained. This relation is found to be simply a generalization of the Hamilton–Jacobi equation. When the latter relation is solved for the finite generating function, the connections for a finite transformation can be determined without the need for integration of the associated infinitesimal transformation.

1. INTRODUCTION

Within the framework of classical dynamical theory, one usually defines canonical transformations on the coordinates and momenta

$$Q = Q(q,p), \qquad P = P(q,p),$$
$$q \equiv (q_1, q_2, \cdots, q_n), \quad p \equiv (p_1, p_2, \cdots, p_n) \quad (1)$$

such that Hamilton's canonical equations remain invariant. We denote the family of all such transformations as

$$\mathcal{T} \equiv \left\{T : \begin{pmatrix} Q \\ P \end{pmatrix} = T\begin{pmatrix} q \\ p \end{pmatrix}\right\}.$$

This family of transformations can be represented by points in a 2-dimensional Euclidean phase space Γ with coordinates $(Q(q,p), P(q,p))$. The point (q,p) is associated with the identity transformation.

The connections (1) can be generated by assuming a generating function F, explicit in one old and one new canonical variable n-tuple. In the notation of Goldstein[1] consider

$$F_1 \equiv F_1(q, Q), \quad p = \frac{\partial F_1}{\partial q}, \quad P = -\frac{\partial F_1}{\partial Q}. \quad (2)$$

It can be easily shown that the transformation T generated in this manner is a member of $\mathcal{T}$. One can then associate with a particular function F_1 a unique point in Γ space.

By a linear Legendre transformation one can produce functions explicit in other variables. For example, consider

$$F_2(q, P) \equiv \left(\mathbb{1} - Q \cdot \frac{\partial}{\partial Q}\right) F_1(q, Q),$$
$$p = \frac{\partial F_2}{\partial q}, \quad Q = \frac{\partial F_2}{\partial P}, \quad (3)$$

where a dot denotes n-tuple contraction.

Similarly, the functions $F_3(p, P)$ and $F_4(p, Q)$ can be generated. It is apparent, however, that, since mapping by Legendre transformation is not one-to-one, functions of one type, F_k, cannot be placed in one-to-one correspondence with points in Γ space. For example, take $F_2(q, P) = q \cdot P$. Then from (3) we have

$$p = \frac{\partial F_2}{\partial q} = P, \quad Q = \frac{\partial F_2}{\partial P} = q,$$

and $F_2(q, P) = q \cdot P$ generates the identity transformation. However, inverting (3), we have

$$F_1(q, Q) = \left(\mathbb{1} - P \cdot \frac{\partial}{\partial P}\right) F_2(q, P) \equiv 0$$

and, hence, representation in terms of functions F_1 alone is not possible. It is now clear that, in order to associate a generating function F with each point in Γ space, one must employ functions of more than one type.

The fact that $\mathcal{T}$ actually constitutes a group of transformations allows the use of Lie representation theory and avoids the ad hoc mathematical description in terms of the generating functions F_k. Consider an arbitrary observable $w(q,p)$ and its associated transformation

$$\begin{pmatrix} Q \\ P \end{pmatrix} = T(\tau)\begin{pmatrix} q \\ p \end{pmatrix}, \quad (4)$$

where

$$T(\tau) \equiv \exp\left[\tau\left(\frac{\partial w}{\partial p} \cdot \frac{\partial}{\partial q} - \frac{\partial w}{\partial q} \cdot \frac{\partial}{\partial p}\right)\right],$$

with τ a real 1-tuple parameter. It can be shown[2] that the operators $T(\tau)$ generate canonical transformations for arbitrary $w(q,p)$. One then associates with each observable $w(q,p)$ a 1-parameter subgroup of transformations $\mathcal{T}_w$ characterized in Γ space by a continuous curve emanating from the identity element $(Q, P) \equiv (q, p)$. An arbitrary canonical transformation can be generated by a representation of the form (4).

Although the more lucid Lie representation is convenient from the point of view of infinitesimal transformations, the generation of finite transformations requires the evaluation of the formal operator $T(\tau)$ in closed form. Since the latter evaluation is not always technically feasible, a procedure analogous to (2) is desirable. It is evident that we must somehow generalize the generating functions F_k to incorporate the parameter τ such that all elements of $\mathcal{C}_w$ may be generated. In order to circumvent any direct evaluation of $T(\tau)$ for a given observable $w(q,p)$, the connection between $w(q,p)$ and the new generating functions $\mathcal{F}_k$ must be made in the neighborhood of the identity transformation.

With this objective in mind, realizing that only the functions F_2 and F_4 can be associated with the identity transformation, we consider the following generalization of the function F_2 which incorporates the parameter τ, subject to the assumption that $\tau = 0$ generates the identity transformation. We define $\mathcal{F}_2 \equiv \mathcal{F}_2(q, P, \tau)$ such that

$$p(q, P, \tau) = \frac{\partial \mathcal{F}_2}{\partial q}, \quad Q(q, P, \tau) = \frac{\partial \mathcal{F}_2}{\partial P}. \tag{5}$$

The function $\mathcal{F}_2(q, P, \tau)$ characterizes the subgroup of transformations $\mathcal{C}_w$ with possible singularities for specific values of τ. In the analysis that follows, we obtain the general relation between the functions $\mathcal{F}_2$ and w. Direct evaluation of $T(\tau)$ is obviated by solving the relation below for $\mathcal{F}_2$.

2. GENERAL RELATION BETWEEN FINITE AND INFINITESIMAL GENERATORS

For an arbitrary observable $g(q,p)$, one defines the canonically transformed observable $G(q,p,\tau)$ by

$$G(q,p,\tau) \equiv T(\tau)g(q,p).$$

It can be easily shown that

$$\frac{dG}{d\tau} = \frac{\partial G}{\partial q}\cdot\frac{\partial w}{\partial p} - \frac{\partial G}{\partial p}\cdot\frac{\partial w}{\partial q} \equiv [G, w], \tag{6}$$

the Poisson bracket of G with w. Since calculation of a Poisson bracket can be made with respect to any pair of canonical variable n-tuples, we have

$$\frac{\partial G}{\partial q}\cdot\frac{\partial w}{\partial p} - \frac{\partial G}{\partial p}\cdot\frac{\partial w}{\partial q} = \frac{\partial G}{\partial Q}\cdot\frac{\partial w}{\partial P} - \frac{\partial G}{\partial P}\cdot\frac{\partial w}{\partial Q}. \tag{7}$$

Using (5) and (6), one obtains

$$\frac{dQ}{d\tau} = \frac{d}{d\tau}\left(\frac{\partial \mathcal{F}_2}{\partial P}\right) = [Q, w].$$

Employing (6) and (7) with slight rearrangement, we have

$$\left(\frac{\partial}{\partial P}\left(\frac{\partial \mathcal{F}_2}{\partial \tau}\right)\right)_{q,\tau} = \left(\frac{\partial w}{\partial P}\right)_Q + \left(\frac{\partial w}{\partial Q}\right)_P\cdot\left(\frac{\partial Q}{\partial P}\right)_{q,\tau} = \left(\frac{\partial w}{\partial P}\right)_{q,\tau}. \tag{8}$$

Similarly, we have

$$\frac{dp}{d\tau} = 0 = \frac{d}{d\tau}\left(\frac{\partial \mathcal{F}_2}{\partial q}\right).$$

Using (6) yields

$$\left(\frac{\partial}{\partial q}\left(\frac{\partial \mathcal{F}_2}{\partial \tau}\right)\right)_{P,\tau} = \left(\frac{\partial w}{\partial q}\right)_{P,\tau}. \tag{9}$$

We have repeatedly used the fact that, under $\mathcal{C}_w$,

$$W(q,p,\tau) \equiv T(\tau)w(q,p) = w(q,p)$$

which, together with the function theorem for Lie representations (see the Appendix), implies

$$w(q,p) = w(Q(q,p,\tau), P(q,p,\tau)). \tag{10}$$

Integrating either (8) or (9) and using the other to simplify the additive function yield

$$\left(\frac{\partial \mathcal{F}_2}{\partial \tau}\right)_{P,q} = w(Q(q,P,\tau),P) + \phi(\tau).$$

Since, for fixed w, all generators of the form $w + \phi(\tau)$ constitute an equivalence class in that they represent the same curve in Γ space, we take $\phi(\tau) \equiv 0$. Hence, using (10), we have

$$\frac{\partial \mathcal{F}_2}{\partial \tau} = w\left(\frac{\partial \mathcal{F}_2}{\partial P}, P\right) = w\left(q, \frac{\partial \mathcal{F}_2}{\partial q}\right), \tag{11}$$

which can be solved for functions $\mathcal{F}_2(q, P, \tau)$. To fix any undetermined constants, we impose the boundary conditions at $\tau = 0$,

$$\mathcal{F}_2(q, P, 0) = q\cdot P = q\cdot p,$$

$$\left.\frac{\partial \mathcal{F}_2}{\partial \tau}\right|_{\tau=0} = w(q, p(q, P, 0)) = w(q, P). \tag{12}$$

The resulting function generates the 1-parameter subgroup of transformations $\mathcal{C}_w$ except at possible singular points in τ. Existence of such singularities is substantiated by the fact that, although the operators $T(\tau)$ can be placed in one-to-one correspondence with points in Γ space, the family of functions $\mathcal{F}_2$ cannot. Since $\mathcal{F}_2$ and $\mathcal{F}_4$ have no common explicit n-tuple dependence, they cannot be singular simultaneously; hence, one can always find a finite generator for an arbitrary observable $w(q,p)$.

It is also noteworthy that, by taking $w = -H$, the Hamiltonian, and $\tau = t$, we see that $T(\tau)$ produces transformations backwards in time from $(q(t), p(t)) \to (q(0), p(0))$. Then one obtains

$$\frac{\partial}{\partial t}(\mathcal{F}_2(q(t), p(0), t)) = -H\left(\frac{\partial \mathcal{F}_2}{\partial p(0)}, p(0)\right) = -H\left(q(t), \frac{\partial \mathcal{F}_2}{\partial q(t)}\right), \tag{13}$$

which is the familiar Hamilton–Jacobi equation. Hence, (11) is simply the generalization of (13) to the arbitrary infinitesimal generator $w(q, p)$. In analogous fashion, $\mathcal{F}_2$ is simply the generalization of the action integral

$$\mathcal{F}_2[Q, P] = \int_0^\tau \left(Q \cdot \frac{dP}{d\tau} + w(Q, P)\right) d\tau + q \cdot p$$

to within a constant additive term, which is necessary to satisfy (12).

Relations similar to (11) can be deduced for each of the other generating functions $\mathcal{F}_k$. For example, one can show in a similar manner that

$$\frac{\partial \mathcal{F}_1}{\partial \tau} = -w\left(q, \frac{\partial \mathcal{F}_1}{\partial q}\right) = -w\left(Q, -\frac{\partial \mathcal{F}_1}{\partial Q}\right). \quad (14)$$

However, to determine a unique function $\mathcal{F}_1$, one must impose boundary conditions in a region where $\mathcal{F}_1$ is nonsingular. Since this is not possible in the neighborhood of the identity, it would be necessary to evaluate $T(\tau)$ at $\tau = \tau_0$ such that representation of $T(\tau_0)$ in terms of $\mathcal{F}_1$ would be nonsingular. But direct evaluation of $T(\tau)$ is precisely what we have tried to avoid. Consequently, (14) is not useful from the standpoint of computational convenience. Similar conclusions hold for the relation in $\mathcal{F}_3$. The first equality in (14) has been deduced previously.[3] To illustrate our method for obtaining finite transformations of the form (4) from the infinitesimal generator $w(q, p)$, we now consider a specific example. Suppose that $w = f(q)p + g(q)$ with $n = 1$ for simplicity. Substitution into (11) yields

$$\frac{\partial \mathcal{F}_2}{\partial \tau} = f(q)\frac{\partial \mathcal{F}_2}{\partial q} + g(q). \quad (15)$$

Under the change of variable

$$z \equiv \int \frac{dq}{f(q)},$$

we obtain

$$\frac{\partial \mathcal{F}_2}{\partial \tau} = \frac{\partial \mathcal{F}_2}{\partial z} + \hat{g}(z),$$

where $\hat{g}(z) \equiv g(q(z))$. The general solution to this equation can be easily shown to be of the form

$$\mathcal{F}_2 = a(\tau + z) - G(z),$$

where

$$G(z) \equiv \int \hat{g}(z)\, dz$$

and a is an arbitrary function of $(\tau + z)$. Invoking (12) immediately gives

$$\mathcal{F}_2 = Pz^{-1}(\tau + z(q)) + G(\tau + z(q)) - G(z(q)), \quad (16)$$

where z^{-1} denotes the inverse function of $z(q)$. Consider, specifically, $f(q) = q^2$ and $g(q) = 0$, from which we obtain $z(q) = -q^{-1}$. Substitution into (16) gives

$$\mathcal{F}_2(q, P, \tau) = Pq(1 - \tau q)^{-1}.$$

This function $\mathcal{F}_2$ generates the connections

$$Q(q, p, \tau) = q(1 - \tau q)^{-1},$$
$$P(q, p, \tau) = p(1 - \tau q)^2. \quad (17)$$

It is easy to show that the latter formulas describe canonical transformations for all values of $\tau \neq q^{-1}$.

To check the validity of (17), we evaluate $T(\tau)$ directly. For $w(q, p) = q^2 p$ we have

$$T(\tau) = e^{\tau G}, \quad G \equiv q^2 \frac{\partial}{\partial q} - 2qp\frac{\partial}{\partial p}. \quad (18)$$

Under the change of variable $(q, p) \to (q, s)$ and $s \equiv qp^{\frac{1}{2}}$, we obtain

$$\left.\frac{\partial}{\partial q}\right|_p = \left.\frac{\partial}{\partial q}\right|_s + p^{\frac{1}{2}}\left.\frac{\partial}{\partial s}\right|_q, \quad \left.\frac{\partial}{\partial p}\right|_q = \frac{qp^{-\frac{1}{2}}}{2}\left.\frac{\partial}{\partial s}\right|_q.$$

Substitution into (18) gives

$$T(\tau) = \exp\left(-\tau \frac{\partial}{\partial(q^{-1})}\right),$$

which is easily recognized as the translation operator in the variable q^{-1}. Hence, we have

$$Q(q, s, \tau) = (q^{-1} - \tau)^{-1}, \quad P(q, s, \tau) = s^2(q^{-1} - \tau)^2.$$

Returning to the variables (q, p) gives (17) immediately.

The above heuristic example demonstrates the usefulness of (11) in obtaining (4) for a given observable $w(q, p)$. As mentioned previously, relations analogous to (11) hold for each $\mathcal{F}_k$; however, only those in terms of $\mathcal{F}_2$ or $\mathcal{F}_4$ are practical from a computational standpoint. Also, the general condition (10) allows solution in terms of either q or P for $\mathcal{F}_2$. Hence, one can solve the simpler of the two relations stated in (11).

ACKNOWLEDGMENT

The author would like to thank Professor G. Rosen for his kind encouragement and many stimulating discussions.

APPENDIX

In the context of the 1-parameter subgroup of canonical transformations $\mathcal{C}_w$, we state the following theorem.

Theorem: For the canonical transformation

$$\begin{pmatrix} Q \\ P \end{pmatrix} = T(\tau)\begin{pmatrix} q \\ p \end{pmatrix},$$

$$T(\tau) \equiv \exp\left[\tau\left(\frac{\partial w}{\partial p}\cdot\frac{\partial}{\partial q} - \frac{\partial w}{\partial q}\cdot\frac{\partial}{\partial p}\right)\right],$$

the arbitrary observable $a(q, p)$ transforms such that

$$A(q, p, \tau) \equiv T(\tau)a(q, p) = a(Q(q, p, \tau), P(q, p, \tau)). \tag{A1}$$

To prove this theorem, we first consider infinitesimal transformations. Then we have

$$A(q, p, \tau) = a(q, p) + \tau[a, w],$$
$$Q(q, p, \tau) = q + \tau[q, w],$$
$$P(q, p, \tau) = p + \tau[p, w].$$

We define $\phi(\lambda)$ such that

$$\phi(\lambda) \equiv a(q, p) + \lambda\tau[a, w] - a(q + \lambda\tau[q, w], p + \lambda\tau[p, w]),$$

and evaluate $\partial\phi/\partial\lambda$ to first order in τ:

$$\frac{\partial\phi}{\partial\lambda} = \tau[a, w] - \tau\left(\frac{\partial a}{\partial q}\bigg|_{\tau=0}\cdot\frac{\partial w}{\partial p} - \frac{\partial a}{\partial p}\bigg|_{\tau=0}\cdot\frac{\partial w}{\partial q}\right) = 0.$$

Hence, for infinitesimal transformations we have

$$\phi(0) = \phi(1) = 0,$$

which yields

$$a(q, p) + \tau[a, w] = a(q + \tau[q, w], p + \tau[p, w]).$$

Since $\mathcal{C}_w$ has the group property, this result can be iterated to yield (A1).

* Work partially supported by National Science Foundation Grant.

[1] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950), p. 240.
[2] G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969), p. 16.
[3] C. Lanczos, *Variational Principles of Mechanics* (U. of Toronto Press, Toronto, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Zeros of the Grand Partition Function for a Lattice Gas

OLE J. HEILMANN*†
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass. 02139

(Received 8 December 1969)

The following three statements about the zeros of the grand partition function of a lattice gas with negative (attractive) interactions are proved: (1) Not all the zeros will be on the unit circle in the high-temperature limit if forces of higher order than 2-body are included; (2) in the low-temperature limit they will, in general, lie on the unit circle; (3) it is possible to have the zeros dense in the complex plane. It is also shown that not all polynomials with positive coefficients and roots on the unit circle are a grand partition function of a lattice gas.

Since Lee and Yang[1] proposed that the roots of the grand partition function for a lattice gas lie on the unit circle for attractive 2-body forces, there has been much speculation about the possibility of a similar theorem for more general forces. Also, the converse theorem that every polynomial with positive coefficients and roots on the unit circle is the grand partition function for some lattice gas has been considered as a conjecture.

It will be shown that the possibility of conjectures of the first type is limited as follows: If higher-order interactions are included, then some of the roots will always lie off the unit circle for sufficiently high temperature if all forces are attractive. Furthermore, it is possible to devise interactions which distribute the roots all over the complex plane in such a manner that they become dense as $n \to \infty$. The second conjecture will be shown to be wrong by a counter-example.

In order to simplify the notation, the mathematically equivalent Ising-type model will be considered in place of the lattice gas. The most general Hamiltonian for n spins $\sigma_1, \sigma_2, \cdots, \sigma_n$, which are 1 or -1, then reads

$$\mathcal{H} = -\sum_{\mu\in\Delta_n} J_\mu \prod_{k\in\mu} \sigma_k - H\sum_{k=1}^{n}\sigma_k. \tag{1}$$

(The summation over μ runs over all subsets of the n spins, which contains at least two spins.) The first term represents interactions between the spins, while the second term is the interaction between the spins and a magnetic field H. We require

$$J_\mu \geq 0, \quad \text{for all } \mu,$$

in order that the system be ferromagnetic.

Theorem: For the canonical transformation

$$\begin{pmatrix} Q \\ P \end{pmatrix} = T(\tau)\begin{pmatrix} q \\ p \end{pmatrix},$$

$$T(\tau) \equiv \exp\left[\tau\left(\frac{\partial w}{\partial p}\cdot\frac{\partial}{\partial q} - \frac{\partial w}{\partial q}\cdot\frac{\partial}{\partial p}\right)\right],$$

the arbitrary observable $a(q, p)$ transforms such that

$$A(q, p, \tau) \equiv T(\tau)a(q, p) = a(Q(q, p, \tau), P(q, p, \tau)). \tag{A1}$$

To prove this theorem, we first consider infinitesimal transformations. Then we have

$$A(q, p, \tau) = a(q, p) + \tau[a, w],$$
$$Q(q, p, \tau) = q + \tau[q, w],$$
$$P(q, p, \tau) = p + \tau[p, w].$$

We define $\phi(\lambda)$ such that

$$\phi(\lambda) \equiv a(q, p) + \lambda\tau[a, w] - a(q + \lambda\tau[q, w], p + \lambda\tau[p, w]),$$

and evaluate $\partial\phi/\partial\lambda$ to first order in τ:

$$\frac{\partial\phi}{\partial\lambda} = \tau[a, w] - \tau\left(\frac{\partial a}{\partial q}\bigg|_{\tau=0}\cdot\frac{\partial w}{\partial p} - \frac{\partial a}{\partial p}\bigg|_{\tau=0}\cdot\frac{\partial w}{\partial q}\right) = 0.$$

Hence, for infinitesimal transformations we have

$$\phi(0) = \phi(1) = 0,$$

which yields

$$a(q, p) + \tau[a, w] = a(q + \tau[q, w], p + \tau[p, w]).$$

Since $\mathcal{C}_w$ has the group property, this result can be iterated to yield (A1).

* Work partially supported by National Science Foundation Grant.

[1] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1950), p. 240.
[2] G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969), p. 16.
[3] C. Lanczos, *Variational Principles of Mechanics* (U. of Toronto Press, Toronto, 1966).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Zeros of the Grand Partition Function for a Lattice Gas

OLE J. HEILMANN*†
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass. 02139

(Received 8 December 1969)

The following three statements about the zeros of the grand partition function of a lattice gas with negative (attractive) interactions are proved: (1) Not all the zeros will be on the unit circle in the high-temperature limit if forces of higher order than 2-body are included; (2) in the low-temperature limit they will, in general, lie on the unit circle; (3) it is possible to have the zeros dense in the complex plane. It is also shown that not all polynomials with positive coefficients and roots on the unit circle are a grand partition function of a lattice gas.

Since Lee and Yang[1] proposed that the roots of the grand partition function for a lattice gas lie on the unit circle for attractive 2-body forces, there has been much speculation about the possibility of a similar theorem for more general forces. Also, the converse theorem that every polynomial with positive coefficients and roots on the unit circle is the grand partition function for some lattice gas has been considered as a conjecture.

It will be shown that the possibility of conjectures of the first type is limited as follows: If higher-order interactions are included, then some of the roots will always lie off the unit circle for sufficiently high temperature if all forces are attractive. Furthermore, it is possible to devise interactions which distribute the roots all over the complex plane in such a manner that they become dense as $n \to \infty$. The second conjecture will be shown to be wrong by a counter-example.

In order to simplify the notation, the mathematically equivalent Ising-type model will be considered in place of the lattice gas. The most general Hamiltonian for n spins $\sigma_1, \sigma_2, \cdots, \sigma_n$, which are 1 or -1, then reads

$$\mathcal{H} = -\sum_{\mu\in\Delta_n} J_\mu \prod_{k\in\mu} \sigma_k - H\sum_{k=1}^{n} \sigma_k. \tag{1}$$

(The summation over μ runs over all subsets of the n spins, which contains at least two spins.) The first term represents interactions between the spins, while the second term is the interaction between the spins and a magnetic field H. We require

$$J_\mu \geq 0, \quad \text{for all } \mu,$$

in order that the system be ferromagnetic.

With

$$z = e^{\beta H}, \tag{2}$$

the partition function can be written as polynomial in z^2:

$$Z = A \cdot z^{-n} \sum_{j=0}^{n} a_j z^{2n-2j}, \tag{3}$$

$$A = \exp\left(\beta \sum_{\mu \in \Delta_n} J_\mu\right), \tag{4}$$

$$a_j = \sum_{\substack{\nu \in \Delta_n \\ \#(\nu)=j}} \exp\left(-2\beta \sum_{\substack{\mu \in \Delta_n \\ \#(\nu \cap \mu)\ \text{odd}}} J_\mu\right) \tag{5}$$

[$\#(\nu)$ is the number of spins in the set ν]. Since a necessary condition for $Z(z) = 0$ to have its roots on the unit circle is $a_j = a_{n-j}$ and since this will in general be fulfilled only if $J_\mu = 0$ for $\#(\mu)$ odd, the summation over μ in the following is restricted to $\#(\mu)$ even.

In order to prove the first statement, we expand a_j in powers of β:

$$a_j = \sum_{\substack{\nu \in \Delta_n \\ \#(\nu)=j}} \left(1 - 2\beta \sum_{\substack{\mu \in \Delta_n \\ \#(\mu)\ \text{even} \\ \#(\mu \cap \nu)\ \text{odd}}} J_\mu + O(\beta^2)\right)$$

$$= \binom{n}{j} - 2\beta \sum_{l=1}^{[\frac{1}{2}n]} \sigma_{2l} \sum_{k\ \text{odd}} \binom{2l}{k}\binom{n-2l}{j-k} + O(\beta^2), \tag{6}$$

$$\sigma_{2l} = \sum_{\substack{\mu \in \Delta_n \\ \#(\mu)=2l}} J_\mu. \tag{7}$$

With

$$x = \tfrac{1}{2}(z + z^{-1}), \tag{8}$$

one has the following formulas:

$$[\tfrac{1}{2}(z - z^{-1})]^2 = x^2 - 1, \tag{9}$$

$$\sum_{j=0}^{n} \binom{n}{j} z^{n-2j} = 2^n x^n, \tag{10}$$

$$\sum_{\substack{k\ \text{odd} \\ 2l \geq k \geq 0}} \binom{2l}{k} z^{2l-2k} = 2^{2l-1}[x^{2l} - (x^2 - 1)^l]$$

$$= -2^{2l-1} \sum_{k=1}^{l} \binom{l}{k} (-1)^k x^{2l-2k}. \tag{11}$$

Using these formulas, we can write Z as a polynomial in x:

$$Z = A 2^n \left(x^n + \beta \sum_{k=1}^{[\frac{1}{2}n]} (-1)^k x^{n-2k} \sum_{l=k}^{[\frac{1}{2}n]} \sigma_{2l} \binom{l}{k} + O(\beta^2)\right). \tag{12}$$

Since z being on the unit circle implies that x is real and since $-1 \leq x \leq 1$, it will be sufficient to show that $Z(x) = 0$ does not have all real roots for β small enough. Since Z (aside from a factor x if n is odd) can be considered as polynomial in x^2, Newton's inequality[2] can be applied to the first three coefficients in this polynomial, giving the following necessary conditions for all the roots of $Z(x)$ to be real:

$$1 \cdot ([\tfrac{1}{2}n] - 1) b_1^2 - [\tfrac{1}{2}n] \cdot 2 \cdot b_0 \cdot b_2 \geq 0, \tag{13}$$

$$b_0 = 1 + O(\beta^2),$$
$$b_k = (-1)^k \beta \sum_{l=k}^{[\frac{1}{2}n]} \sigma_{2l} \binom{l}{k} + O(\beta^2), \quad k > 0. \tag{14}$$

Since the first term in (13) is of the order β^2, while second is of the order β, if $\sigma_{2l} > 0$ for some $l > 1$, the inequality in (13) will always fail for sufficiently small β, if we have that $J_\mu \geq 0$ for all μ and $J_\mu > 0$ for some μ such that $\#(\mu) > 2$. This concludes the proof of the following theorem:

Theorem 1: In the Ising model with nonnegative interactions of even order, if at least one interaction of higher order than two is different from zero, then for sufficiently high temperature the zeros of the partition function will not all lie on the unit circle.

For small T, however, one will generally have all the roots lying on the unit circle, if $J_\mu = 0$ for $\#(\mu)$ odd. In order to prove this, notice that, if in (5) the summation in the exponential contains at least one $J_\mu > 0$ for each set ν with $\#(\nu) = j$, then the corresponding a_j can be made arbitrarily small by making β sufficiently large. This condition can be stated as follows:

Condition 1: A subset ν of n spins is said to fulfill this condition if there exists a subset μ such that $\#(\mu \cap \nu)$ is odd and $J_\mu > 0$.

Suppose every ν, for which $1 \leq \#(\nu) \leq n - 1$, fulfills Condition 1; then for sufficiently small temperature

$$\sum_{j=1}^{n-1} a_j < 2. \tag{15}$$

Again, writing Z as polynomial in x, we have

$$Z = \sum_{j=0}^{n} a_j T_{|n-2j|}(x), \tag{16}$$

where $T_k(x)$ is the kth-order Tschebycheff polynomial

$$T_k(x) = \cos[k \arccos(x)]. \tag{17}$$

Since (15) implies

$$\left|\sum_{j=1}^{n-1} a_j T_{|n-2j|}(x)\right| < 2, \quad \text{for} \quad -1 \leq x \leq 1, \tag{18}$$

one has, for sgn Z,

$$\operatorname{sgn}[Z(x_l)] = (-1)^l, \quad (19)$$

for

$$x_l = \cos(l\pi/n), \quad l = 0, 1, \cdots, n. \quad (20)$$

This proves that $Z(z) = 0$ has $2n$ roots on the unit circle. Since the same will hold if the system of n spins is composed of several noninteracting systems of spins, each of which has its zeros on the unit circle, a set ν may alternatively fulfill the following condition:

Condition 2: A subset ν of n spins is said to fulfill this condition if $J_\mu = 0$ for all subsets μ of the n spins, which are neither contained completely in ν nor in the complement of ν with respect to the n spins.

The theorem can then be stated as:

Theorem 2: In the Ising model for n spins with nonnegative interactions of even order, if every subset of the n spins fulfills either Condition 1 or Condition 2, then for sufficiently low temperature all the zeros of the partition function will lie on the unit circle.

To prove that roots can lie everywhere in complex plane, we consider first a system of $2n$ spins with only $2n$th-order interaction $J^{(2n)}$ and define

$$\gamma = \exp[-\beta J^{(2n)}]. \quad (21)$$

The partition function for this case is

$$Z = Az^{-2n}[\tfrac{1}{2}(1+\gamma)(z^2-1)^{2n} + \tfrac{1}{2}(1-\gamma)(z^2-1)^{2n}], \quad (22)$$

with the roots

$$z_p^2 = (\alpha_p + 1)/(\alpha_p - 1),$$
$$\alpha_p = \exp(\delta/2n)\exp[i(1+2p)/2n], \quad (23)$$
$$\delta = \log[(1-\gamma)/(1+\gamma)], \quad p = 0, 1, \cdots, 2n-1.$$

If the exponentials are evaluated in a Taylor series, one finds that the root of maximum modulus becomes, for fixed γ and large n,

$$z_0^2 \simeq 4n/(\delta + i\pi). \quad (24)$$

Next we consider the case of $2n$ groups of m spins each with very strong 2-body interactions connecting all the spins within each group. The only other interaction shall be $2n$-order interactions between one spin from each group. In the limit where the 2-body forces become infinite, the partition functions become identical to (22) with z^m in place of z and $J^{(2n)}$ being the sum of the $2n$-body interactions. By taking $n = m^m$ and then letting $m \to \infty$, it is easily seen that one has the following theorem:

Theorem 3: In the Ising model with nonnegative interactions of even order in the limit of an infinite number of spins, the zeros of the partition function may become dense in the complex plane.

Finally, for a counterexample to the second conjecture, consider

$$(y+1)^3(y^2 - \tfrac{4}{3}y + 1) = y^5 + \tfrac{5}{3}y^4 + \tfrac{5}{3}y + 1. \quad (25)$$

This is clearly a polynomial with roots on the unit circle and nonnegative coefficients

$$a_1 = \tfrac{5}{3}, \quad a_2 = 0.$$

In order to make $a_2 = 0$ consistent with (5), some of the J_μ have to be ∞. Since $n = 5$, only 2-body and 4-body interactions have to be considered. Making one of the 4-body interactions ∞ is not satisfactory since this implies $a_1 \leq 1$. Next, trying only with 2-body interactions, we see that $J_{1,2} = J_{1,3} = J_{1,4} = \infty$ is a choice which makes a_1 as large as possible. In this case too, however, one has $a_1 < 1$, which concludes the proof of the following theorem:

Theorem 4: Not all polynomials with positive coefficients and zeros on the unit circle are the partition function for some Ising model with nonpositive (ferromagnetic) interactions.

ACKNOWLEDGMENT

The author is very indebted to Professor E. Lieb for important suggestions and encouragement.

* Work partially supported by National Science Foundation Grant GP-9414 and by Statens Naturvidenskabelige Forskningsraad, J. Nr. 511-208/69.

† On leave from University of Copenhagen. Permanent address: Chemistry Laboratory III, The H.C. Ørsted Institute, Copenhagen, Denmark.

[1] T. D. Lee and C. N. Yang, Phys. Rev. **87**, 410 (1952).

[2] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge U.P., Cambridge, 1959), Theorem 144.

Exactly Solvable Electrodynamic Two-Body Problem

R. A. RUDD* AND ROBERT NYDEN HILL
Department of Physics, University of Delaware, Newark, Delaware

(Received 12 March 1970)

The equations of motion of the physical 2-body problems of relativistic electrodynamics are differential–difference equations. However, in the unphysical 2-body problem in which one particle responds only to retarded fields and the other only to advanced fields, the equations of motion are differential equations. These differential equations are solved in the center-of-momentum frame by the elementary method of finding integrating factors. Solutions in an arbitrary Lorentz frame are found by a method which parametrizes the world lines on the particle velocities. All of the constants of the motion contain interaction contributions; this appears to be a characteristic feature of relativistic particle dynamics.

I. INTRODUCTION

Existing intuition with respect to the interactions of charged particles is rooted primarily in the exactly solvable problems of a single particle in an external field (the nonrelativistic 2-particle problem is reducible to this form). The only previously known exact solutions of a 2-body problem of relativistic electrodynamics are the circular-orbit solutions found by Schild[1] for the case of half-retarded plus half-advanced interactions. Other known exact results take the form of existence and uniqueness theorems[2] and include a demonstration by Driver that, for sufficiently large values of the product of initial relative velocity and initial separation, instantaneous specification of positions and velocities guarantees existence and uniqueness for all time of the solution to the problem of two charges interacting via retarded fields (without radiation damping) where the motion is confined to one dimension as a consequence of initial conditions.[3] Approximate methods which have been used include expansion of the equations of motion in powers of c^{-1} about the nonrelativistic limit, expansion of the solutions in powers of the mass ratio of the two charges,[4] and expansion of the equations of motion in powers of $e^2/(mc^2r)$ (where r is the instantaneous interparticle separation in some Lorentz frame) about the free-particle limit.[5]

It is the differential–difference structure of the equations of motion which makes most electrodynamic 2-body problems hard to handle. We consider here the one electrodynamic 2-body problem which is without this difficulty—the problem in which one particle responds only to retarded fields and the other particle responds only to advanced fields. Despite the unphysical nature of this problem, the results shed light on certain aspects of relativistic particle dynamics, such as the presence of interaction contributions to all of the constants of the motion and the presence of an asymptotically surviving interaction contribution to the constant associated with Lorentz invariance.[6] We restrict ourselves to the case in which the particle motion is confined to one dimension as a consequence of initial conditions. Section II is devoted to the integration of the equations of motion in the center-of-momentum frame. The resulting world lines are given by Eqs. (11)–(15).

The possibility of an instantaneous interaction description of particle motions in relativistic dynamics has been explored by Currie and by Hill.[5] One result of this exploration is a convenient Lorentz-invariant parametric representation of the world lines of a 1-dimensional 2-particle problem.[7] This representation permits exploration of a Hamiltonian formulation of the dynamics, which is necessarily an instantaneous-interaction description, because Hamilton's equations are differential equations. Section III obtains this parametric representation for the dynamics of Sec. II; the general results are given by Eqs. (16), (30)–(32), (34), (35), (37), and (41), with the explicit parametric representation of the world lines in the center-of-momentum frame given by Eqs. (45) and (52). These results conform to the general framework of Ref. 7.

II. AN EXACTLY SOLVABLE ELECTRODYNAMIC 2-BODY PROBLEM

We consider two-point particles of masses m_1 and m_2 and charges e_1 and e_2 in three dimensions whose motion is confined to one dimension as a consequence of the initial data. In this case, the magnetic and radiation fields due to one particle vanish at the location of the other. We consider only the case $e_1e_2 > 0$, set $c = 1$, and assume that always $x_1 > x_2$. The Lienard–Wiechert expressions[8] for the fields of a point charge yield, for the retarded and advanced fields $(E_1)_{\text{ret}}$ and

$(E_1)_{\rm adv}$ at particle 1 due to particle 2,

$$(E_1)_{\rm ret} = \frac{e_2}{[x_1(t_1) - x_2(t_{2r}')]^2}\left(\frac{1 + v_2(t_{2r})}{1 - v_2(t_{2r})}\right), \quad (1a)$$

$$(E_1)_{\rm adv} = \frac{e_2}{[x_1(t_1) - x_2(t_{2a})]^2}\left(\frac{1 - v_2(t_{2a})}{1 + v_2(t_{2a})}\right), \quad (1b)$$

where

$$t_{2r} = t_1 - [x_1(t_1) - x_2(t_{2r})], \quad (1c)$$

$$t_{2a} = t_1 + [x_1(t_1) - x_2(t_{2a})]. \quad (1d)$$

The fields $(E_2)_{\rm ret}$ and $(E_2)_{\rm adv}$, at particle 2 due to particle 1, are

$$(E_2)_{\rm ret} = -\frac{e_1}{[x_1(t_{1r}) - x_2(t_2)]^2}\left(\frac{1 - v_1(t_{1r})}{1 + v_1(t_{1r})}\right), \quad (2a)$$

$$(E_2)_{\rm adv} = -\frac{e_1}{[x_1(t_{1a}) - x_2(t_2)]^2}\left(\frac{1 + v_1(t_{1a})}{1 - v_1(t_{1a})}\right), \quad (2b)$$

where

$$t_{1r} = t_2 - [x_1(t_{1r}) - x_2(t_2)], \quad (2c)$$

$$t_{1a} = t_2 + [x_1(t_{1a}) - x_2(t_2)]. \quad (2d)$$

We distinguish two exactly solvable cases. In case (A), particle 1 responds only to advanced fields and particle 2 only to retarded fields. In case (B), which is case (A) time reversed, particle 1 responds only to retarded interactions and particle 2 responds only to advanced interactions.

By inserting the appropriate fields from Eqs. (1) and (2) in the relativistic equations of motion

$$\frac{d}{dt_i}\left(\frac{m_i v_i(t_i)}{[1 - v_i^2(t_i)]^{\frac{1}{2}}}\right) = e_i E_i, \quad (3)$$

we obtain

$$\frac{d}{dt_1}\left(\frac{m_1 v_1(t_1)}{[1 - v_1^2(t_1)]^{\frac{1}{2}}}\right) = \frac{e_1 e_2}{[x_1(t_1) - x_2(t_2)]^2}\left(\frac{1 + \theta v_2(t_2)}{1 - \theta v_2(t_2)}\right), \quad (4a)$$

$$\frac{d}{dt_2}\left(\frac{m_2 v_2(t_2)}{[1 - v_2^2(t_2)]^{\frac{1}{2}}}\right) = -\frac{e_1 e_2}{[x_1(t_1) - x_2(t_2)]^2}\left(\frac{1 + \theta v_1(t_1)}{1 - \theta v_1(t_1)}\right), \quad (4b)$$

with the light-cone condition

$$t_1 - t_2 = \theta[x_1(t_1) - x_2(t_2)]. \quad (4c)$$

In case (A), $\theta \equiv -1$; in case (B), $\theta \equiv +1$.

The light-cone condition (4c) implies that differentiation with respect to t_1 is related to differentiation with respect to t_2 by

$$\frac{dt_2}{dt_1} = \frac{1 - \theta v_1(t_1)}{1 - \theta v_2(t_2)}. \quad (5)$$

Conserved quantities H, P, and K can be found by the elementary method of looking for integrating factors, if (5) is used to rewrite the equations in terms of a single time. The conserved quantities, which we identify respectively as energy, momentum, and the constant of the motion associated with Lorentz invariance, are

$$H = m_1(1 - v_1^2)^{-\frac{1}{2}} + m_2(1 - v_2^2)^{-\frac{1}{2}} + e_1 e_2 (x_1 - x_2)^{-1}, \quad (6)$$

$$P = m_1 v_1 (1 - v_1^2)^{-\frac{1}{2}} + m_2 v_2 (1 - v_2^2)^{-\frac{1}{2}} - \theta e_1 e_2 (x_1 - x_2)^{-1}, \quad (7)$$

and

$$K = m_1(x_1 - v_1 t_1)(1 - v_1^2)^{-\frac{1}{2}} + m_2(x_2 - v_2 t_2)(1 - v_2^2)^{-\frac{1}{2}} + \tfrac{1}{2} e_1 e_2 (x_1 - x_2)^{-1}[x_1 + x_2 + \theta(t_1 + t_2)]. \quad (8)$$

By Lorentz-transforming x_i, t_i, and v_i in the conventional manner, one can show that H and P form a 2-vector while K is an invariant (or, equivalently, the one nonvanishing component of an antisymmetric 2-tensor). Algebraic elimination among Eqs. (6)–(8) in the center-of-momentum frame $P = 0$ yields

$$m_1(1 - v_1^2)^{-\frac{1}{2}} + e_1 e_2 (m_2/E)^2 [x_1 - (K/E) - (e_1 e_2/E)]^{-1} = \tfrac{1}{2}[E + (m_1^2/E) - (m_2^2/E)] \quad (9)$$

and

$$m_2(1 - v_2^2)^{-\frac{1}{2}} + e_1 e_2 (m_1/E)^2 [-x_2 + (K/E) - (e_1 e_2/E)]^{-1} = \tfrac{1}{2}[E - (m_1^2/E) + (m_2^2/E)], \quad (10)$$

where $E = (H^2 - P^2)^{\frac{1}{2}}$ is the invariant center-of-momentum frame energy. Equations (9) and (10) have the form of energy conservation laws for a particle in a fixed Coulomb field with the unusual feature that the strength and location of the fixed Coulomb field depend on E.

Equations (9) and (10) can be readily integrated to yield the world lines. The result can be written concisely by introducing the parameters

$$\mu \equiv E^4 + m_1^4 + m_2^4 - 2E^2 m_1^2 - 2E^2 m_2^2 - 2m_1^2 m_2^2,$$
$$b_1 \equiv 4\mu^{-1} m_1 m_2^2 e_1 e_2, \quad (11)$$
$$b_2 \equiv 4\mu^{-1} m_1^2 m_2 e_1 e_2,$$

and the new variables

$$y_1 \equiv x_1 - K/E - (e_1 e_2/\mu E)[(E^2 - m_1^2)^2 - m_2^4],$$
$$y_2 \equiv x_2 - K/E + (e_1 e_2/\mu E)[(E^2 - m_2^2)^2 - m_1^4]. \quad (12)$$

The turning points are at $y_1 = b_1$ and $y_2 = -b_2$; the solutions are

$$t_1 - t_{10} = \theta_1\{\mu^{-\frac{1}{2}}(E^2 + m_1^2 - m_2^2)(y_1^2 - b_1^2)^{\frac{1}{2}} + 8\mu^{-\frac{3}{2}} e_1 e_2 m_1^2 m_2^2 E \tanh^{-1}[y_1^{-1}(y_1^2 - b_1^2)^{\frac{1}{2}}]\}, \quad (13)$$

$$t_2 - t_{20} = \theta_2\{\mu^{-\frac{1}{2}}(E^2 - m_1^2 + m_2^2)(y_2^2 - b_2^2)^{\frac{1}{2}} - 8\mu^{-\frac{3}{2}} e_1 e_2 m_1^2 m_2^2 E \tanh^{-1}[y_2^{-1}(y_2^2 - b_2^2)^{\frac{1}{2}}]\}, \quad (14)$$

where $\theta_i = -1$ for $t_i < t_{i0}$ and $\theta_i = +1$ for $t_i > t_{i0}$. The integration constant t_{i0} is the turning time of the ith particle.

The set of integration constants H, P, K, t_{10}, t_{20} cannot be independent since we began with a pair of second-order ordinary differential equations whose solution should involve only four constants. The time-translation invariance implies that the sum of turning times $t_{10} + t_{20}$ is independent of H, P, and K; hence, the difference $t_{10} - t_{20}$ must be dependent on H and P. The calculation of the Appendix shows that

$$t_{10} - t_{20} = \theta\{2e_1e_2E\mu^{-1}(E^2 - m_1^2 - m_2^2) - 8\mu^{-\frac{3}{2}}e_1e_2m_1^2m_2^2E \tanh^{-1}[\mu^{\frac{1}{2}}(E^2 - m_1^2 - m_2^2)^{-1}]\}. \quad (15)$$

III. PARAMETRIC FORM OF THE WORLD LINES

Differential conditions which guarantee the Lorentz invariance of instantaneous action-at-a-distance relativistic dynamics have been given by Currie and by Hill.[5] The general solution of these conditions for the special case of two particles in one dimension has been obtained by Hill, and the resulting equations of motion integrated to obtain a convenient parametric form for the world lines.[7] In this section, we consider the specialization of these general results to the 2-body problems of Sec. II.

We begin by summarizing the relevant general results from Ref. 7. Two functions $f(\xi, \zeta)$ and $g(\eta, \zeta)$ define a Lorentz-invariant four-constant family of pairs of world lines

$$x_1 = \frac{1}{2}\left[c_1 + c_2 + \Phi^{\frac{1}{2}}\frac{\partial f}{\partial \xi} + \Phi^{-\frac{1}{2}}\left(f - \xi\frac{\partial f}{\partial \xi}\right)\right], \quad (16a)$$

$$t_1 = \frac{1}{2}\left[c_1 - c_2 + \Phi^{\frac{1}{2}}\frac{\partial f}{\partial \xi} - \Phi^{-\frac{1}{2}}\left(f - \xi\frac{\partial f}{\partial \xi}\right)\right], \quad (16b)$$

$$x_2 = \frac{1}{2}\left[c_1 + c_2 - \Phi^{\frac{1}{2}}\frac{\partial g}{\partial \eta} - \Phi^{-\frac{1}{2}}\left(g - \eta\frac{\partial g}{\partial \eta}\right)\right], \quad (16c)$$

$$t_2 = \frac{1}{2}\left[c_1 - c_2 - \Phi^{\frac{1}{2}}\frac{\partial g}{\partial \eta} + \Phi^{-\frac{1}{2}}\left(g - \eta\frac{\partial g}{\partial \eta}\right)\right]. \quad (16d)$$

Here, ξ and η are Lorentz-invariant parameters for the two world lines; c_1, c_2, Φ, and ζ are the four constants of the motion. The Lorentz invariance can be easily verified:

Under the transformation $t_i \to t_i' = t_i + t_0$ (time translation),

$$c_1 \to c_1' = c_1 + t_0,$$
$$c_2 \to c_2' = c_2 - t_0,$$

and Φ and ζ are left unchanged. Under the transformation $x_i \to x_i' = x_i + x_0$ (space translation),

$$c_i \to c_i' = c_i + x_0,$$

and Φ and ζ are left unchanged. Under the transformation

$$x_i \to x_i' = (x_i - vt_i)(1 - v^2)^{-\frac{1}{2}}$$

and

$$t_i \to t_i' = (t_i - vx_i)(1 - v^2)^{-\frac{1}{2}}$$

(pure Lorentz transformation),

$$c_1 \to c_1' = c_1(1 - v)^{\frac{1}{2}}(1 + v)^{-\frac{1}{2}},$$
$$c_2 \to c_2' = c_2(1 - v)^{-\frac{1}{2}}(1 + v)^{\frac{1}{2}},$$
$$\Phi \to \Phi' = \Phi(1 - v)(1 + v)^{-1},$$

and ζ is left unchanged.

The velocities of the particles are given by

$$v_1 = (\Phi - \xi)/(\Phi + \xi), \quad (17a)$$
$$v_2 = (\Phi - \eta)/(\Phi + \eta). \quad (17b)$$

Because Φ is a constant of the motion, it is clear from (17) that parametrizing the world lines on ξ and η is essentially a Lorentz-invariant way of parametrizing the world lines on the particle velocities. The choice of a numerical value for Φ picks out an inertial frame. The fields E_i at the locations of the particles, obtained from $E_i = m_ia_ie_i^{-1}(1 - v_i^2)^{-\frac{3}{2}}$, are

$$E_1 = -\frac{1}{2}\frac{m_1}{e_1}\xi^{-\frac{3}{2}}\left(\frac{\partial^2 f}{\partial \xi^2}\right)^{-1} \quad (18a)$$

and

$$E_2 = \frac{1}{2}\frac{m_2}{e_2}\eta^{-\frac{3}{2}}\left(\frac{\partial^2 g}{\partial \eta^2}\right)^{-1}. \quad (18b)$$

We define functions $\alpha(\xi, \eta, \zeta)$ and $\beta(\xi, \eta, \zeta)$ by

$$\alpha \equiv f - \xi\frac{\partial f}{\partial \xi} + g - \eta\frac{\partial g}{\partial \eta}, \quad (19a)$$

$$\beta \equiv \frac{\partial f}{\partial \xi} + \frac{\partial g}{\partial \eta}. \quad (19b)$$

Then it is easily shown by equating (16b) and (16d) that the relation between ξ and η for which the particles' world points are simultaneous in the frame specified by Φ is given by

$$\alpha = \Phi\beta. \quad (20)$$

It can also be shown from (16) and (19) that the light-cone condition $t_1 - t_2 = x_1 - x_2$ is equivalent to

$$\alpha(\xi, \eta, \zeta) = 0, \quad (21)$$

while $t_1 - t_2 = -(x_1 - x_2)$ is equivalent to

$$\beta(\xi, \eta, \zeta) = 0. \quad (22)$$

Equations (21) and (22) can also be obtained from (20) by letting Φ approach 0 or ∞ to get the limiting inertial frames in which the world points connected by one or the other light-cone conditions are simultaneous.

The fields given by Eqs. (1) and (2) can be rewritten in parametric variables by the use of (16), (17), (19), (21), and (22). The result is

$$(E_1)_{\text{ret}} = 4e_2[\eta_{\text{ret}}(\xi, \zeta)]^{-1}\{\beta[\xi, \eta_{\text{ret}}(\xi, \zeta), \zeta]\}^{-2}, \quad (23\text{a})$$

$$(E_1)_{\text{adv}} = 4e_2\eta_{\text{adv}}(\xi, \zeta)\{\alpha[\xi, \eta_{\text{adv}}(\xi, \zeta), \zeta]\}^{-2}, \quad (23\text{b})$$

$$(E_2)_{\text{ret}} = -4e_1\xi_{\text{ret}}(\eta, \zeta)\{\alpha[\xi_{\text{ret}}(\eta, \zeta), \eta, \zeta]\}^{-2}, \quad (24\text{a})$$

$$(E_2)_{\text{adv}} = -4e_1[\xi_{\text{adv}}(\eta, \zeta)]^{-1}\{\beta[\xi_{\text{adv}}(\eta, \zeta), \eta, \zeta]\}^{-2}, \quad (24\text{b})$$

where $\eta_{\text{ret}}(\xi, \zeta)$ and $\xi_{\text{adv}}(\eta, \zeta)$ are obtained by solving $\alpha = 0$ for η and for ξ, respectively, while $\eta_{\text{adv}}(\xi, \zeta)$ and $\xi_{\text{ret}}(\eta, \zeta)$ are obtained by solving $\beta = 0$ for η and for ξ. Equations which specify $f(\xi, \zeta)$ and $g(\eta, \zeta)$ for various electrodynamic 2-body problems now follow from equating (18a) to an appropriate linear combination of (23a) and (23b) and (18b) to a linear combination of (24a) and (24b). The resulting equations are again differential–difference equations, except in the two cases solved in Sec. II.

We consider case (A) $[E_1 = (E_1)_{\text{adv}}, E_2 = (E_2)_{\text{ret}}]$ first. The equations to be solved are

$$\frac{\partial^2 f_A}{\partial \xi^2} = -\tfrac{1}{8}m_1(e_1e_2)^{-1}\xi^{-\frac{3}{2}}\eta^{-1}\alpha^2, \quad (25\text{a})$$

$$\frac{\partial^2 g_A}{\partial \eta^2} = -\tfrac{1}{8}m_2(e_1e_2)^{-1}\xi^{-1}\eta^{-\frac{3}{2}}\alpha^2, \quad (25\text{b})$$

with η and ξ related by $\beta = 0$. The relation between ξ and η can be made explicit by differentiating $\beta = 0$ and using (25) to obtain

$$\frac{d\eta}{d\xi} = -\left(\frac{\partial^2 f_A}{\partial \xi^2}\right)\left(\frac{\partial^2 g_A}{\partial \eta^2}\right)^{-1} = -\frac{m_1}{m_2}\,\xi^{-\frac{1}{2}}\eta^{\frac{1}{2}}. \quad (26)$$

Integration yields

$$m_1\xi^{\frac{1}{2}} + m_2\eta^{\frac{1}{2}} = \Lambda^{-1}(\zeta), \quad (27)$$

where the integration constant Λ is an arbitrary function of ζ. Next, we obtain α. By differentiating (19a) and using (25), it follows that

$$d\alpha = \tfrac{1}{8}\alpha^2(e_1e_2)^{-1}(m_1\xi^{-\frac{3}{2}}\eta^{-1}\,d\xi + m_2\xi^{-1}\eta^{-\frac{3}{2}}\,d\eta). \quad (28)$$

Equation (28) can be integrated with the aid of (26) and (27) to obtain

$$\alpha = -4e_1e_2[m_1\xi^{-\frac{1}{2}} + m_2\eta^{-\frac{1}{2}} - c(\zeta)\Lambda(\zeta)]^{-1}, \quad (29)$$

where the integration constant c depends arbitrarily on ζ. Equations (27) and (29) can be used to express the right-hand side of Eq. (25a) for $\partial^2 f/\partial\xi^2$ completely in terms of ξ. Integrating twice with respect to ξ then yields

$$f_A(\xi, \zeta) = 4e_1e_2\Lambda^2[2m_1m_2^2\nu^{-1}\xi^{\frac{1}{2}} + \tfrac{1}{2}\nu^{-1}\Lambda^{-1}(c - m_1^2 - m_2^2) + 2m_1^2m_2^2\nu^{-\frac{3}{2}}(1 - c\Lambda^2\xi)\ln|\Gamma_f(\xi)|] + \xi f_1 + f_2, \quad (30)$$

where

$$\Gamma_f(\xi) \equiv \frac{2m_1c\Lambda^2\xi^{\frac{1}{2}} - (c + m_1^2 - m_2^2)\Lambda + \nu^{\frac{1}{2}}}{2m_1c\Lambda^2\xi^{\frac{1}{2}} - (c + m_1^2 - m_2^2)\Lambda - \nu^{\frac{1}{2}}} \quad (31)$$

and

$$\nu \equiv \Lambda^2(m_1^4 + m_2^4 + c^2 - 2m_1^2m_2^2 - 2m_1^2c - 2m_2^2c). \quad (32)$$

The integration constants f_1 and f_2 are arbitrary functions of ζ. From Eqs. (19) and $\beta = 0$, it follows that

$$g_A = \alpha - f_A + (\xi - \eta)\partial f_A/\partial\xi. \quad (33)$$

Equation (27) can be used to express the right-hand side of (33) completely in terms of η. The result is

$$g_A(\eta, \zeta) = 4e_1e_2\Lambda^2[2m_1^2m_2\nu^{-1}\eta^{\frac{1}{2}} + \tfrac{1}{2}\nu^{-1}\Lambda^{-1}(c - m_1^2 - m_2^2) + 2m_1^2m_2^2\nu^{-\frac{3}{2}}(1 - c\Lambda^2\eta)\ln|\Gamma_g(\eta)|] - \eta f_1 - f_2, \quad (34)$$

where

$$\Gamma_g(\eta) \equiv \frac{2m_2c\Lambda^2\eta^{\frac{1}{2}} - (c - m_1^2 + m_2^2)\Lambda + \nu^{\frac{1}{2}}}{2m_2c\Lambda^2\eta^{\frac{1}{2}} - (c - m_1^2 + m_2^2)\Lambda - \nu^{\frac{1}{2}}}. \quad (35)$$

Next, we consider case (B) $[E_1 = (E_1)_{\text{ret}}, E_2 = (E_2)_{\text{adv}}]$. The equations to be solved are now

$$\frac{\partial^2 f_B}{\partial \xi^2} = -\tfrac{1}{8}m_1(e_1e_2)^{-1}\xi^{-\frac{3}{2}}\eta\beta^2, \quad (36\text{a})$$

$$\frac{\partial^2 g_B}{\partial \eta^2} = -\tfrac{1}{8}m_2(e_1e_2)^{-1}\xi\eta^{-\frac{3}{2}}\beta^2. \quad (36\text{b})$$

If we introduce new dependent and independent variables according to $\xi' = \xi^{-1}$, $\eta' = \eta^{-1}$, $f' = \xi^{-1}f$, and $g' = \eta^{-1}g$, we find that $\alpha' = \beta$, $\beta' = \alpha$, and that cases (A) and (B) are interchanged. Hence, the solutions of (36) can be found from those already obtained for (25); they are

$$f_B(\xi, \zeta) = \xi f_A(\xi^{-1}, \zeta), \quad (37\text{a})$$

$$g_B(\eta, \zeta) = \eta g_A(\eta^{-1}, \zeta). \quad (37\text{b})$$

We see from (39) that f and $\partial f/\partial\xi$ are infinite when either $\Gamma_f(\xi) = 0$ or $\Gamma_f(\xi) = \infty$. It is clear from the parametric form of the world lines given in (16) that these infinities of f and $\partial f/\partial\xi$ must correspond to $x_1 = \infty$, $t_1 = \mp\infty$ before and after the collision.[9]

We use a subscript b to refer to $t = -\infty$ and a subscript a to refer to $t = +\infty$. In order to distinguish ξ_a and ξ_b, we note that with $\Phi > 0$, $v_{1b} < v_{1a}$, when combined with (17a) yields $\xi_b > \xi_a$. Hence

$$\xi_{bA}^{\frac{1}{2}} = \xi_{aB}^{-\frac{1}{2}} = (2m_1c\Lambda^2)^{-1}[(c + m_1^2 - m_2^2)\Lambda + \nu^{\frac{1}{2}}], \tag{38a}$$

$$\xi_{aA}^{\frac{1}{2}} = \xi_{bB}^{-\frac{1}{2}} = (2m_1c\Lambda^2)^{-1}[(c + m_1^2 - m_2^2)\Lambda - \nu^{\frac{1}{2}}]. \tag{38b}$$

Similarly, $v_{2b} > v_{2a}$ implies $\eta_b < \eta_a$; then

$$\eta_{bA}^{\frac{1}{2}} = \eta_{aB}^{-\frac{1}{2}} = (2m_2c\Lambda^2)^{-1}[(c - m_1^2 + m_2^2)\Lambda - \nu^{\frac{1}{2}}], \tag{39a}$$

$$\eta_{aA}^{\frac{1}{2}} = \eta_{bB}^{-\frac{1}{2}} = (2m_2c\Lambda^2)^{-1}[(c - m_1^2 + m_2^2)\Lambda + \nu^{\frac{1}{2}}]. \tag{39b}$$

Asymptotically as $t \to \pm\infty$, H and P approach their free-particle values

$$H_{\text{free}} = \sum_{i=1}^{2} m_i(1 - v_i^2)^{-\frac{1}{2}},$$

$$P_{\text{free}} = \sum_{i=1}^{2} m_i v_i(1 - v_i^2)^{-\frac{1}{2}}.$$

The invariant center-of-momentum-frame energy also approaches its free-particle value

$$E_{\text{free}} = (H_{\text{free}}^2 - P_{\text{free}}^2)^{\frac{1}{2}};$$

and, by using (17), we then find that

$$\begin{aligned} E^2 &= m_1^2 + m_2^2 + m_1m_2(\xi_b^{\frac{1}{2}}\eta_b^{-\frac{1}{2}} + \xi_b^{-\frac{1}{2}}\eta_b^{\frac{1}{2}}) \\ &= m_1^2 + m_2^2 + m_1m_2(\xi_a^{\frac{1}{2}}\eta_a^{-\frac{1}{2}} + \xi_a^{-\frac{1}{2}}\eta_a^{\frac{1}{2}}). \end{aligned} \tag{40}$$

By using (38) and (39) in (40), we discover that

$$c = E^2. \tag{41}$$

Equations (11), (32), and (41) then imply that

$$\nu = \mu\Lambda^2. \tag{42}$$

The other three functions of integration in f and g [$\Lambda(\zeta)$, $f_1(\zeta)$, and $f_2(\zeta)$] cannot be identified asymptotically and are a manifestation of the redundancy in the parametric representation.[10]

An asymptotic evaluation also shows that

$$\begin{aligned} P &= -\tfrac{1}{2}(m_1\xi_b^{\frac{1}{2}} + m_2\eta_b^{\frac{1}{2}})\Phi^{-\frac{1}{2}} + \tfrac{1}{2}(m_1\xi_b^{-\frac{1}{2}} + m_2\eta_b^{-\frac{1}{2}})\Phi^{\frac{1}{2}} \\ &= -\tfrac{1}{2}(m_1\xi_a^{\frac{1}{2}} + m_2\eta_a^{\frac{1}{2}})\Phi^{-\frac{1}{2}} + \tfrac{1}{2}(m_1\xi_a^{-\frac{1}{2}} + m_2\eta_a^{-\frac{1}{2}})\Phi^{\frac{1}{2}}. \end{aligned} \tag{43}$$

The value Φ_{cp} which Φ takes on in the center-of-momentum frame can be found by setting $P = 0$ in (43) and solving for Φ; if (38), (39), and (41) are then used, one finds that

$$\Phi_{cpA}^{-1} = \Phi_{cpB} = c\Lambda^2 = E^2\Lambda^2. \tag{44}$$

The parametric form of the orbits in the center-of-momentum frame can now be found. We consider case (A) first. The use of (30), (31), (41), (42), and (44) in (16a) and (16b) produces

$$\begin{aligned} x_1 = \tfrac{1}{2}[&(c_1 + E^{-1}\Lambda^{-1}f_1) + (c_2 + E\Lambda f_2)] \\ &+ 2e_1e_2\mu^{-1}E^{-1}\{-m_2^2(E^2 + m_1^2 - m_2^2) \\ &+ \tfrac{1}{2}E^2(E^2 - m_1^2 - m_2^2) \\ &+ m_2^2\mu\Lambda\xi^{\frac{1}{2}}[\Lambda\xi^{\frac{1}{2}}(E^2 + m_1^2 - m_2^2) \\ &- m_1(1 + E^2\Lambda^2\xi)]^{-1}\}, \end{aligned} \tag{45a}$$

$$\begin{aligned} t_1 = \tfrac{1}{2}[&(c_1 + E^{-1}\Lambda^{-1}f_1) - (c_2 + E\Lambda f_2)] \\ &+ 2e_1e_2\mu^{-1}E^{-1}\{-\tfrac{1}{2}E^2(E^2 - m_1^2 - m_2^2) \\ &+ m_1m_2^2(E^2 + m_1^2 - m_2^2)(1 - E^2\Lambda^2\xi) \\ &\times [\Lambda\xi^{\frac{1}{2}}(E^2 + m_1^2 - m_2^2) - m_1(1 + E^2\Lambda^2\xi)]^{-1} \\ &- 4m_1^2m_2^2E^2\mu^{-\frac{1}{2}} \ln|\Gamma_f(\xi)|\}. \end{aligned} \tag{45b}$$

The demonstration that the parametric form of the center-of-momentum-frame world line given in (45) agrees with the form given by (13) can be made by noting that y_1, as defined by (12), equals b_1 [defined in Eq. (11)] at the turning point, where $v_1 = 0$. It follows from (17a) and (44) that the center-of-momentum-frame turning point occurs at $\xi = \xi_{tpA}$ in case (A) where

$$\xi_{tpA} = \Phi_{cpA} = E^{-2}\Lambda^{-2}. \tag{46}$$

The use of (46) in (45a) shows that the value of x_1 at the turning point is

$$\begin{aligned} x_{1tp} = \tfrac{1}{2}[&(c_1 + E^{-1}\Lambda^{-1}f_1) + (c_2 + E\Lambda f_2)] \\ &+ b_1 + e_1e_2\mu^{-1}E(E^2 - m_1^2 - m_2^2). \end{aligned} \tag{47}$$

Thus, we must have

$$y_1 = x_1 - x_{1tp} + b_1. \tag{48}$$

It can now be shown, by using (11), (45a), (47), (48), and the definition of θ_i which follows (14), that

$$\theta_1(y_1^2 - b_1^2)^{\frac{1}{2}} = \frac{2e_1e_2\mu^{-\frac{1}{2}}E^{-1}m_1m_2^2(1 - E^2\Lambda^2\xi)}{\Lambda\xi^{\frac{1}{2}}(E^2 + m_1^2 - m_2^2) - m_1(1 + E^2\Lambda^2\xi^2)}. \tag{49}$$

It follows from (31), (38a), (45a), and (47)–(49) that

$$\begin{aligned} \Gamma_f(\xi) &= (\xi^{\frac{1}{2}} - \xi_{aA}^{\frac{1}{2}})/(\xi^{\frac{1}{2}} - \xi_{bA}^{\frac{1}{2}}) \\ &= -\xi_{aA}^{\frac{1}{4}}\xi_{bA}^{-\frac{1}{4}}[y_1 - \theta_1(y_1^2 - b_1^2)^{\frac{1}{2}}]^{\frac{1}{2}} \\ &\quad \times [y_1 + \theta_1(y_1^2 - b_1^2)^{\frac{1}{2}}]^{-\frac{1}{2}}. \end{aligned} \tag{50}$$

The use of (38), (49), and (50) in (45b) yields the form (13) for the world line with

$$t_{10} = \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) - (c_2 + E\Lambda f_2)] - e_1e_2\mu^{-1}E(E^2 - m_1^2 - m_2^2) + 8e_1e_2m_1^2m_2^2E\mu^{-\frac{3}{2}} \tanh^{-1}[\mu^{\frac{1}{2}}(E^2 + m_1^2 - m_2^2)^{-1}]. \tag{51}$$

A considerable amount of the labor involved in carrying out the details of algebraic manipulation required to obtain the world line of particle 2 can be avoided by noting, from inspection of (16), (30), (31), (34), and (35), that the interchange of m_1 and m_2 and the replacement of ξ by η carry f into $g + 2(\eta f_1 + f_2)$,

$$x_1 - \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) + (c_2 + E\Lambda f_2)]$$

into

$$-x_2 + \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) + (c_2 + E\Lambda f_2)],$$

and

$$t_1 - \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) - (c_2 + E\Lambda f_2)]$$

into

$$-t_2 + \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) - (c_2 + E\Lambda f_2)].$$

Hence, (45) is replaced by

$$x_2 = \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) + (c_2 + E\Lambda f_2)] + 2e_1e_2\mu^{-1}E^{-1}\{m_1^2(E^2 - m_1^2 + m_2^2) - \tfrac{1}{2}E^2(E^2 - m_1^2 - m_2^2) - m_1^2\mu\Lambda\eta^{\frac{1}{2}}[\Lambda\eta^{\frac{1}{2}}(E^2 - m_1^2 + m_2^2) - m_2(1 + E^2\Lambda^2\eta)]^{-1}\}, \tag{52a}$$

$$t_2 = \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) - (c_2 + E\Lambda f_2)] + 2e_1e_2\mu^{-1}E^{-1}\{\tfrac{1}{2}E^2(E^2 - m_1^2 - m_2^2) - m_1^2m_2(E^2 - m_1^2 + m_2^2)(1 - E^2\Lambda^2\eta) \times [\Lambda\eta^{\frac{1}{2}}(E^2 - m_1^2 + m_2^2) - m_2(1 + E^2\Lambda^2\eta)]^{-1} + 4m_1^2m_2^2E^2\mu^{-\frac{1}{2}}\ln|\Gamma_g(\eta)|\}. \tag{52b}$$

The turning point is at

$$\eta_{\text{tpA}} = \Phi_{\text{cpA}} = E^{-2}\Lambda^{-2}, \tag{53}$$

where

$$x_{2\text{tp}} = \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) + (c_2 + E\Lambda f_2)] - b_2 - e_1e_2\mu^{-1}E(E^2 - m_1^2 - m_2^2). \tag{54}$$

The remaining calculations go through as for particle 1; we have

$$y_2 = x_2 - x_{2\text{tp}} - b_2, \tag{55}$$

$$\theta_2(y_2^2 - b_2^2)^{\frac{1}{2}} = -\frac{2e_1e_2\mu^{-\frac{1}{2}}E^{-1}m_1^2m_2(1 - E^2\Lambda^2\eta)}{\Lambda\eta^{\frac{1}{2}}(E^2 - m_1^2 + m_2^2) - m_2(1 + E^2\Lambda^2\eta)}, \tag{56}$$

and

$$\Gamma_g(\eta) = (\eta^{\frac{1}{2}} - \eta_{b\text{A}}^{\frac{1}{2}})/(\eta^{\frac{1}{2}} - \eta_{a\text{A}}^{\frac{1}{2}}) = -\eta_{b\text{A}}^{\frac{1}{4}}\eta_{a\text{A}}^{-\frac{1}{4}}[y_2 - \theta_2(y_2^2 - b_2^2)^{\frac{1}{2}}]^{\frac{1}{2}} \times [y_2 + \theta_2(y_2^2 - b_2^2)^{\frac{1}{2}}]^{-\frac{1}{2}}. \tag{57}$$

Using these in (52b) produces an expression for the world line of the form (14) with

$$t_{20} = \tfrac{1}{2}[(c_1 + E^{-1}\Lambda^{-1}f_1) - (c_2 + E\Lambda f_2)] + e_1e_2\mu^{-1}E(E^2 - m_1^2 - m_2^2) - 8e_1e_2m_1^2m_2^2E\mu^{-\frac{3}{2}}\tanh^{-1}[\mu^{\frac{1}{2}}(E^2 - m_1^2 + m_2^2)^{-1}]. \tag{58}$$

The result (15) for the difference in turning times follows from (11), (51), and (58).

The same change of variables ($\xi' = \xi^{-1}$, $\eta' = \eta^{-1}$, $f' = \xi^{-1}f$, and $g' = \eta^{-1}g$) which was used to obtain f_{B} and g_{B} from f_{A} and g_{A} can be used to obtain the world lines for case (B) from the world lines for case (A). Under this change of variables,

$$\frac{\partial f'}{\partial \xi'} = f - \xi\frac{\partial f}{\partial \xi}, \quad f' - \xi'\frac{\partial f'}{\partial \xi'} = \frac{\partial f}{\partial \xi},$$

$$\frac{\partial g'}{\partial \eta'} = g - \eta\frac{\partial g}{\partial \eta}, \quad \text{and} \quad g' - \eta'\frac{\partial g'}{\partial \eta'} = \frac{\partial g}{\partial \eta}.$$

Taking account of (44), we see that the effect of this change of variables, which interchanges cases (A) and (B), is to interchange $\Phi^{\frac{1}{2}}\partial f/\partial\xi$ with $\Phi^{-\frac{1}{2}}(f - \xi\partial f/\partial\xi)$ and $\Phi^{\frac{1}{2}}\partial g/\partial\eta$ with $\Phi^{-\frac{1}{2}}(g - \eta\partial g/\partial\eta)$. It now follows from (16) that this interchange carries $x_i - \frac{1}{2}(c_1 + c_2)$ into itself while carrying $t_i - \frac{1}{2}(c_1 - c_2)$ into its negative. Hence, the form of the world lines for case (B) is the same as for case (A), but the sign of $t_{10} - t_{20}$ is reversed. This agrees with the results of Sec. II.

It can be readily verified that the f and g obtained in this section exhibit the behavior for ξ near ξ_b and for η near η_b which was deduced in Ref. 7 [Eq. (151)]. A Hamiltonian formulation of the present dynamics can be obtained by inserting f and g into the general results of Ref. 7. If this is done, it will be found that the constants H and P of Sec. II, which are made unique by the demand of asymptotic reduction to free-particle form, agree with the H and P of Ref. 7. However, the constant K of Sec. II, which cannot be made unique by the demand of asymptotic reduction to free-particle form, differs from the K of Ref. 7 by a constant which depends on ζ.

APPENDIX: COMPUTATION OF THE DIFFERENCE IN TURNING TIMES $t_{10} - t_{20}$

Comparison of the sum of Eqs. (9) and (10) with Eq. (6) yields

$$\frac{(m_2/E)^2}{x_1(t_1) - K/E - e_1e_2/E} - \frac{(m_1/E)^2}{x_2(t_2) - K/E + e_1e_2/E} = \frac{1}{x_1(t_1) - x_2(t_2)}. \quad \text{(A1)}$$

It follows from clearing fractions in (A1) and introducing the variables y_i defined by (12) that

$$m_1^2y_1^2 + (E^2 - m_1^2 - m_2^2)y_1y_2 + m_2^2y_2^2 + (2m_1m_2e_1e_2)^2\mu^{-1} = 0. \quad \text{(A2)}$$

Equations (A2) can be solved for either y_1 or y_2, with the results

$$2m_1^2y_1 = -(E^2 - m_1^2 - m_2^2)y_2 + \mu^{\frac{1}{2}}\varphi_2(y_2^2 - b_2^2)^{\frac{1}{2}}, \quad \text{(A3)}$$

$$2m_2^2y_2 = -(E^2 - m_1^2 - m_2^2)y_1 + \mu^{\frac{1}{2}}\varphi_1(y_1^2 - b_1^2)^{\frac{1}{2}}, \quad \text{(A4)}$$

where the φ_i are ± 1. For t_1, $t_2 \to \pm\infty$, $y_1 \to +\infty$ and $y_2 \to -\infty$. Consistency between (A3) and (A4) then requires that $\varphi_1 = \varphi_2$ in these limits; continuity implies that φ_i can change sign only at the turning points $y_1 = b_1$ and $y_2 = -b_2$.

It can be shown from Eqs. (A2)–(A4) that

$$\varphi_1\mu^{-\frac{1}{2}}(E^2 + m_1^2 - m_2^2)(y_1^2 - b_1^2)^{\frac{1}{2}} - \varphi_2\mu^{-\frac{1}{2}}(E^2 - m_1^2 + m_2^2)(y_2^2 - b_2^2)^{\frac{1}{2}} = y_1 - y_2 \quad \text{(A5)}$$

and

$$-[y_1 + \varphi_1(y_1^2 - b_1^2)^{\frac{1}{2}}][y_2 + \varphi_2(y_2^2 - b_2^2)^{\frac{1}{2}}] = 8(m_1m_2e_1e_2)^2\mu^{-2}(\mu^{\frac{1}{2}} + E^2 - m_1^2 - m_2^2). \quad \text{(A6)}$$

Equation (A6) and the relations

$$\varphi_i \tanh^{-1} [y_i^{-1}(y_i^2 - b_i^2)^{\frac{1}{2}}] = \ln |b_i^{-1}[y_i + \varphi_i(y_i^2 - b_i^2)^{\frac{1}{2}}]|,$$

$$\tanh^{-1} [\mu^{\frac{1}{2}}(E^2 - m_1^2 - m_2^2)^{-1}] = \ln [(2m_1m_2)^{-1}(\mu^{\frac{1}{2}} + E^2 - m_1^2 - m_2^2)]$$

now yield the result

$$\sum_{i=1}^{2} \varphi_i \tanh^{-1} [y_i^{-1}(y_i^2 - b_i^2)^{\frac{1}{2}}] = \tanh^{-1} [\mu^{\frac{1}{2}}(E^2 - m_1^2 - m_2^2)^{-1}]. \quad \text{(A7)}$$

By the use of (12), (A5), and (A7) in the light-cone condition (4c), it follows that

$$\begin{aligned} t_1 - t_2 &= \theta\varphi_1\{\mu^{-\frac{1}{2}}(E^2 + m_1^2 - m_2^2)(y_1^2 - b_1^2)^{\frac{1}{2}} \\ &\quad + 8\mu^{-\frac{3}{2}}e_1e_2m_1^2m_2^2E \tanh^{-1} [y_1^{-1}(y_1^2 - b_1^2)^{\frac{1}{2}}]\} \\ &\quad - \theta\varphi_2\{\mu^{-\frac{1}{2}}(E^2 - m_1^2 + m_2^2)(y_2^2 - b_2^2)^{\frac{1}{2}} \\ &\quad - 8\mu^{-\frac{3}{2}}e_1e_2m_1^2m_2^2E \tanh^{-1} [y_2^{-1}(y_2^2 - b_2^2)^{\frac{1}{2}}]\} \\ &\quad + \theta\{2e_1e_2E\mu^{-1}(E^2 - m_1^2 - m_2^2) \\ &\quad - 8\mu^{-\frac{3}{2}}e_1e_2m_1^2m_2^2E \tanh^{-1} [\mu^{\frac{1}{2}}(E^2 - m_1^2 - m_2^2)^{-1}]\}. \end{aligned} \quad \text{(A8)}$$

A comparison of (A8) with the result of subtracting Eq. (14) from Eq. (13) shows that $\varphi_i = \theta\theta_i$ and yields Eq. (15).

[*] Section III is based on a thesis submitted to the University of Delaware in partial fulfillment of the requirements for a M.S. degree awarded in June, 1969.

[1] A. Schild, Phys. Rev. **131**, 2762 (1963); A. Schild and J. A. Schlosser, J. Math. Phys. **9**, 913 (1968).

[2] R. D. Driver, Am. Math. Soc. Notices **7**, 522 (1960); Ann. Phys. (N.Y.) **21**, 122 (1963); R. D. Driver and M. J. Norris, *ibid.* **42**, 347 (1967); R. D. Driver, in *International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics*, edited by J. P. La Salle and S. Lefshetz (Academic, New York, 1963).

[3] R. D. Driver, Phys. Rev. **178**, 2051 (1969). The fact that instantaneous initial data does not determine a unique solution for many differential–difference equations makes this result far from obvious.

[4] J. L. Synge, Proc. Roy. Soc. (London) **A177**, 118 (1940).

[5] R. N. Hill, J. Math. Phys. **8**, 201 (1967). The question of the Lorentz covariance of the instantaneous interaction equations of motion which result from expansions in c^{-1} and/or e^2 is examined in this paper and by D. G. Currie, Phys. Rev. **142**, 817 (1966).

[6] The presence of these asymptotic interaction contributions was first remarked by H. Van Dam and E. P. Wigner, Phys. Rev. **142**, 838 (1966). See also J. H. Cooke, *ibid.* **174**, 1602 (1968).

[7] R. N. Hill, J. Math. Phys. **11**, 1918 (1970).

[8] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley, Reading, Mass., 1951), p. 177.

[9] The asymptotic region is discussed in more detail in Sec. V of Ref. 7.

[10] The redundancy of the parametric representation is discussed in more detail in the last part of Sec. I [following Eq. (44c)] of Ref. 7. Inspection of Eqs. (45) and (52) of the present paper shows that the choice of f_1 and f_2 fixes a "zero point" for the constants c_1 and c_2 which are changed by space–time translations; inspection of Eqs. (44), (45), and (52) shows that the choice of Λ fixes the scale for ξ and for the constant Φ which specifies a particular Lorentz frame.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Comparison of SCF and k_ν Functions for the Helium Series

DAVID M. BISHOP AND BARBARA E. SCHNEIDER
Department of Chemistry, University of Ottawa, Ottawa 2, Canada

(Received 29 January 1970)

Wavefunctions for He, Li^+, Be^{2+}, and O^{6+} are presented. They were determined by using reduced modified Bessel functions of the second kind, $k_\nu(qr)$. The z dependence of energies calculated using such functions for high values of z is found to be the same as for the Hartree-Fock functions.

Bishop and Somorjai[1] have recently shown that the ground-state energies of the helium series calculated with $k_\nu(qr)$ functions are in excellent agreement with the best Hartree-Fock values.[2] The functions $k_\nu(qr)$ are reduced modified Bessel functions of the second kind and are defined by

$$k_\nu(qr) = (qr)^\nu K_\nu(qr),$$

where $K_\nu(x)$ is the normal modified Bessel function of the second kind. The parameters ν and q were optimized variationally[1] for He, Li^+, Be^{2+}, and O^{6+} for a total wavefunction of the form

$$\Psi = \psi(r_1)\psi(r_2),$$

where

$$\psi(r_1) = Nk_\nu(qr_1)$$

and r_1 and r_2 are distances of the two electrons from the nucleus. The normalization constant N is given by

$$N^{-2} = 2^{2\nu+2}q^{-3}\pi[\Gamma(\nu + \tfrac{3}{2})]^2B(2\nu + \tfrac{3}{2}, \tfrac{3}{2}),$$

where $B(x, y)$ is the beta function.

The excellence of the energy results has encouraged us to verify whether or not the wavefunction itself is equally satisfactory. We have therefore evaluated the radial function

$$P_{\rm BS}(r) = rNk_\nu(qr)$$

for which

$$4\pi\int_0^\infty P_{\rm BS}^2(r)\,dr = 1,$$

for several values of r and using the optimized values of q and ν which were found before.[1] In Table I, we display for various r values both $P_{\rm BS}(r)$ and the difference between $P_{\rm BS}(r)$ and the best Hartree-Fock values, given by Roothaan, Sachs, and Weiss[2], $P_{\rm HF}(r)$. The latter values were normalized such that

$$4\pi\int_0^\infty P_{\rm HF}^2(r)\,dr = 1$$

and were obtained from the equations

$$P_{\rm HF}(r) = r\sum \frac{C_iR_i(r)}{4\pi},$$

$$R_i(r) = [(2n_i!)]^{-\frac{1}{2}}(2\zeta_i)^{n_i+\frac{1}{2}}r^{n_i-1}e^{-\zeta_i r}.$$

It should be noted that the n_i's in Table VI of Ref. 2, are, in fact, $n_i - 1$ and that for $z = 3$ of the same table it appears to us that the first two columns of coefficients (C_i) should be interchanged.

Table I shows that the difference between the two wavefunctions (Δ) is strikingly small, in spite of the fact that $k_\nu(qr)$ function contains only two parameters, whereas the Hartree-Fock function for He, for example, contains two ζ values and eleven linear coefficients. It would appear that, as well as producing good energies, the $k_\nu(qr)$ basis set also gives quite accurate wavefunctions.

It was apparent from the previous work[1] that, as z, the nuclear charge, increased, the ground-state energies of these two-electron systems were improved; we have now determined the energies for some high z values (10-100) to determine whether we approach the same z limit as the Hartree-Fock case. The optimized q and ν values together with the energies are given in Table II. We find that the difference between our energies and $-z^2 + \frac{5}{8}z$ is practically constant (0.1089 a.u.), and we may therefore write

$$E = -z^2 + \tfrac{5}{8}z - 0.1089$$

for high z values. This can be compared with the result of Linderberg[3] who has calculated the Hartree-Fock energy to be

$$E = -z^2 + \tfrac{5}{8}z - 0.11100317 - 0.00105525z^{-1}.$$

It is apparent that for high z values our energies will approach the Hartree-Fock values.

The evidence given here supports the contention

Table I. Radial functions for He, Li^{+}, Be^{2+}, and O^{6+}.

r(a.u.)	He $P_{BS}(r)$	He Δ[a]	Li^{+} $P_{BS}(r)$	Li^{+} Δ	Be^{2+} $P_{BS}(r)$	Be^{2+} Δ	O^{6+} $P_{BS}(r)$	O^{6+} Δ
0.01	0.01375	0.00061	0.02599	0.00063	0.04042	0.00060	0.11343	0.00044
0.02	0.02662	0.00085	0.05000	0.00079	0.07715	0.00062	0.20876	0.00010
0.03	0.03880	0.00091	0.07237	0.00072	0.11074	0.00041	0.28874	−0.00034
0.04	0.05038	0.00085	0.09327	0.00054	0.14150	0.00011	0.35538	−0.00068
0.05	0.06141	0.00071	0.11282	0.00030	0.16966	−0.00021	0.41036	−0.00089
0.10	0.10961	−0.00041	0.19326	−0.00094	0.27785	−0.00134	0.55538	−0.00027
0.15	0.14825	−0.00148	0.25036	−0.00150	0.34373	−0.00128	0.56656	0.00093
0.20	0.17921	−0.00214	0.28955	−0.00135	0.37939	−0.00050	0.51501	0.00134
0.25	0.20380	−0.00235	0.31480	−0.00073	0.39350	0.00047	0.43955	0.00098
0.30	0.22304	−0.00218	0.32920	0.00009	0.39244	0.00130	0.36051	0.00029
0.35	0.23775	−0.00174	0.33518	0.00093	0.38096	0.00184	0.28768	−0.00043
0.40	0.24861	−0.00122	0.33468	0.00164	0.36261	0.00206	0.22500	−0.00099
0.45	0.25621	−0.00041	0.32925	0.00217	0.34000	0.00200	0.17331	−0.00137
0.50	0.26102	0.00033	0.32014	0.00250	0.31505	0.00172	0.13189	−0.00157
0.55	0.26348	0.00105	0.30836	0.00262	0.28916	0.00129	0.09940	−0.00162
0.60	0.26394	0.00171	0.29471	0.00256	0.26332	0.00078	0.07431	−0.00156
0.65	0.26272	0.00228	0.27984	0.00235	0.23820	0.00024	0.05518	−0.00143
0.70	0.26010	0.00275	0.26424	0.00202	0.21428	−0.00030	0.04074	−0.00126
0.75	0.25630	0.00311	0.24833	0.00161	0.19182	−0.00079	0.02993	−0.00107
0.80	0.25154	0.00336	0.23241	0.00114	0.17100	−0.00124	0.02189	−0.00089
0.85	0.24599	0.00352	0.21672	0.00065	0.15188	−0.00162	0.01595	−0.00073
0.90	0.23981	0.00357	0.20144	0.00014	0.13445	−0.00193	0.01158	−0.00059
0.95	0.23313	0.00353	0.18670	−0.00035	0.11868	−0.00217	0.00839	−0.00047
1.00	0.22607	0.00342	0.17259	−0.00082	0.10448	−0.00235	0.00606	−0.00038
1.05	0.21873	0.00323	0.15918	−0.00126	0.09177	−0.00246	0.00436	−0.00031
1.10	0.21120	0.00299	0.14651	−0.00167	0.08043	−0.00253	0.00314	−0.00025
1.15	0.20355	0.00269	0.13458	−0.00203	0.07035	−0.00254	0.00225	−0.00020
1.20	0.19584	0.00236	0.12341	−0.00234	0.06143	−0.00252	0.00161	−0.00016
1.25	0.18813	0.00200	0.11299	−0.00261	0.05355	−0.00247	0.00115	−0.00013
1.30	0.18047	0.00161	0.10329	−0.00284	0.04661	−0.00239	0.00082	−0.00010
1.35	0.17289	0.00121	0.09429	−0.00302	0.04052	−0.00229	0.00059	−0.00008
1.40	0.16542	0.00080	0.08597	−0.00316	0.03517	−0.00218	0.00042	−0.00006
1.45	0.15810	0.00039	0.07830	−0.00327	0.03050	−0.00206	0.00030	−0.00004
1.50	0.15094	−0.00002	0.07123	−0.00334	0.02641	−0.00193	0.00021	−0.00002
1.55	0.14396	−0.00042	0.06473	−0.00339	0.02285	−0.00180	0.00015	−0.00001
1.60	0.13718	−0.00081	0.05877	−0.00340	0.01975	−0.00167	0.00011	0.00000
1.65	0.13060	−0.00119	0.05330	−0.00339	0.01706	−0.00155	0.00007	0.00001
1.70	0.12424	−0.00155	0.04831	−0.00336	0.01472	−0.00142	0.00005	0.00002
1.75	0.11809	−0.00189	0.04375	−0.00331	0.01269	−0.00131	0.00004	0.00002
1.80	0.11217	−0.00221	0.03959	−0.00324	0.01093	−0.00119	0.00003	0.00003
1.85	0.10647	−0.00250	0.03580	−0.00317	0.00941	−0.00109	0.00002	0.00003
1.90	0.10100	−0.00278	0.03235	−0.00308	0.00810	−0.00099	0.00001	0.00003
1.95	0.09575	−0.00304	0.02922	−0.00298	0.00696	−0.00089		
2.00	0.09072	−0.00327	0.02637	−0.00287	0.00598	−0.00081		
2.05	0.08590	−0.00348	0.02378	−0.00276	0.00513	−0.00073		
2.10	0.08130	−0.00366	0.02144	−0.00265	0.00441	−0.00066		
2.15	0.07691	−0.00383	0.01932	−0.00253	0.00378	−0.00059		
2.20	0.07272	−0.00397	0.01740	−0.00242	0.00324	−0.00053		
2.25	0.06872	−0.00410	0.01566	−0.00230	0.00278	−0.00047		
2.30	0.06492	−0.00420	0.01409	−0.00218	0.00238	−0.00042		
2.35	0.06130	−0.00429	0.01268	−0.00207	0.00204	−0.00038		
2.40	0.05786	−0.00435	0.01140	−0.00196	0.00174	−0.00034		
2.45	0.05459	−0.00441	0.01024	−0.00185	0.00149	−0.00030		
2.50	0.05149	−0.00444	0.00920	−0.00174	0.00127	−0.00027		
2.55	0.04855	−0.00446	0.00826	−0.00164	0.00109	−0.00024		
2.60	0.04576	−0.00447	0.00742	−0.00155	0.00093	−0.00021		
2.65	0.04312	−0.00447	0.00665	−0.00145	0.00079	−0.00019		
2.70	0.04061	−0.00445	0.00597	−0.00136	0.00068	−0.00016		
2.75	0.03824	−0.00442	0.00535	−0.00128	0.00058	−0.00014		
2.80	0.03600	−0.00439	0.00480	−0.00120	0.00049	−0.00013		
2.85	0.03388	−0.00434	0.00430	−0.00112	0.00042	−0.00011		
2.90	0.03188	−0.00429	0.00385	−0.00105	0.00036	−0.00010		
2.95	0.02998	−0.00423	0.00345	−0.00098	0.00031	−0.00008		
3.00	0.02819	−0.00416	0.00309	−0.00092	0.00026	−0.00007		
3.05	0.02651	−0.00409	0.00277	−0.00086	0.00022	−0.00006		
3.10	0.02491	−0.00401	0.00248	−0.00080	0.00019	−0.00006		
3.15	0.02341	−0.00393	0.00222	−0.00075	0.00016	−0.00005		
3.20	0.02200	−0.00385	0.00198	−0.00070	0.00014	−0.00004		
3.25	0.02066	−0.00376	0.00177	−0.00066	0.00012	−0.00003		
3.30	0.01940	−0.00367	0.00159	−0.00061				
3.35	0.01822	−0.00358	0.00142	−0.00057				
3.40	0.01710	−0.00348	0.00127	−0.00054				

[a] $\Delta = P_{BS}(r) - P_{HF}(r)$.

TABLE II. Parameter values and energies for some He-like ions.

z	ν	q	E(a.u.)	Δ[a]
10	0.471211	9.500317	−93.85882	0.10882
20	0.485611	19.498010	−387.60884	0.10884
30	0.490409	29.497253	−881.35885	0.10885
40	0.492801	39.496710	−1575.10886	0.10886
50	0.494241	49.496481	−2468.85886	0.10886
60	0.495201	59.496331	−3562.60886	0.10886
70	0.495889	69.496323	−4856.35886	0.10886
80	0.496401	79.496149	−6350.10886	0.10886
90	0.496801	89.496090	−8043.85886	0.10886
100	0.497121	99.496045	−9937.60886	0.10886

[a] $\Delta = -z^2 + \frac{5}{8}z - E$.

that $k_\nu(qr)$ functions will be useful for solving atomic problems at the Hartree–Fock level. They are simple, contain only two parameters for optimization, and present no particular problem for integration.

Note added in proof: Though there are undoubtedly misprints in Table VI of Ref. 2. for $z = 3$, our remedy, the interchange of the first two columns of coefficients, is probably not correct and, therefore, the safest procedure would be to ignore column 5 of our Table I.

[1] D. M. Bishop and R. L. Somorjai, J. Math. Phys. **11**, 1150 (1970).
[2] C. C. J. Roothaan, L. M. Sachs, and A. W. Weiss, Rev. Mod. Phys. **32**, 186 (1960).
[3] J. Linderberg, Phys. Rev. **121**, 816 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Global Singularities and the Taub-NUT Metric*

DAVID A. FEINBLUM
Department of Physics, State University of New York, Albany, New York,
and
Department of Nuclear Physics, Weizmann Institute of Science, Rehovot, Israel

(Received 5 January 1970)

Several examples of singular behavior which are not apparent in the line element are discussed. It is shown that the presence of such singularities depends on the manner in which coordinate patches are assembled to form the entire manifold. It is shown that the usual connection of a coordinate patch with the Taub metric to a patch with the NUT metric is singular in the sense of this particle.

1. INTRODUCTION

The method of obtaining a geometry for a model universe that satisfies the Einstein equations may be outlined quite roughly as follows: (1) Obtain a metric which satisfies the equations, perhaps using several coordinate patches; (2) piece these patches together to form a "complete" manifold. The term "complete" is usually taken as meaning *geodesically complete*,[1] but here it is used in the sense of a *complete uniform space*.[2] The latter notion is felt by the author to express most precisely the desired notion and to be more convenient in practical cases.[3] The term *global singularity* is used to indicate some property of the model universe which causes it to fail to be a smooth manifold with a well-behaved causal structure and which is not observable just by inspection of the metric tensor. Not much attention has been given to global problems beyond the recognition that, if the universe is regarded as the union of several pieces, the universe must be well behaved at the boundary between them. A distinct exception to this last remark is some of the work of Misner,[4] who has explicitly remarked on the need for global considerations in the construction of model universes. In this article, several types of global singularities are discussed, and consideration is given to how one might recognize a strange (but not necessarily singular) local behavior as an indication suggesting the presence of a global singularity.

The operation of "piecing the patches" together, mentioned above, is a little more elaborate than one might deduce from the statement as given. If the manifold is not quite well behaved, there may be a need to "complete" it by the addition of singular points.[1–3] A corollary to the consideration of certain global singularities, which provoke the need for such completion, is the possibility of discussing how one would obtain the singular points correctly if (as would occur in practice) they were not provided in advance. Finally, it will be seen how these singularities may appear in an actual assembly of a universe in a case which has appeared in the literature.

2. THE CONICAL SINGULARITY

The 2-dimensional surface with line element

$$ds^2 = dr^2 + r^2\, d\theta^2 \tag{1}$$

TABLE II. Parameter values and energies for some He-like ions.

z	ν	q	E(a.u.)	Δ^a
10	0.471211	9.500317	−93.85882	0.10882
20	0.485611	19.498010	−387.60884	0.10884
30	0.490409	29.497253	−881.35885	0.10885
40	0.492801	39.496710	−1575.10886	0.10886
50	0.494241	49.496481	−2468.85886	0.10886
60	0.495201	59.496331	−3562.60886	0.10886
70	0.495889	69.496323	−4856.35886	0.10886
80	0.496401	79.496149	−6350.10886	0.10886
90	0.496801	89.496090	−8043.85886	0.10886
100	0.497121	99.496045	−9937.60886	0.10886

[a] $\Delta = -z^2 + \frac{5}{8}z - E$.

that $k_\nu(qr)$ functions will be useful for solving atomic problems at the Hartree–Fock level. They are simple, contain only two parameters for optimization, and present no particular problem for integration.

Note added in proof: Though there are undoubtedly misprints in Table VI of Ref. 2. for $z = 3$, our remedy, the interchange of the first two columns of coefficients, is probably not correct and, therefore, the safest procedure would be to ignore column 5 of our Table I.

[1] D. M. Bishop and R. L. Somorjai, J. Math. Phys. **11**, 1150 (1970).
[2] C. C. J. Roothaan, L. M. Sachs, and A. W. Weiss, Rev. Mod. Phys. **32**, 186 (1960).
[3] J. Linderberg, Phys. Rev. **121**, 816 (1961).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Global Singularities and the Taub-NUT Metric*

DAVID A. FEINBLUM
Department of Physics, State University of New York, Albany, New York,
and
Department of Nuclear Physics, Weizmann Institute of Science, Rehovot, Israel

(Received 5 January 1970)

Several examples of singular behavior which are not apparent in the line element are discussed. It is shown that the presence of such singularities depends on the manner in which coordinate patches are assembled to form the entire manifold. It is shown that the usual connection of a coordinate patch with the Taub metric to a patch with the NUT metric is singular in the sense of this particle.

1. INTRODUCTION

The method of obtaining a geometry for a model universe that satisfies the Einstein equations may be outlined quite roughly as follows: (1) Obtain a metric which satisfies the equations, perhaps using several coordinate patches; (2) piece these patches together to form a "complete" manifold. The term "complete" is usually taken as meaning *geodesically complete*,[1] but here it is used in the sense of a *complete uniform space*.[2] The latter notion is felt by the author to express most precisely the desired notion and to be more convenient in practical cases.[3] The term *global singularity* is used to indicate some property of the model universe which causes it to fail to be a smooth manifold with a well-behaved causal structure and which is not observable just by inspection of the metric tensor. Not much attention has been given to global problems beyond the recognition that, if the universe is regarded as the union of several pieces, the universe must be well behaved at the boundary between them. A distinct exception to this last remark is some of the work of Misner,[4] who has explicitly remarked on the need for global considerations in the construction of model universes. In this article, several types of global singularities are discussed, and consideration is given to how one might recognize a strange (but not necessarily singular) local behavior as an indication suggesting the presence of a global singularity.

The operation of "piecing the patches" together, mentioned above, is a little more elaborate than one might deduce from the statement as given. If the manifold is not quite well behaved, there may be a need to "complete" it by the addition of singular points.[1–3] A corollary to the consideration of certain global singularities, which provoke the need for such completion, is the possibility of discussing how one would obtain the singular points correctly if (as would occur in practice) they were not provided in advance. Finally, it will be seen how these singularities may appear in an actual assembly of a universe in a case which has appeared in the literature.

2. THE CONICAL SINGULARITY

The 2-dimensional surface with line element

$$ds^2 = dr^2 + r^2\, d\theta^2 \qquad (1)$$

and periodicity $(r, \theta) = (r, \theta + n\phi)$, with n integral and ϕ fixed, but unequal to 2π (that is, the ordinary cone), exhibits a well-understood but, nonetheless, unusual singular behavior. It is worthwhile to restate these well-known data in a fashion which emphasizes the global nature of the behavior. The local geometry, as determined from the line element, is flat wherever it is defined. Consequently, the curvature-tensor components as calculated in a local Cartesian coordinate system (and therefore in all coordinate systems) vanish identically, and their limits—as one considers points approaching the cone vertex—vanish. Since the polar coordinates used exhibit a coordinate singularity at $r = 0$, one would normally obtain the local geometry by examining patches for decreasing r (usually by inspection) rather than by using a new coordinate patch. For this manifold, it has been observed that the former procedure yields a misleading impression and the latter procedure (because of the singularity) is impossible to perform. This difficulty shows the existence of the singularity, but reveals nothing of the character of it.

A geometry in which we are more interested possesses the metric

$$ds^2 = \epsilon(-dr^2 + r^2\, d\theta^2). \tag{2}$$

Again, θ is periodic with period ϕ. $\epsilon = \pm 1$, so that either direction may be the timelike one. If the r direction is timelike, all observers traveling on radial geodesics collide at the origin, and the geometry implodes. If it is the θ direction which is timelike, an observer with world line $r =$ a small constant traverses a closed timelike curve of small duration per orbit, while time is stationary for the observer at $r = 0$. This behavior (and the general behavior of the geodesics) does not change even if the period of θ is 2π. Cones with different periods, nonetheless, have the same local geometry. Consequently, the geodesic equations are the same in all cases, and the only influence of the period is to affect how much θ must change to generate one orbit around the cone. Figure 1 assumes a period of 2π; this yields a geometry which possesses an isometric embedding of the constant r and constant θ curves into a plane. The geodesics may be obtained very easily by noting that, in the neighborhood of any point but the origin, the geometry is flat. Because of this, Cartesian coordinates may be constructed locally; the transformation

$$x = r\cosh\theta,$$
$$y = r\sinh\theta,$$

suffices. The equation of a geodesic is of the form

$$r\sinh\theta = mr\cosh\theta + b. \tag{3}$$

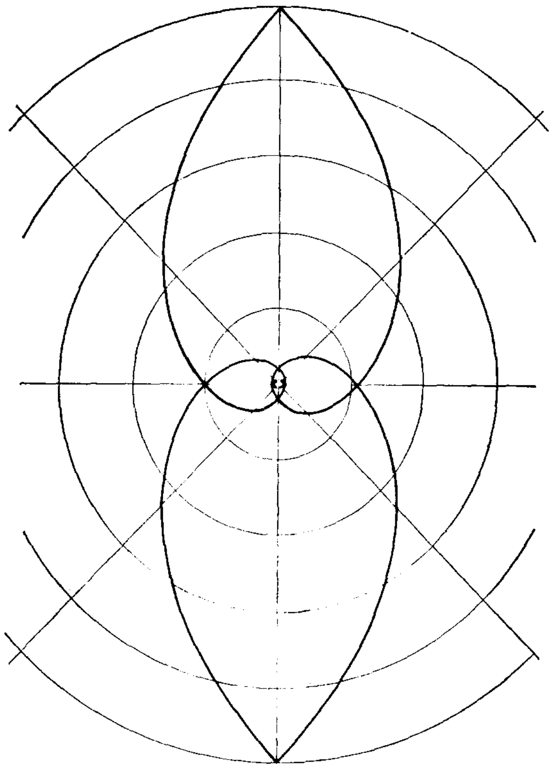


FIG. 1. A 2-dimensional geometry (the "flat cone") with metric $ds^2 = -dr^2 + r^2\, d\theta^2$ showing several null geodesics.

Another form of the geodesic equation is

$$(1 - m)e^{\theta} - (1 + m)e^{-\theta} = b/r. \tag{4}$$

The constants m and b are fixed once the initial point and initial slope of the geodesic are determined. There is a geodesic with fixed θ (when $b = 0$, $m = \tanh\theta_0$) but none with fixed r. The null geodesics satisfy

$$r = be^{\pm\theta}. \tag{5}$$

All geodesics reach the origin in finite affine length. All but radial geodesics spiral infinitely often as $r \to 0$. Consequently, it is not possible to extend any nonradial geodesic through the origin.

If one examines the thickening of a geodesic (i.e., the region swept out by perturbing the initial conditions of a given geodesic), he finds an unexpected result. Consider the geodesic obtained by perturbing $\mathbf{b}$ to $\mathbf{b}(1 + \eta)$ and m to $m + \delta$. Retain θ as the independent variable, and denote the perturbed radial coordinate by ρ. Then $\rho(\theta) \to r(\theta)$ as δ and η tend to 0. One finds that

$$\frac{\rho}{r} = 1 + \eta + \frac{\rho}{b} 2\delta\cosh\theta$$
$$= 1 + \eta + \frac{2\delta\cosh\theta(b + \eta)}{(1 - m - \delta)e^{\theta} - (1 + m + \delta)e^{-\theta}}. \tag{6}$$

Unless m is within δ of ± 1, one has, for $\theta \to +\infty$,

$$\frac{\rho}{r} \to 1 + \eta + \delta \frac{1+\eta}{1-m-\delta}$$

and, for $\theta \to -\infty$,

$$\frac{\rho}{r} \to 1 + \eta - \delta \frac{1+\eta}{1+m+\delta}. \tag{7}$$

If one investigates the null geodesics, one finds that zero denominators only occur as θ diverges in the sense which causes r to diverge and *not* in the sense which causes r to approach 0. Thus, the amount of radial thickening tends to a constant fraction of the radial distance to the origin (see Fig. 2). The implication of this is that one can find two geodesics with the same end point such that neither geodesic ever remains in the thickening of the other. All that is necessary is that one geodesic spirals in toward $r = 0$ as θ increases, while the other does so for decreasing θ (Figs. 1 and 2). This type of result is never obtained in a regular geometry. However, thickenings of radial geodesics behave more reasonably. Perturbations of a radial geodesic yield (see Fig. 3)

$$\rho = \frac{\eta}{\delta} \frac{1}{e^{-\theta} - e^{\theta}}. \tag{8}$$

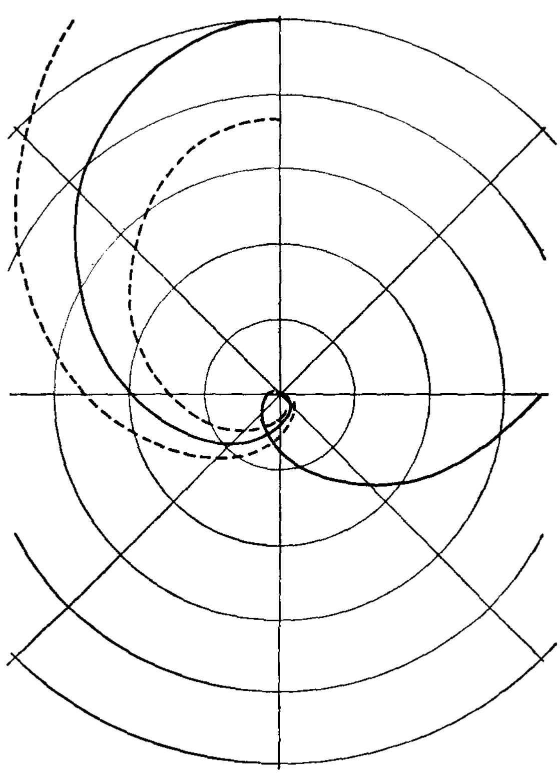


FIG. 2. A spacelike geodesic and a thickening of it in the flat cone. Another spacelike geodesic is shown, which shares an end point with the first geodesic but which does not remain in any thickening of it.

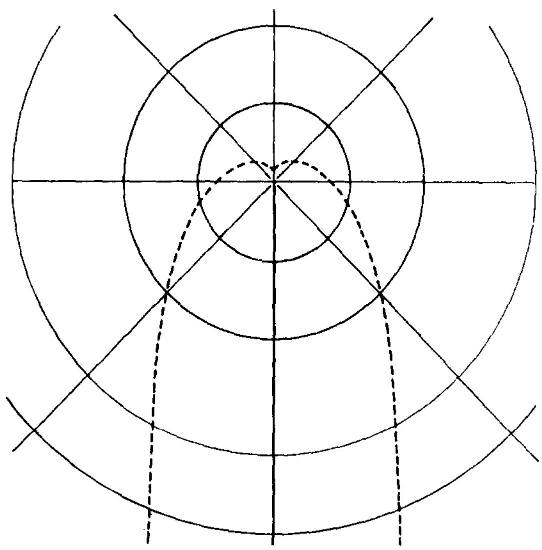


FIG. 3. A thickening of a radial geodesic in the flat cone.

Each geodesic eventually stays in each thickening of a radial geodesic, but the converse does not hold.

An important application of these elementary remarks is that there are geometries G which one can render incomplete by deleting a closed set C, complete by the technique proposed in Ref. 1, and *not* get the geometry which is obtained by deleting from G only the interior of C. If one wishes a revision of this technique, which adds back the vertex and only this point from a cone with the vertex removed, one can modify the prescription as follows. As in Ref. 1, we add an end point to each incomplete geodesic and determine equivalence classes of geodesics. We declare, as in Ref. 1, that two equivalent geodesics share an end point.[5] The modification consists in the definition of when two geodesics are equivalent. The following definition works in the case of a cone: Geodesics γ_1 and γ_2 are equivalent if there is a γ_3 such that both γ_1 and γ_2 enter and remain in each thickening of γ_3. There is more than ad hoc reasoning underlying this proposal. The idea of using the concept of thickening is valid if the thickening of a geodesic contains the geodesic in its interior even in the completed space. The discussion of this section shows that this is false on a cone. However, the proposed relation merely assumes that each boundary point is reached by one geodesic whose thickening is a neighborhood even after completion. On the cone, each radial geodesic has the required property. Whether there will always be such a geodesic for each boundary point is by no means clear, however. If there is none, the boundary consists of too many points. Indeed, the relation may not be an equivalence relation in this case, which would at least serve as a warning.

Some higher-dimensional versions of this singularity may be treated quite easily with the aid of the results for the 2-dimensional case. Consider the metric

$$ds^2 = \epsilon(-dr^2 + r^2\,d\theta^2) + dz^2 + dw^2. \tag{9}$$

The substitutions $x = r\cosh\theta$ and $y = r\sinh\theta$ lead to the metric

$$ds^2 = \epsilon(dx^2 - dy^2) + dz^2 + dw^2. \tag{10}$$

Choose the coordinates so that $y = z = w = 0$ at the initial point in the geodesic. One finds, by converting the expressions for the geodesic in the (x, y, z, w) system to the (r, θ, z, w) system,

$$\begin{aligned} r\sinh\theta &= mr\cosh\theta + b,\\ (m^{-1} + cz^{-1})\tanh\theta &= 1,\\ (m^{-1} + fw^{-1})\tanh\theta &= 1. \end{aligned} \tag{11}$$

The projections of the geodesics onto the (r, θ) plane follow the equations of the 2-dimensional case. Since $\tanh\theta \to 1$ very rapidly as $\theta \to \pm\infty$, in the regions of interest the geodesic structure is nearly the same as in the 2-dimensional case. It is now possible to find two geodesics γ_1 and γ_2 with the same end point which never intersect and such that γ_i does not remain in each thickening of γ_j, $i, j = 1, 2$, $i \neq j$. However, γ_i does enter each thickening of γ_j infinitely often, as there is no focusing in the z and w directions similar to that found in the (r, θ) plane. For example, if one induces the perturbation $b \to b + \delta$, $m \to m + \eta$, $c \to c + \alpha$ (for simplicity, the initial point is not perturbed in the z or w directions and the initial slope is not perturbed in the w direction), one finds

$$\frac{\xi}{z} = +\frac{\gamma}{c} - \frac{\eta}{cm(m + \eta)}, \tag{12}$$

where ξ is the perturbed coordinate value in the z direction. The perturbations in the r and θ directions are the same as in the 2-dimensional case. The geodesics for which θ is constant do behave in a different fashion than in the 2-dimensional case, because z and w depend linearly on r. The tendency toward constancy of z and w as r gets small, exhibited by all other geodesics, is totally absent, but nothing unexpected occurs with respect to thickenings of this latter class. If the end point is $e = (0, 0, z, w)$, for any given thickening of a constant-θ geodesic, the "ball"

$$\left\{x : \int |g_{\mu\nu}\,dx^\mu\,dx^\nu|^{\frac{1}{2}} < \Delta^2\right\} \tag{13}$$

is entirely within the thickening for some $\Delta > 0$.

One other conical singularity may be discussed very easily. Consider the metric

$$ds^2 = \epsilon(-dr^2 + r^2\,d\theta^2) + r^2\sin^2\theta\,d\phi^2, \tag{14}$$

with $0 \leq \theta \leq \pi$ and a periodicity of 2π in the ϕ coordinate. If $\epsilon = +1$, then the entire 2-sphere ($r =$ const) degenerates to a point as r goes to zero. However, for any initial point p and any vector $\mathbf{v}$ at that point, there is a plane P which contains the geodesic starting at p in the direction of $\mathbf{v}$. This is immediately seen because of the symmetry of the sphere. The angular coordinates may be chosen so that $\theta = \frac{1}{2}\pi$ on P. This duplicates the 2-dimensional cone, except that the angular coordinate is now called ϕ.

If $\epsilon = -1$ or if the metric has the form

$$ds^2 = dr^2 + r^2(d\theta^2 - \sin^2\theta\,d\phi^2), \tag{15}$$

a small neighborhood of $\theta = 0$ or $\theta = \pi$ is nearly isometric with the (r, θ, z) subspace of the 4-dimensional case treated above. One exhibits the quasi-isometry by defining $z = r$, $\bar{r} = r\theta$, and $\bar{\theta} = \phi$. To the degree of approximation needed to set $\sin\theta = \theta$, one has the desired exhibition.

3. A "SPHERE" THAT IS A CYLINDER

The topology of the NUT region of Taub–NUT space has been claimed to be $S^3 \times R$ for two separate reasons: (1) The NUT metric itself requires these identifications; (2) this is the topology of Taub space, and the NUT region is a smooth continuation of Taub space. In this section, it is shown that one cannot read the identification out of the line element and that, therefore, the first argument is not compelling. The second argument is considered in Sec. 5. We proceed from elementary situations and increase the complexity until an example is considered which includes Taub–NUT space itself. It is also shown that the presence or absence of a global singularity may depend crucially on the choices of identifications made. It is this property which is responsible for the use of the term "global" in describing these singularities.

Spherical coordinates (r, θ, ϕ) contain a source of confusion. One declares that θ is restricted to $[0, \pi]$ and that ϕ is restricted to $[0, 2\pi]$. He then declares that the ϕ coordinate is periodic. Not stated, but tacitly assumed, is the requirement that two points with zero distance between them are the same. In this manner, one realizes that $(r, 0, 0)$ is the same point as $(r, 0, \phi)$. However, there is the chance to confuse a sphere with a cylinder, for a cylinder admits a coordinate system with the same periodicity conditions.

Consider a collection of cylinders each with line element

$$ds^2 = d\theta^2 + d\phi^2, \quad 0 \leq \theta \leq \pi, \quad 0 \leq \phi \leq 2\pi, \tag{16}$$

and periodicity conditions $(\theta, 0) = (\theta, 2\pi)$. Embed them into a 3-geometry (ψ, θ, ϕ) such that each cylinder is specified by the relation $\psi = \text{const}$. Perform this embedding in a manner such that the line element of the 3-geometry is

$$ds^2 = d\psi^2 - 2\cos\theta\, d\psi\, d\phi + d\phi^2 + d\theta^2. \quad (17)$$

The 3-geometry has at least a coordinate singularity at $\theta = 0, \pi$ since the distance between the circles $\theta = 0$, $\psi = \psi_0$ and $\theta = 0$, $\psi = \psi_1$ vanishes. The line element may be rewritten as

$$ds^2 = (d\psi - \cos\theta\, d\phi)^2 + (d\theta^2 + \sin^2\theta\, d\phi^2). \quad (18)$$

One might be tempted to view each $\psi = \psi_0$ surface as a 2-sphere, due to the form of the last parenthetical expression, but he is quickly disabused because the length of the circles $\psi = \psi_0$, $\theta = 0, \pi$ does not vanish. However, the transformations $\bar\psi = \psi + \phi$, $\tilde\psi = \psi - \phi$ result in the line elements

$$\begin{aligned} ds^2 &= (d\bar\psi - 2\cos^2\tfrac12\theta\, d\phi)^2 + (d\theta^2 + \sin^2\theta\, d\phi^2) \\ &= (d\tilde\psi + 2\sin^2\tfrac12\theta\, d\phi)^2 + (d\theta^2 + \sin^2\theta\, d\phi^2). \end{aligned} \quad (19)$$

In the $\bar\psi$ and $\tilde\psi$ coordinate systems, there seems little reason not to interpret the geometry of the surface $\bar\psi = \bar\psi_0$, $\theta > \frac12\pi$ or the surface $\tilde\psi = \tilde\psi_0$, $\theta < \frac12\pi$ as a distorted hemisphere. If one makes the transformation $\psi^* = \psi + \phi f(\theta)$, where $f'(0) = f'(\pi) = f''(0) = f''(\pi) = 0$, $f(0) = 1$, and $f(\pi) = -1$, he finds $\psi^* \to \bar\psi$ as $\theta \to 0$ and $\psi^* \to \tilde\psi$ as $\theta \to \pi$. Certainly, terms involving $d\psi^*\, d\theta$ are generated, but they vanish at $\theta = 0, \pi$ and are not a cause for concern. In the $(\psi^*, 0, \phi)$ system, one would certainly take the surface $\psi^* = \psi^*_0$ to be (topologically) a sphere.

There is no paradox. If one insists on the original equivalence relation, which, to be careful, must be written $(\psi, \theta, 0) = (\psi, \theta, 2\pi)$, he finds that in the other systems one must set $(\bar\psi, \theta, 0) = (\bar\psi + 2\pi, \theta, 2\pi)$, $(\tilde\psi, \theta, 0) = (\tilde\psi - 2\pi, \theta, 2\pi)$, and $(\psi^*, \theta, 0) = (\psi^* + 2\pi f, \theta, 2\pi)$. With these relations, there is a (global) isometry between the four metrics-cum-coordinate system on the space. The reason for the vanishing of the length of the curve $\tilde\psi = \tilde\psi_0$, $\theta = 0$, $0 \le \phi \le 2\pi$, is that the vectors ∂_ψ and ∂_ϕ are anti-parallel when $\theta = 0$. If $\tilde\psi$ is held fixed, then $d\psi = d\phi$ and the curve is one single point, as one might expect (see Fig. 4).

We do not intend to imply that one cannot use the natural identification in the $\bar\psi$, $\tilde\psi$, or ψ^* coordinate systems, but they do lead to different spaces. For example, in the $\bar\psi$ system, the surface $\bar\psi = \bar\psi_0$ is a well-behaved surface, and a simply connected neighborhood of $\theta = 0$ exists in this space, while in the ψ space there is none.

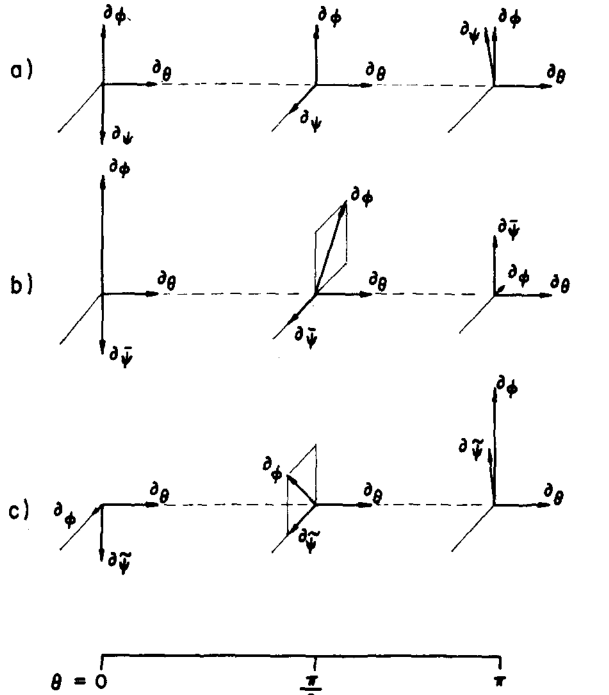


FIG. 4. The vectors (a) $(\partial_\theta, \partial_\phi, \partial_\psi)$, (b) $(\partial_\theta, \partial_\phi, \partial_{\bar\psi})$, and (c) $(\partial_\theta, \partial_\phi, \partial_{\tilde\psi})$, as a function of θ for the geometry with metric

$$ds^2 = (d\psi - \cos\theta\, d\phi)^2 + d\theta^2 + \sin^2\theta\, d\phi^2.$$

In each case the basis vectors are the orthonormal set which, in the first coordinate system, is given by

$$e_1 = \partial_\theta, \quad e_2 = \partial_\phi, \quad e_3 = (\sin\theta)^{-1}(\partial_\psi + \cos\theta\, \partial_\phi).$$

The symbol ∂_θ represents the vector with (θ, ϕ, ψ) components $(1, 0, 0)$ and similarly for the other vectors.

Now apply these remarks to the Taub metric.[6] There are coordinate systems in which this metric is denoted by the line elements[7]

$$\begin{aligned} ds^2 = -U^{-1}\, dt^2 &+ (2l)^2 U (d\tilde\psi + 2\sin^2\tfrac12\theta\, d\phi)^2 \\ &+ W(d\theta^2 + \sin^2\theta\, d\phi^2) \end{aligned} \quad (20)$$

and

$$\begin{aligned} ds^2 = -U^{-1}\, dt^2 &+ (2l)^2 U (d\bar\psi - 2\cos^2\tfrac12\theta\, d\phi)^2 \\ &+ W(d\theta^2 + \sin^2\theta\, d\phi^2). \end{aligned}$$

In these expressions, l is a constant and U and W are nonvanishing smooth functions of t (on an interval $t_1 < t < t_2$). On each slice $t = t_0$, the manifold resulting from the natural identification on the ϕ coordinate is homeomorphic (indeed, uniformly isomorphic[8]) with the manifolds previously discussed in the case that $(2l)^2 U = W = 1$. If one further imposes a periodicity on $\tilde\psi$ (or $\bar\psi$), one obtains a solid torus. Restricting θ to $\theta \le \frac12\pi$ (respectively, $\theta \ge \frac12\pi$) changes nothing. If one pieces the solid tori together, as required by the coordinate transformation $\bar\psi = \tilde\psi + 2\phi$, the resulting object has been shown[9] to be a 3-sphere. The geometry on the sphere is nonsingular at the poles, e.g., at the north pole, the geometry of a constant ψ surface is given by the line element

$$ds^2 = d\theta^2 + \theta^2\, d\phi^2$$

and identification $(\theta, \phi) = (\theta, \phi + 2\pi)$. Replacing θ by r, one sees that the geometry is smooth.

There is yet another coordinate system obtained by setting $\psi = \tilde{\psi} + \phi$, which gives line element

$$ds^2 = -U^{-1}\, dt^2 + (2l)^2 U(d\psi - \cos\theta\, d\phi)^2 + W(d\theta^2 + \sin^2\theta\, d\phi^2). \quad (21)$$

The natural identification yields nonsingular 2-cylinders for the constant-t and constant-ψ surfaces, or tori if the θ coordinate is also required to be periodic. However, one obtains a coordinate singularity in ψ which becomes a real singularity if a periodicity is imposed on the ψ coordinate. Thus, one sees that there may be more than one way to connect a locally specified geometry into a well-defined manifold and that the apparent naturalness of a given choice may depend crucially on the coordinate system used. The consequences of recognizing this possibility of discovering several choices of manifolds for one line element may be most important with the line element here considered at large values of t. In this region[10] $U(t)$ is negative and ψ becomes timelike, while t becomes spacelike. Also ϕ is spacelike in some places and timelike in others. These observations also hold when the $\tilde{\psi}$ or $\tilde{\psi}$ coordinate replaces ψ. The metric in this region was first considered by Newman, Tamborino, and Unti[11]; it has been called the NUT metric. With the identifications which have, to this point, been imposed on the NUT metric, one obtains an extension of the Taub space. Consequently, one may call the global object Taub–NUT *space*. It has been shown[12] that there are several unsatisfactory—and often unexpected—features possessed by this space from the standpoint of considering it a cosmological model. As it can be shown that these features may largely be traced to the spherical topology on the constant-t 3-surfaces in the NUT ($U < 0$) region, it is comforting to realize that one may connect the manifold in a different manner. Just as in the Taub ($U > 0$) region, the (ψ, θ, ϕ) coordinate system is no longer suggestive of the 3-sphere; however, the changed sign of U makes it less clear what conditions *should* be imposed. In the next section, the odd results obtained from a 3-sphere with $(-, +, +)$ metric are examined; then, the question of how compelling the need is to consider the Taub and the NUT regions part of the same manifold is considered.

4. CAUSALITY SINGULARITIES

It is well known that, in the neighborhood of any 3-slice S through a 4-manifold M, one can create a patch in which the slice is spacelike and the time coordinate is orthogonal to the slice—indeed,[13] one may require that $g_{00} = -1$. One might suspect that, by relaxing the last condition, it would be possible to construct a global time orthogonal coordinate system. Whenever this can be done, there is a very nice causal structure for the manifold. However, if M is compact, no such happy result occurs. Indeed, either the causal properties of M are very bad or there are problems with the spacelike directions which are, perhaps, worse. The following result, similar to ones given by Misner[10] and Avez,[14] demonstrates this assertion.

Theorem 1: Let M be an isochronous n-manifold of signature[15] $n - 2$ and S an $(n - 1)$-manifold which divides M into two disjoint pieces. Let Γ be a simple closed timelike curve which intersects S. Then S is not everywhere spacelike in M.

Proof: Let the disjoint pieces of $M - S$ be P and Q. Let $\mathbf{v}$ be the tangent to Γ in the forward time direction at some point $x \in \Gamma \cap S$. Say $\mathbf{v}$ points from P into Q. Assume S is everywhere spacelike. Construct a continuous field $\mathbf{n}$ of unit normal vectors on S such that $\mathbf{n}(x) \cdot \mathbf{v} > 0$. Then $\mathbf{n}$ is everywhere timelike and points forward. Propagate a continuous field of tangent vectors $\mathbf{t}$ on Γ such that $\mathbf{t}(x) = \mathbf{v}$. Everywhere $\mathbf{t}$ is timelike and forward pointing. Γ must intersect S at some point $y \neq x$ because, if τ is a parameter on Γ, (a) $\tau(x) = 0$ and τ increases in the direction of $\mathbf{v}$ at x and (b) for small positive a, $\Gamma(a) \in Q$ while $\Gamma(-a) \in P$. But, $\mathbf{t}(y) \cdot \mathbf{n}(y) < 0$ since Γ goes from Q to P with increasing τ. This is a contradiction; therefore, S is not everywhere spacelike.∎

Now, assume that one attempts to place a continuous nonvanishing timelike vector field on the manifold. If all attempts fail, as they do on such manifolds as the 2-sphere, there is a causality singularity. Two types of this singularity were discussed in connection with the 2-dimensional cone (depending on which coordinate is timelike). Suppose, to the contrary, that there is a continuous nonvanishing timelike vector field. Then, it has been shown[12] that there is at least one closed timelike curve. The emphasis here has been on the timelike direction, i.e., we have *first* thought of placing a well-behaved timelike vector field on the manifold and only *then* is thought given to the spacelike directions. If this attitude is maintained, one obtains the following result.

Theorem 2: Let M be a compact isochronous differentiable n-manifold which admits a global continuous nonvanishing vector field v. For each metric $\mathbf{g}$ for M such that $\mathbf{v}$ is a timelike Killing field, there is a closed timelike curve through each point of M. Any $(n-1)$-manifold which separates M into two disjoint pieces is not everywhere spacelike.

Proof: Let ϕ be the flow[16] generated by $\mathbf{v}$. The flow ϕ is a map from $M \times R \to M$ such that $\phi_s\phi_t(x) = \phi_{s+t}(x)$ and $\phi_0(x) = x$. Consider an arbitrarily small neighborhood D of x, and define $\phi_t(D)$ as $\{\phi_t(x): x \in D\}$. Choose t such that $\phi_t(D) \cap D = \Phi$. Then, consider the sets $\phi_{nt}(D)$ for all nonnegative n. If all of them could be disjoint, the volume of M would be infinite (since ϕ preserves volumes) and M would not be compact. This contradicts a hypothesis. Therefore, for some m, $k\phi_{mt}(D) \cap \phi_{kt}(D) \neq \Phi$. But, this implies that $\phi_0(D) \cap \phi_{(k-m)t}(D) \neq \Phi$. Therefore, the integral curve Γ through x $(= \{\phi_t(x): t > 0 \text{ for all } t\})$ passes through each neighborhood of x. For sufficiently small D, it is possible to deform Γ so that it remains smooth and timelike. The final assertion follows from the application of Theorem 1.▮

It should be remarked that one can first obtain spacelike $(n-1)$-slices and *then* seek a well-behaved timelike vector field. If this is done, the difficulties are transferred to the timelike directions. The difficulties indicated by Theorem 2 are characteristic of the topology of the manifold as a whole. A corollary to Theorem 2 with application to the NUT metric is that the placing of a global timelike vector field on the 3-sphere precludes the possibility of a spacelike 2-slice. These remarks seem rather paradoxical; consequently, the special case of the 3-sphere is examined in detail.

One has at his disposal theorems demonstrating the existence of a global vector field on the 3-sphere—indeed, of the existence of three such orthonormal fields[17] and, consequently, of the existence of a global field of triads and a global quatratic form[18] of signature +1. The prototype 3-sphere is the set

$$x^2 + y^2 + u^2 + v^2 = 1$$

in 4-dimensional Euclidean space. Since[19] the position vector of a general point on the sphere has components (x, y, u, v), one may obtain three vectors tangent to the sphere by seeking the normals to the position vector. With a little care to get three orthonormal vectors, one obtains

$$\begin{aligned}\mathbf{V}_1 &= (y, -x, v, -u),\\ \mathbf{V}_2 &= (v, u, -y, -x),\\ \mathbf{V}_3 &= (u, -v, -x, y).\end{aligned} \tag{22}$$

Normalization obtains automatically since the original sphere is of unit radius. If an integral curve of field $\mathbf{V}_1$ is denoted by $\gamma(t) = [x(t) \cdots v(t)]$ with $\mathbf{V}_1(t) = \dot{\gamma}(t)$, one finds that

$$\dot{x} = y, \quad \dot{y} = -x \quad \text{and} \quad \dot{u} = v, \quad \dot{v} = -u$$

with solution vector

$$(\alpha \sin(t+\beta), \alpha \cos(t+\beta), \gamma \sin(t+\theta), \gamma \cos(t+\theta)). \tag{23}$$

Also, $\alpha^2 + \gamma^2 = 1$, which suggests setting $\alpha = \sin\phi$ and $\gamma = \cos\phi$. One can choose the point for which $t = 0$ so that $\beta = 0$. Then, using the trigonometric identities for the sine and cosine of a sum, one may write the position vector of a point on γ as

$$\gamma(t) = i \sin t + j \cos t,$$

with

$$\mathbf{i} = (\sin\phi, 0, \cos\phi\cos\theta, -\cos\phi\sin\theta)$$

and

$$\mathbf{j} = (0, \sin\phi, \cos\phi\sin\theta, \cos\phi\cos\theta). \tag{24}$$

Since $\mathbf{i}$ and $\mathbf{j}$ are easily verified to be orthonormal constant vectors, one finds that γ is a great circle of the sphere. An orthogonal mapping of the 4-space onto itself can be constructed which sends the members of $\mathbf{V}_1$ in a 1-to-1 fashion onto the members of $\mathbf{V}_2$ or $\mathbf{V}_3$. Consequently, the integral curves of these fields are great circles also. At each point on the sphere, the elements of the three fields form an orthonormal triad which generates a quadratic differential form. Any of the vector fields, say $\mathbf{V}_1$, may be chosen to be timelike and the other two spacelike. However, there is one bad feature of these integral curves which we now investigate. Any two 2-planes through the origin in 4-space intersect either at the origin only or on a straight line containing the origin. Consequently, any two integral curves cross in two points or none. One could, alternatively, take the 2-surface generated by the integral curves of $\mathbf{V}_2$ and $\mathbf{V}_3$ passing through some $\gamma(t)$ (for each fixed t and one γ) and thus obtain well-behaved spacelike surfaces. But only half of γ may be so used, for that is sufficient to fill the 3-sphere with 2-surfaces (that are, in fact, 2-spheres), and one finds that there are required to be two points on γ at which the forward time direction changes discontinuously. Any global field of vectors may be written as $\mathbf{V} = a\mathbf{V}_1 + b\mathbf{V}_2 + c\mathbf{V}_3$ with a, b, and c continuous scalar fields. Consequently, any partition (or "fibration") of the 3-sphere into simple closed curves—i.e., topological circles—shares the properties just obtained. No complete "open curve"—i.e., continuous image of the real line—can exist in a compact space without getting "infinitely close" to itself.[20] Consequently, basing the concept of timelike direction on such curves causes even more badly behaved results than do circles. Finally, the use of self-intersecting curves is precluded by the lack of a unique direction to assign as timelike at each point of self-intersection.

5. THE $U = 0$ SURFACE IN TAUB SPACE

In Taub space the 3-sphere topology of the constant t surfaces causes no embarrassment. In NUT space, however, the corresponding surfaces contain a timelike direction with awkward consequences as described in the previous section. This suggests that one might wish to perform different identifications in NUT space than in Taub space. It is hard to see how one might justify such an action unless the combined Taub–NUT space were to exhibit singular behavior at the junction. This does happen if the term "singular" is used in the sense of this article. The Taub-space metric may be presented in any of the forms (20) or (21) with

$$U = (t - t_1)(t_2 - t)W^{-1},$$
$$W = t^2 + l^2,$$

$t_2 > t_1$ and l a positive constant. It has been shown[21] that there is an infinite spiraling of geodesics as one approaches the surface $t = t_1$ or t_2. This is suggestive of a conical singularity. The suggestion is made even stronger if one investigates it in more detail. Consider $t \leqslant t_2$, and define

$$A^2 = t_2^2 + l^2,$$
$$B^2 = t_2 - t_1, \quad (25)$$
$$\rho = (2A/B)(t_2 - t)^{\frac{1}{2}}.$$

Then the line element is

$$ds^2 = d\rho^2 - l^2 B^4 A^{-4} \rho^2\, d\phi^2 + A^2(d\theta^2 + \sin^2\theta\, d\phi^2), \quad (26)$$

with only the lowest-order term in each coefficient. To this approximation, there is no difference among the coordinates ψ, $\bar{\psi}$, $\tilde{\psi}$, and ψ^* of Sec. 3. After scaling ψ by the factor $l^{-2}B^{-4}A^4$ and making an obvious coordinate transformation, one can see that (26) is equivalent to (9).

Alternatively, one can use the Geroch technique, modified as proposed in Sec. 2. However, the regions $t = t_2$ and $t = t_1$ were originally analyzed by a general technique developed by the present author. Since the coordinates (20) or (21) are time orthogonal in Taub space, the prescription reduces to the following rule: Reverse the sign of g_{00} and complete the resulting (positive-definite) space. The procedure claims that the boundary points so obtained by this prescription are the correct ones if the manifold can be cut into a finite number of pieces such that either (a) the completed piece is compact and geodesically convex or (b) almost flat in a sense which can be made precise. In the case of Taub space, the region $0 < \psi < \frac{3}{4}\alpha$, $0 < \theta < \frac{3}{4}\pi$ satisfies the requirements, where α is the period of the coordinate. Similarly, a few other patches, obtained in an obvious fashion, suffice to cover the remainder of Taub space. When $U = 0$, there is no term in the line element involving $d\rho$.

All three methods of analyzing $t = t_2$ or t_1 agree that there is a 2-surface at each of $t = t_1, t_2$. Consequently, the periodicity of all the spatial coordinates shows that there is a type of conical singularity, in particular, that of the example of (9).

The first two techniques used to analyze $t = t_1$, t_2 may be applied directly in the NUT region and, again, suggest a conical singularity if ψ is assumed periodic. The general technique requires time orthogonality on each of the patches (although it is not necessary that it be the same system for all patches). However, this has only been achieved approximately. It is hard to be sure that $t = t_1$, t_2—viewed as a subset of the NUT region—is 2 dimensional, for each of the techniques used has some uncertainty. But, since they still agree, one has some confidence. If so, there are conical singularities at $t = t_1, t_2$ in the NUT region also.

If one deletes the periodicity requirement on ψ, it is no longer clear that $t = t_1, t_2$ is 2 dimensional. All that is sure is that $t = t_1, t_2, -a \leq \psi \leq b$ is 2 dimensional for each finite a and b. However, there may be points which formally correspond to $\psi = \infty$. This does happen in the special case for which the NUT metric reduces to the Schwarzschild one.

Whether this happens in general is as yet an open question, but, if it does, one presumes that it will be possible to extend the NUT region into one with the Taub metric, although not with the same connections. This might lead to a truly nonsingular model, since $t = 0$ is not a singularity except for the Schwarzschild special case.

ACKNOWLEDGMENTS

The author would like to thank Professor W. Guillemin and Professor R. O'Neill for their valuable discussions. He also wishes to thank Professor I. Robinson for a critical review of a portion of this article.

* Supported by a State University Research Foundation grant-in-aid.

[1] See R. Geroch [J. Math. Phys. 9, 450 (1968)] for the most advanced techniques to date in this direction and for a review of earlier literature.

[2] This topic is covered in various topology texts; see, e.g., John L. Kelly, *General Topology* (Van Nostrand, Princeton, N.J., 1955), p. 190. A brief discussion is given in Ref. 3.

[3] D. Feinblum, "A New Technique for the Analysis of Singularities," in *Relativity and Gravitation*, C. Kuper and A. Peres, Eds. (Gordon and Breach, in press).
[4] What is described here is a rough version of the techniques of Ref. 1. The most important omission is that only incomplete geodesics are examined, and these only near the (to be added) end point. For general use, the proposed modification would have to be similarly amended, but it has a more serious deficiency, mentioned in the text, which may be harder to remove. It is not at all clear to the present author that completions based on geodesics can be given a completely satisfactory definition.
[5] C. W. Misner, J. Math. Phys. 4, 924 (1963).
[6] A. H. Taub, Ann. Math. 53, 472 (1951).
[7] Reference 4, Eq. 2. The coordinates used here are those of Ref. 4, but the notation corresponds to Ref. 10.
[8] Reference 2, p. 180.
[9] See Ref. 1, Section V.
[10] C. W. Misner and A. H. Taub, Zh. Eksp. Teor. Fiz. 55, 233 (1968) [Sov. Phys. JETP 28, 122 (1969)]. Also, C. W. Misner, "Taub-NUT Space as a Counterexample to Almost Anything" in *Lectures in Applied Mathematics*, J. Ehlers, Ed. (American Mathematical Society, Providence, R.I., 1967), Vol 8.
[11] E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. 4, 915 (1963).
[12] See Refs. 1 and 10.
[13] R. Geroch, J. Math. Phys. 8, 782 (1967); R. Bass and L. Witten, Rev. Mod. Phys. 29, 452 (1957); E. H. Kronheimer and R. Penrose, Proc. Cambridge Phil. Soc. 63, 481 (1967).
[14] A. Avez, Compt. Rend. 254, 3984 (1962).
[15] That is, with one timelike coordinate.
[16] For a definition of a flow on a manifold, see R. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass. 1968). The method of proof presented here was shown to the author by V. Guillemin (private communication).
[17] G. Whithead, Ann. Math. 43, 132 (1942), Theorem 11 and the immediately preceding discussion.
[18] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, N.J. 1951).
[19] The arguments used here were very kindly supplied by R. O'Neill (private communication).
[20] Precisely, the real line does not have an isometric embedding in any compact space, although it may have a 1-to-1 isometric immersion. Immersions are discussed in any text on differentiable manifolds.
[21] See Misner and Taub, Ref. 10.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Stark Effect in Hydrogen Atoms for Nonuniform Fields*

J. D. BEKENSTEIN†‡ AND J. B. KRIEGER
Department of Physics, Polytechnic Institute of Brooklyn, Brooklyn, New York 11201

(Received 5 March 1970)

The correction for the energy eigenvalues of the Schrödinger equation for a hydrogenic atom in a nonuniform field resulting from the inhomogeneity of the field is expressed in terms of expectation values involving the eigenfunctions of the system for a uniform field. Only the first-order terms in the inhomogeneity of the field are retained. An examination of the symmetry of the eigenfunctions for the uniform field, followed by an application of Gauss' law, shows that the correction depends only on one component of the field gradient tensor, regardless of the symmetry of the field, except for states with magnetic quantum number $m = \pm 1$. For the latter states we find the degeneracy is removed provided that the field is not cylindrically symmetric. We evaluate the correction by applying Feynman's theorem to a pair of 1-dimensional eigenvalue equations similar to those obtained in the separation of the uniform field problem in parabolic coordinates. All the necessary eigenvalues are calculated by the WKB method that has been previously employed in obtaining the eigenvalues for the uniform field problem. As the final result we present an expression for the zz component of the quadrupole tensor of the electron labeled according to parabolic quantum numbers. Finally, we discuss the use of this expression in the study of line broadening caused by interatomic interactions (pressure broadening).

I. INTRODUCTION

Recently, a great amount of theoretical work has been done on the Stark effect in hydrogen.[1-3] In all these treatments, however, only the case of a uniform electric field has been treated. The effect of an inhomogeneity in the applied field on the Stark spectrum can be shown, by an order of magnitude estimate, to be unobservable for laboratory fields. On the other hand, the effect of the inhomogeneity in the field caused by the interatomic interaction in the hydrogen gas cannot be neglected.

To be more precise, the shift in energy caused by a gradient in the field is of the order of the product of the field gradient by a typical area of the atom (playing the role of the quadrupole moment). We are here using atomic units (a.u.), i.e., $e = \hbar = m = 1$. A typical laboratory field is 10^6 V/cm or 2×10^{-4} a.u. (1 a.u. $= 5.142 \times 10^9$ V/cm)[4] and can change appreciably over a distance of 0.1 cm or 2×10^7 a.u. (1 a.u. $= 0.53 \times 10^{-8}$ cm). The typical gradient is then of the order of 10^{-11} a.u. while a typical area is n^4 a.u., where n is the principal quantum number of the particular state of the atom. Even for $n = 10$ (a very excited state), the shift in energy will be only 10^{-7} a.u. of energy or about 2.7×10^{-6} eV. By comparison, the fine structure splitting of hydrogen is of the order of 10^{-4} eV so that the observation of the inhomogeneous-field Stark effect is a difficult matter.

[3] D. Feinblum, "A New Technique for the Analysis of Singularities," in *Relativity and Gravitation*, C. Kuper and A. Peres, Eds. (Gordon and Breach, in press).

[4] What is described here is a rough version of the techniques of Ref. 1. The most important omission is that only incomplete geodesics are examined, and these only near the (to be added) end point. For general use, the proposed modification would have to be similarly amended, but it has a more serious deficiency, mentioned in the text, which may be harder to remove. It is not at all clear to the present author that completions based on geodesics can be given a completely satisfactory definition.

[5] C. W. Misner, J. Math. Phys. **4**, 924 (1963).

[6] A. H. Taub, Ann. Math. **53**, 472 (1951).

[7] Reference 4, Eq. 2. The coordinates used here are those of Ref. 4, but the notation corresponds to Ref. 10.

[8] Reference 2, p. 180.

[9] See Ref. 1, Section V.

[10] C. W. Misner and A. H. Taub, Zh. Eksp. Teor. Fiz. **55**, 233 (1968) [Sov. Phys. JETP **28**, 122 (1969)]. Also, C. W. Misner, "Taub-NUT Space as a Counterexample to Almost Anything" in *Lectures in Applied Mathematics*, J. Ehlers, Ed. (American Mathematical Society, Providence, R.I., 1967), Vol 8.

[11] E. Newman, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).

[12] See Refs. 1 and 10.

[13] R. Geroch, J. Math. Phys. **8**, 782 (1967); R. Bass and L. Witten, Rev. Mod. Phys. **29**, 452 (1957); E. H. Kronheimer and R. Penrose, Proc. Cambridge Phil. Soc. **63**, 481 (1967).

[14] A. Avez, Compt. Rend. **254**, 3984 (1962).

[15] That is, with one timelike coordinate.

[16] For a definition of a flow on a manifold, see R. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass. 1968). The method of proof presented here was shown to the author by V. Guillemin (private communication).

[17] G. Whithead, Ann. Math. **43**, 132 (1942), Theorem 11 and the immediately preceding discussion.

[18] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, N.J. 1951).

[19] The arguments used here were very kindly supplied by R. O'Neill (private communication).

[20] Precisely, the real line does not have an isometric embedding in any compact space, although it may have a 1-to-1 isometric immersion. Immersions are discussed in any text on differentiable manifolds.

[21] See Misner and Taub, Ref. 10.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Stark Effect in Hydrogen Atoms for Nonuniform Fields*

J. D. BEKENSTEIN†‡ AND J. B. KRIEGER
Department of Physics, Polytechnic Institute of Brooklyn, Brooklyn, New York 11201

(Received 5 March 1970)

The correction for the energy eigenvalues of the Schrödinger equation for a hydrogenic atom in a nonuniform field resulting from the inhomogeneity of the field is expressed in terms of expectation values involving the eigenfunctions of the system for a uniform field. Only the first-order terms in the inhomogeneity of the field are retained. An examination of the symmetry of the eigenfunctions for the uniform field, followed by an application of Gauss' law, shows that the correction depends only on one component of the field gradient tensor, regardless of the symmetry of the field, except for states with magnetic quantum number $m = \pm 1$. For the latter states we find the degeneracy is removed provided that the field is not cylindrically symmetric. We evaluate the correction by applying Feynman's theorem to a pair of 1-dimensional eigenvalue equations similar to those obtained in the separation of the uniform field problem in parabolic coordinates. All the necessary eigenvalues are calculated by the WKB method that has been previously employed in obtaining the eigenvalues for the uniform field problem. As the final result we present an expression for the zz component of the quadrupole tensor of the electron labeled according to parabolic quantum numbers. Finally, we discuss the use of this expression in the study of line broadening caused by interatomic interactions (pressure broadening).

I. INTRODUCTION

Recently, a great amount of theoretical work has been done on the Stark effect in hydrogen.[1-3] In all these treatments, however, only the case of a uniform electric field has been treated. The effect of an inhomogeneity in the applied field on the Stark spectrum can be shown, by an order of magnitude estimate, to be unobservable for laboratory fields. On the other hand, the effect of the inhomogeneity in the field caused by the interatomic interaction in the hydrogen gas cannot be neglected.

To be more precise, the shift in energy caused by a gradient in the field is of the order of the product of the field gradient by a typical area of the atom (playing the role of the quadrupole moment). We are here using atomic units (a.u.), i.e., $e = \hbar = m = 1$. A typical laboratory field is 10^6 V/cm or 2×10^{-4} a.u. (1 a.u. $= 5.142 \times 10^9$ V/cm)[4] and can change appreciably over a distance of 0.1 cm or 2×10^7 a.u. (1 a.u. $= 0.53 \times 10^{-8}$ cm). The typical gradient is then of the order of 10^{-11} a.u. while a typical area is n^4 a.u., where n is the principal quantum number of the particular state of the atom. Even for $n = 10$ (a very excited state), the shift in energy will be only 10^{-7} a.u. of energy or about 2.7×10^{-6} eV. By comparison, the fine structure splitting of hydrogen is of the order of 10^{-4} eV so that the observation of the inhomogeneous-field Stark effect is a difficult matter.

On the other hand, interatomic fields can be much stronger than 10^{-4} a.u. and can vary considerably over distances of a few a.u., so that the field gradients are much larger than those associated with laboratory fields. It follows that the effect of interatomic interactions on the spectrum may be observable. It is thus physically meaningful to derive the first correction term for the effect of an inhomogeneous field on the spectrum.

In this work we first construct the Schrödinger equation for the nonrelativistic problem of a hydrogen atom in an electric field including the first-order terms of the inhomogeneity of the field. Using nondegenerate perturbation theory, we express the correction for the effect of the inhomogeneity in terms of expectation values of the inhomogeneity in the external potential taken with respect to eigenfunctions of the system for the uniform field problem. An examination of the symmetry of these eigenfunctions, together with the use of Gauss's law for the external field, shows that the correction can be expressed completely in terms of the zz component of the quadrupole tensor of the electron and the zz component of the field gradient tensor, regardless of the symmetry of the field, except for states with magnetic quantum number $m = \pm 1$.

We rewrite the quadrupole moment in terms of simple integrals in parabolic coordinates. Instead of evaluating these directly by using explicit expressions for the eigenfunctions, we choose to evaluate them by repeatedly applying Feynman's theorem[5] to a pair of equations which are slight generalizations of those obtained from the separation of the uniform field problem in parabolic coordinates. We evaluate the eigenvalues of the equations by the WKB method that we used previously in obtaining the Stark-effect eigenvalues for uniform fields.[3] (Reference 3 will henceforth be referred to as I.) In this manner, we obtain a formula for the zz component of the quadrupole tensor for states labeled with parabolic quantum numbers.

Finally, using the above formula, we discuss the calculation of the broadening of Stark spectrum lines caused by the interaction of hydrogen atoms with their neighbors in the gas and give numerical estimates for this effect (pressure broadening).

II. CORRECTIONS FOR THE FIELD INHOMOGENEITY

Let ϕ be the electrostatic potential describing the field in which a hydrogenic atom of nuclear charge Ze, nuclear mass M, and electron mass m is immersed. We choose the arbitrary constant of ϕ in such a way that ϕ vanishes at the center of mass of the atom. If $\mathbf{r}'$ is the position vector in the center-of-mass system, we can expand $\phi(\mathbf{r}')$ as follows:

$$\phi(\mathbf{r}') = \nabla\phi(0)\cdot\mathbf{r}' + \tfrac{1}{2}\sum_{i,j}\frac{\partial^2\phi(0)}{\partial x_i'\partial x_j'}x_i'x_j' + \cdots, \quad (1)$$

where x_j' is a Cartesian component of $\mathbf{r}'$. We choose the z' axis of our coordinate system in the direction of $\nabla\phi(0)$. Then,

$$\phi(\mathbf{r}') = -\mathcal{E}z' - \tfrac{1}{2}\sum_{i,j}\frac{\partial\mathcal{E}_j}{\partial x_i'}x_i'x_j' + \cdots, \quad (2)$$

where $\boldsymbol{\mathcal{E}} = -\nabla\phi(0) = (0, 0, \mathcal{E})$.

We now introduce $\mathbf{r}$, the vector from the nucleus to the electron, so that if the subscript e denotes the electron and the subscript n the nucleus,

$$\mathbf{r}_e' = \frac{M}{m+M}\mathbf{r} \quad \text{and} \quad \mathbf{r}_n' = -\frac{m}{m+M}\mathbf{r}$$

give the positions of the nucleus and the electron with respect to the center of mass. The potential energy for the problem is (with $e = +|e|$)

$$V = e\mathcal{E}(z_e' - Zz_n') + \tfrac{1}{2}e\sum_{i,j}\frac{\partial\mathcal{E}_j}{\partial x_i'}(x_{ej}'x_{ei}' - Zx_{nj}'x_{ni}') - \frac{Ze^2}{|\mathbf{r}_e' - \mathbf{r}_n'|},$$

which can be rewritten as

$$V = -eFz - \tfrac{1}{2}e\sum_{i,j}\nabla_{ij}x_ix_j - \frac{Ze^2}{r}, \quad (3)$$

with

$$F \equiv -\mathcal{E}\left(1 + \frac{m}{m+M}(Z-1)\right), \quad (4)$$

$$\nabla_{ij} \equiv -\frac{\partial\mathcal{E}_j}{\partial x_i'}\left(1 - \frac{2mM + (Z+1)m^2}{(m+M)^2}\right). \quad (5)$$

Here z is the third component of $\mathbf{r}$. The above definition of F agrees with that used in I.

We write the kinetic energy in center-of-mass coordinates as $-(\hbar^2/2\mu)\nabla^2$ with $\mu = mM/(m+M)$, thus neglecting relativistic corrections. Again, neglecting spin effects, we have the eigenvalue equation

$$-\frac{\hbar^2}{2\mu}\nabla^2\psi - \frac{Ze^2}{r}\psi - \left(eFz + \tfrac{1}{2}e\sum_{i,j}\nabla_{ij}x_ix_j\right)\psi = E\psi. \quad (6)$$

For a given quantum state, E is a function of F and ∇_{ij}. For small ∇_{ij} we have, using nondegenerate perturbation theory,

$$E(F, \nabla_{ij}) = E(F, 0) + \sum_{i,j}\langle\psi(F,0)| -\tfrac{1}{2}ex_ix_j|\psi(F,0)\rangle\nabla_{ij}. \quad (7)$$

Equation (7) is valid provided that the matrix elements of the perturbation connecting originally degenerate states are all zero. It is well known (see I, for example) that the only degeneracies of the uniform field problem are those with respect to the sign of the magnetic quantum number m (eigenfunctions with $\pm m$ are degenerate). In Sec. III we show that for $m^2 \neq 1$ all matrix elements of the perturbation linking these degenerate states are zero. The case $m^2 = 1$ is treated in Sec. IV.

III. SIMPLIFICATION OF CORRECTION TERMS ($m^2 \neq 1$)

In I we separated Eq. (6) with $\nabla_{ij} = 0$ in parabolic coordinates (ζ, η, φ) defined, in terms of spherical polar coordinates (r, θ, φ), by

$$\zeta = r(1 + \cos\theta), \quad \eta = r(1 - \cos\theta), \quad \varphi = \varphi, \quad r = \tfrac{1}{2}(\zeta + \eta) \tag{8}$$

or, in terms of Cartesian coordinates, by

$$x = (\zeta\eta)^{\frac{1}{2}} \cos\varphi, \quad y = (\zeta\eta)^{\frac{1}{2}} \sin\varphi, \quad z = \tfrac{1}{2}(\zeta - \eta). \tag{9}$$

By choosing the eigenfunction of the form

$$\psi(F, 0) = \frac{A}{(2\pi)^{\frac{1}{2}}} \frac{f(\zeta)}{(\zeta)^{\frac{1}{2}}} \frac{g(\eta)}{(\eta)^{\frac{1}{2}}} e^{im\varphi}, \quad m = 0, \pm 2, \pm 3, \cdots, \tag{10}$$

we obtain as in I the equations satisfied by f and g, i.e.,

$$\begin{aligned} -\frac{\hbar^2}{2\mu}\frac{d^2 f}{d\zeta^2} - \left(\frac{eF}{8}\zeta + \frac{Ze^2 + \beta}{4\zeta} - \frac{m^2 - 1}{4\zeta^2}\frac{\hbar^2}{2\mu}\right) f = \frac{E}{4} f, \\ -\frac{\hbar^2}{2\mu}\frac{d^2 f}{d\eta^2} - \left(-\frac{eF}{8}\eta + \frac{Ze^2 - \beta}{4\eta} - \frac{m^2 - 1}{4\eta^2}\frac{\hbar^2}{2\mu}\right) g = \frac{E}{4} g. \end{aligned} \tag{11}$$

The probability density $\psi^*\psi$ does not depend on φ and is thus cylindrically symmetric about the z axis. It is clear from Eqs. (9) and (10) that matrix elements of xy, xz, yz, x^2, y^2, and z^2 connecting originally degenerate states will vanish when the integration over φ is performed for states with $m^2 \neq 1$. Thus, nondegenerate perturbation theory is valid in this case. Furthermore, from Eqs. (9) and (10) we see that expectation values of xz, yz, and xy vanish when the φ integration is performed and, hence, only $\langle x^2\rangle$, $\langle y^2\rangle$, and $\langle z^2\rangle$ are of interest. Finally, we note that, since ψ is factorable as given in Eq. (10), we have

$$\langle x^2\rangle = \langle\zeta\eta\rangle\langle\cos^2\varphi\rangle = \tfrac{1}{2}\langle\zeta\eta\rangle = \langle\zeta\eta\rangle\langle\sin^2\varphi\rangle = \langle y^2\rangle. \tag{12}$$

It follows from Eqs. (6), (7), and (12) that

$$E(F, \nabla_{ij}) = E(F, 0) - \tfrac{1}{2}e[\langle z^2\rangle\nabla_{zz} + \langle x^2\rangle(\nabla_{yy} + \nabla_{xx})],$$

which is further simplified by noting the Gauss' law, i.e., $\nabla \cdot \mathcal{E} = 0$ for the external field, $\mathcal{E}$ gives, according to Eq. (5),

$$\nabla_{xx} + \nabla_{yy} = -\nabla_{zz} \tag{13}$$

so that the correction for the energy to first order in the gradient is

$$\Delta E = -\tfrac{1}{2}e\langle z^2 - x^2\rangle\nabla_{zz} = -\tfrac{1}{4}e\langle 3z^2 - r^2\rangle\nabla_{zz}, \tag{14}$$

where we have used

$$\langle x^2\rangle = \tfrac{1}{2}\langle x^2 + y^2\rangle = \tfrac{1}{2}\langle r^2 - z^2\rangle.$$

Equation (13) is identical to that for the classical interaction energy of a quadrupole moment with a cylindrically symmetric field.[6] It must be noted, however, that nowhere in our derivation did we assume the external field to be cylindrically symmetric. The simplification leading to Eq. (14) results from the symmetry of the eigenfunctions, not from the symmetry of the field.

We now proceed to express $\langle z^2 - x^2\rangle$ in terms of simple integrals. Noting from Eq. (11) that both f and g may be chosen real, we define

$$\begin{aligned} J &= \int_0^\infty f^2\zeta\, d\zeta, & J' &= \int_0^\infty g^2\eta\, d\eta, \\ K &= \int_0^\infty f^2\zeta^2\, d\zeta, & K' &= \int_0^\infty g^2\eta^2\, d\eta, \\ L &= \int_0^\infty \frac{f^2}{\zeta}\, d\zeta, & L' &= \int_0^\infty \frac{g^2}{\eta}\, d\eta, \end{aligned} \tag{15}$$

where f and g are separately normalized, i.e.,

$$\int_0^\infty f^2(\zeta)\, d\zeta = \int_0^\infty g^2(\eta)\, d\eta = 1.$$

We also note that the volume element in parabolic coordinates is

$$d^3r = \tfrac{1}{4}(\zeta + \eta)\, d\zeta\, d\eta\, d\varphi.$$

Thus from Eq. (12) we have

$$\langle x^2\rangle = \tfrac{1}{2}\langle\zeta\eta\rangle = \tfrac{1}{8}A^2\int_0^{2\pi}\frac{d\varphi}{2\pi}\int_0^\infty\int_0^\infty(\zeta + \eta)f^2g^2\, d\zeta\, d\eta,$$

which from Eq. (15) becomes

$$\langle x^2\rangle = \tfrac{1}{2}\langle\zeta\eta\rangle = \tfrac{1}{8}A^2(J + J'). \tag{16}$$

Similarly,

$$\langle z^2\rangle = \tfrac{1}{4}(\langle\zeta^2\rangle + \langle\eta^2\rangle - 2\langle\zeta\eta\rangle).$$

Again, from the definitions,

$$\langle \zeta^2 \rangle = \tfrac{1}{4}A^2(KL' + J),$$
$$\langle \eta^2 \rangle = \tfrac{1}{4}A^2(K'L + J'),$$

so that

$$\langle z^2 - x^2 \rangle = \tfrac{1}{16}A^2(KL' + K'L - 3J - 3J').$$

The normalization integral is

$$\int \psi^* \psi \, d^3r = \tfrac{1}{4}A^2(L' + L) = 1.$$

By using this, our final result is

$$\langle z^2 - x^2 \rangle = \frac{1}{4}\frac{(KL' + K'L - 3J - 3J')}{L + L'}. \quad (17)$$

IV. EIGENVALUES FOR THE CASE $m^2 = 1$

The treatment of the preceding section is invalid for the case $m^2 = 1$ because for this case the perturbation in Eq. (6) is not diagonal with respect to the eigenfunctions given by Eq. (10). We thus take the following linear combinations of the eigenfunctions in Eq. (10) for $m^2 = 1$:

$$\psi_\pm = \frac{A}{\sqrt{\pi}} \frac{f(\zeta)}{\sqrt{\zeta}} \frac{g(\eta)}{\sqrt{\eta}} \begin{Bmatrix} \cos \varphi \\ \sin \varphi \end{Bmatrix}. \quad (18)$$

We also rotate our coordinate system about the z axis (direction of the field) so that the quadrupole tensor of the field becomes diagonal in the (x, y) plane, i.e., $\nabla_{xy} = 0$.

It is then easily verified that $\langle + |z^2| - \rangle$, $\langle + |xz| - \rangle$, $\langle + |yz| - \rangle$, $\langle + |x^2| - \rangle$, and $\langle + |y^2| - \rangle$ all vanish once the integration over φ is performed [see Eq. (9)]. Hence the perturbation is diagonal with respect to the degenerate states given by Eq. (18), and thus nondegenerate perturbation theory is appropriate.

It also follows that $\langle xz \rangle_\pm = \langle yz \rangle_\pm = 0$; since $\nabla_{xy} = 0$, we have, from Eq. (7),

$$\begin{aligned}\Delta E &= -\tfrac{1}{2}e(\langle z^2 \rangle \nabla_{zz} + \langle x^2 \rangle \nabla_{xx} + \langle y^2 \rangle \nabla_{yy}) \\ &= -\tfrac{1}{2}[\langle z^2 \rangle \nabla_{zz} + \tfrac{1}{2}\langle x^2 + y^2 \rangle (\nabla_{xx} + \nabla_{yy}) \\ &\qquad + \tfrac{1}{2}\langle x^2 - y^2 \rangle (\nabla_{xx} - \nabla_{yy})].\end{aligned}$$

Using Gauss' law (13), we have

$$\begin{aligned}\Delta E = -\tfrac{1}{2}e[\langle z^2 - \tfrac{1}{2}(x^2 + y^2) \rangle \nabla_{zz} \\ + \tfrac{1}{2}\langle x^2 - y^2 \rangle (\nabla_{xx} - \nabla_{yy})]\end{aligned}$$

or

$$\Delta E = -\tfrac{1}{4}e[\langle 3z^2 - r^2 \rangle \nabla_{zz} + \langle x^2 - y^2 \rangle (\nabla_{xx} - \nabla_{yy})]. \quad (19)$$

The first term in Eq. (19) is identical to expression (14) for the case $m^2 \neq 1$. The second term breaks the degeneracy between the plus and minus states. That this is so is seen from the following calculations:

$$\begin{aligned}\langle x^2 - y^2 \rangle_+ &= \langle \zeta\eta \rangle \int_0^{2\pi} \frac{(\cos^2 \phi - \sin^2 \phi)}{\pi} \cos^2 \phi \, d\phi \\ &= \langle \zeta\eta \rangle \int_0^{2\pi} \cos 2\phi \frac{(1 + \cos 2\phi)}{2\pi} d\phi = \tfrac{1}{2}\langle \zeta\eta \rangle.\end{aligned}$$

On the other hand,

$$\begin{aligned}\langle x^2 - y^2 \rangle_- &= \langle \zeta\eta \rangle \int_0^{2\pi} \frac{(\cos^2 \phi - \sin^2 \phi)}{\pi} \sin^2 \phi \, d\phi \\ &= \langle \zeta\eta \rangle \int_0^{2\pi} \cos 2\phi \frac{(1 - \cos 2\phi)}{2\pi} = -\tfrac{1}{2}\langle \zeta\eta \rangle.\end{aligned}$$

This last result is equal and opposite that for the plus state. Hence

$$\Delta E_\pm = -\tfrac{1}{4}e[Q\nabla_{zz} \pm P(\nabla_{xx} - \nabla_{yy})], \quad (20)$$

where $Q = \langle 3z^2 - r^2 \rangle$ is the zz component of the quadrupole moment tensor for the given state and $P = \tfrac{1}{2}\langle \zeta\eta \rangle$.

Both Q and P are averages of quantities independent of φ and because the term in the eigenfunctions containing φ [both those given by Eq. (10) and Eq. (18)] is normalized, their values are the same no matter which type of eigenfunction is used to compute them. Thus, by comparing Eq. (14) with Eq. (17), we have

$$Q = \frac{1}{2}\frac{(KL' + K'L - 3J - 3J')}{L + L'} \quad (21)$$

and, from Eq. (16) and the normalization integral,

$$P = \frac{J + J'}{2(L + L')}. \quad (22)$$

V. EVALUATION OF INTEGRALS

The terms in Eqs. (21) and (22) could be evaluated by using explicit expressions for $f(\zeta)$ and $g(\eta)$ and performing the required integrations. This is very difficult to do for a general state. Instead, we shall evaluate these terms by making use of Feynman's theorem which states that, if $H(\lambda)$ is an Hermitian operator which is a continuous function of the parameter λ and if $\psi(\lambda)$ is one of its eigenfunctions with eigenvalue $E(\lambda)$, then

$$\frac{dE(\lambda)}{d\lambda} = \langle \psi(\lambda) | \frac{\partial H(\lambda)}{\partial \lambda} | \psi(\lambda) \rangle.$$

Thus we consider the following generalization of the first of Eqs. (11):

$$-\frac{\hbar^2}{2\mu}\frac{d^2 f}{d\zeta^2} - \left(eG\zeta^2 + \frac{eF}{8}\zeta + \frac{Ze^2 + \beta}{4\zeta} - \frac{m^2 - 1}{4\zeta^2}\frac{\hbar^2}{2\mu}\right) f = \frac{E_\pm}{4} f, \quad (23)$$

subject to the condition

$$\int_0^\infty f^2(\zeta)\,d\zeta = 1.$$

Equation (23) is an eigenvalue equation for f with eigenvalue $\frac{1}{4}E_+$ which goes over into the first Eq. (11) as $G \to 0$. Applying Feynman's theorem to it, we have

$$\begin{aligned} J &= \lim_{G\to 0} -\frac{2}{e}\frac{\partial E_+}{\partial F}, \\ K &= \lim_{G\to 0} -\frac{1}{4e}\frac{\partial E_+}{\partial G}, \\ L &= \lim_{G\to 0} -\frac{\partial E_+}{\partial \beta}. \end{aligned} \tag{24}$$

In an analogous fashion, we generalize the second of Eqs. (11) to

$$-\frac{\hbar^2}{2\mu}\frac{d^2g}{d\eta^2} - \left(eG\eta^2 - \frac{eF}{8}\eta + \frac{Ze^2-\beta}{4\eta} - \frac{m^2-1}{4\zeta^2}\frac{\hbar^2}{2\mu}\right)g = \frac{E_-}{4}g \tag{25}$$

and obtain

$$\begin{aligned} J' &= \lim_{G\to 0} \frac{2}{e}\frac{\partial E_-}{\partial F}, \\ K' &= \lim_{G\to 0} -\frac{1}{4e}\frac{\partial E_-}{\partial G}, \\ L' &= \lim_{G\to 0} \frac{\partial E_-}{\partial \beta}. \end{aligned} \tag{26}$$

In taking the derivative, β must be treated as an independent variable. We thus only need to solve for the eigenvalues of Eqs. (23) and (25) to first order in G. In I we obtained the eigenvalues for the more restricted problem given by Eqs. (11). We presented there a generalization of Dunham's[7] WKB quantization rule for Eqs. (11) which we can immediately adapt for our problem. We thus have

$$\begin{aligned} &\oint R\,d\zeta - \frac{\hbar^2}{64\mu}\oint \frac{[d(\zeta^2R^2)/d\zeta]^2}{R^5}\frac{d\zeta}{\zeta^4} \\ &\quad - \frac{\hbar^4}{8192\mu^2}\oint\left\{\frac{49[d(\zeta^2R^2)/d\zeta]^4}{R^{11}} - \frac{16\zeta[d(\zeta^2R^2)/d\zeta](\zeta d/d\zeta)^3(\zeta^2R^2)}{R^7}\right\}\frac{d\zeta}{\zeta^8} \\ &\quad + O(\hbar^6) = \frac{(n_1+\frac{1}{2})\hbar 2\pi}{\sqrt{2\mu}}, \end{aligned} \tag{27}$$

with

$$n_1 = 0, 1, \cdots$$

and

$$R^2 = eG\zeta^2 + \frac{eF}{8}\zeta + \frac{E_\pm}{4} + \frac{Ze^2+\beta}{4\zeta} - \frac{m^2}{4\zeta^2}\frac{\hbar^2}{2\mu},$$

where the contour integrals are to encircle the region on the positive real ζ axis for which $R^2 > 0$, which is nearest to the origin (see I). For the η equation the same rule applies except for the replacements

$$F \to -F, \quad \beta \to -\beta, \quad n_1 \to n_2, \quad E_+ \to E_-.$$

The quantization rule given by Eq. (27) may be evaluated by expanding the integrals in powers of F and G, but, because the limit $G \to 0$ is to be taken at the end, we do not need terms in G of order higher than the first. In addition, for simplicity, we shall keep only terms in the first order in F. It is possible to prove that the quantization rule will give the eigenvalues correctly to these orders.[8] After evaluating the integrals resulting from the expansion, we obtain

$$\begin{aligned} &\frac{\pi}{2^{\frac{1}{2}}}(Ze^2 \pm \beta) \\ &\quad \pm \frac{eF}{E_\pm^2}\left(\frac{3}{22}\pi 2^{\frac{1}{2}}(Ze^2\pm\beta)^2 + (m^2-1)\frac{1}{16}\pi\frac{\hbar^2}{\mu}2^{\frac{1}{2}}\right) \\ &\quad - \frac{eG}{E_\pm^3}\left(\frac{105}{24}\pi 2^{\frac{1}{2}}(Ze^2\pm\beta)^3 + \frac{1}{4}(3m^2-7)\pi\frac{\hbar^2}{\mu}2^{\frac{1}{2}}(Ze^2\pm\beta)E_\pm\right) \\ &= \left(\begin{matrix} n_1 \\ n_2 \end{matrix} + \frac{1}{2}|m| + \frac{1}{2}\right)\frac{\hbar 2\pi}{(2\mu)^{\frac{1}{2}}}(-2E_\pm)^{\frac{1}{2}}, \end{aligned} \tag{28}$$

with the upper (lower) symbols referring to the $\zeta(\eta)$ equation.

In Eq. (28) we substitute $E_\pm = E_0 + E_1eF + E_2eG$, separate different orders in F and G, and obtain E_0, E_1, and E_2. The result for the ζ equation is

$$\begin{aligned} E_+ = &-\frac{(Ze^2+\beta)^2\mu}{2n^2\hbar^2} \\ &- \frac{1}{4}\frac{\hbar^2}{(Ze^2+\beta)\mu}(3n'^2 - m^2 + 1)eF \\ &- \frac{\hbar^4}{(Ze^2+\beta)^2\mu^2}n'^2(10n'^2 - 6m^2 + 14)eG, \end{aligned} \tag{29}$$

with $n' = 2n_1 + |m| + 1$. For the η equation, replace E_- for E_+, $-F$ for F, $-\beta$ for β, and n' by $n'' \equiv 2n_2 + |m| + 1$. If $\alpha_\pm \equiv (Ze^2 \pm \beta)\mu/\hbar^2$, we obtain, by combining Eqs. (24), (26), and (29),

$$\begin{aligned} J &= (1/2\alpha_+)(3n'^2 - m^2 + 1), \\ J' &= (1/2\alpha_-)(3n''^2 - m^2 + 1), \\ L &= \alpha_+/n'^2, \\ L' &= \alpha_-/n''^2, \\ K &= (1/2\alpha_+^2)n'^2(5n'^2 - 3m^2 + 7), \\ K' &= (1/2\alpha_-^2)n''^2(5n''^2 - 3m^2 + 7). \end{aligned}$$

In letting $G \to 0$ in Eqs. (24) and (26), Eqs. (23) and (25) go over into Eqs. (11) and $E_\pm \to E(F, 0)$. Therefore, in $\alpha_\pm$ we must substitute the value of β resulting from the simultaneous solution of Eqs. (11). We obtained β in I; here we keep only its zeroth-order value because we seek P and Q to only this order in F. Thus $\beta = Ze^2(n_1 - n_2)/n$.

Substituting the above results in Eqs. (20), (21), and (22), multiplying numerator and denominator of the resulting expressions by $n'n''$, and simplifying with the use of the identities

$$n = n'' + (n_1 - n_2) = n' + (n_2 - n_1),$$

we obtain for the general m case

$$\Delta E = -\tfrac{1}{4}eQ\nabla_{zz} - \tfrac{1}{4}eP(\nabla_{xx} - \nabla_{yy})\epsilon(m), \quad (30)$$

with

$$\begin{aligned} Q &\equiv \langle 3z^2 - r^2 \rangle \\ &= \tfrac{1}{8}a_0^2(5nn'^3 + 5nn''^3 + 8n^2 - 18n^2n'n'') + O(F), \\ P &\equiv \tfrac{1}{2}\langle \zeta\eta \rangle = \tfrac{3}{4}a_0^2n^2n'n'' + O(F), \\ \epsilon(m) &= 1, \quad m = 1, \\ &= 0, \quad m^2 \neq 1, \\ &= -1, \quad m = -1, \end{aligned}$$

and $a_0 = \hbar^2/Ze^2\mu$, the Bohr radius for the atom with nuclear charge Ze. It may be seen that for the ground state ($n = n' = n'' = 1$) $\Delta E = 0$ if the $O(F)$ terms are neglected. This was expected since the ground state is nondegenerate (spin degeneracy aside) and spherically symmetric in any coordinate system.

We also see from Eq. (30) that the splitting for $m^2 = 1$ is $\tfrac{3}{8}ea_0^2n^2n'n''(\nabla_{xx} - \nabla_{yy})$. This splitting (if measurable) serves as a measurement of the departure of the field from cylindrical symmetry. We recall that the x and y directions are not arbitrary but have been chosen so that $\nabla_{xy} = 0$. Hence $\nabla_{xx} - \nabla_{yy}$ is uniquely determined by the structure of the field. For a cylindrically symmetric field, $\nabla_{xx} - \nabla_{yy} = 0$; hence, if it is nonzero, its magnitude, as determined from the splitting of the $m^2 = 1$ states, measures the asymmetry of the field.

It may be noted that nowhere have we used explicit expressions for the eigenfunctions of the problem. The above calculations illustrate the power of Feynman's theorem for calculating expectation values when accompanied by some independent method for calculating the necessary eigenvalues.

VI. LINE BROADENING

In this section we shall give an estimate of the line broadening of the Stark spectrum which is caused by interatomic interactions. We consider here only the case of a gas made up mostly of hydrogen atoms in their ground state. This is a possible situation since the binding energy of the H_2 molecule is smaller than the energy difference between the ground and first excited states of the hydrogen atom. Thus over a certain range of temperature the situation considered will occur.

Aside from the natural broadening, there is a contribution to the intrinsic broadening of lines caused by collisions of the excited atoms with other atoms in the gas. This effect, called collision or pressure broadening, has been studied in detail.[9] We expect that collisions causing broadening occur if an excited atom in a quantum state characterized by the principal quantum number n and another unexcited atom approach to within a distance of n^2a_0 (characteristic size of the excited atom). If the closest distance of approach is much larger than this, the radiating atom will not be disrupted and no intrinsic line broadening will occur. However, as we shall see, an apparent broadening of the line will still appear in such a case.

If the gas is exposed to a uniform electric field $\mathcal{E}$, an electric dipole moment is induced in each atom which then becomes the source of a perturbing field. The atoms are moving with respect to each other, but, because their speed is in all cases small compared to the speed of the electron in the excited atom, the adiabatic approximation is valid. Accordingly, the excited atom will remain in the same quantum state, but its energy will shift back and forth as the perturbing atoms fly by.

A detecting apparatus, a photographic plate for example, will not resolve the oscillation of a spectral line and so will show a line broadening. But this broadening is only apparent; the lines are not truly broadened by the perturbation.

We now proceed to calculate the effect quantitatively. The energy shift of an H atom in the ground state because of a uniform field $\mathcal{E}$ (in the z direction) is (see I, for example)

$$\Delta E = -\frac{9}{4}\frac{\hbar^6}{e^6\mu^3}\mathcal{E}^2 + O(\mathcal{E}^4).$$

Thus its dipole moment along the field is

$$p = \frac{9}{2}\frac{\hbar^6}{e^6\mu^3}\mathcal{E} = \frac{9}{2}a_0^3\mathcal{E}. \quad (31)$$

If this atom is at position $\mathbf{r}_0$, the potential due to it at $\mathbf{r}$ will be

$$\phi = p(z - z_0)/|\mathbf{r} - \mathbf{r}_0|^3.$$

The z component of the field and the zz component of the gradient caused by it are

$$\mathcal{E}_z = -\frac{p}{|\mathbf{r}-\mathbf{r}_0|^3} + \frac{3p(z-z_0)^2}{|\mathbf{r}-\mathbf{r}_0|^5},$$

$$\frac{\partial \mathcal{E}_z}{\partial z} = \frac{9p(z-z_0)}{|\mathbf{r}-\mathbf{r}_0|^5} - \frac{15p(z-z_0)^3}{|\mathbf{r}-\mathbf{r}_0|^7},$$

with

$$\frac{\partial \mathcal{E}_x}{\partial x} = \frac{\partial \mathcal{E}_y}{\partial y}.$$

For a fixed $|\mathbf{r}-\mathbf{r}_0|$, we have

$$-\frac{p}{|\mathbf{r}-\mathbf{r}_0|^3} < \mathcal{E}_z < \frac{2p}{|\mathbf{r}-\mathbf{r}_0|^3}, \tag{32}$$

$$-\frac{6p}{|\mathbf{r}-\mathbf{r}_0|^4} < \frac{\partial \mathcal{E}_z}{\partial z} < \frac{6p}{|\mathbf{r}-\mathbf{r}_0|^4}. \tag{33}$$

We can take the mean deviation of $\mathcal{E}_z$ and $\partial\mathcal{E}_z/\partial z$ from their averages to be half of the above ranges.

To obtain an upper bound for the apparent broadening, we use Eqs. (32) and (33) in the expression for the energy shift in an inhomogeneous field to first order in the field and first order in the field gradient [see I and Eq. (30)]:

$$\Delta E = -\frac{3}{2}\frac{\hbar^2}{e\mu} n(n_1-n_2)F - \frac{1}{4} eQ\nabla_{zz}.$$

The apparent broadening results from replacing F above by $\frac{3}{2}p/|\mathbf{r}-\mathbf{r}_0|^3$, ∇_{zz} by $6p/|\mathbf{r}-\mathbf{r}_0|^4$, and estimating $|\mathbf{r}-\mathbf{r}_0|$ as $\rho^{-\frac{1}{3}}$ where ρ is the number of atoms per unit volume. Thus the broadening is

$$\delta E \simeq \tfrac{81}{8} e a_0^4 n\,|n_1-n_2|\,\mathcal{E}\rho + \tfrac{27}{32} e a_0^5(5nn'^5 + 5nn''^3 + 8n^2 - 18n^2n'n'')\mathcal{E}\rho^{\frac{4}{3}}, \tag{34}$$

where we have used Eqs. (30), (31), and the definition $a_0 = \hbar^2/e^2\mu$ to simplify the results. Since the kind of broadening that we are considering depends on the polarization of the atoms in the gas, we find that it vanishes when the applied field $\mathcal{E}$ vanishes.

For comparison we estimate the collision broadening. The mean free path of the excited atom is, from kinetic theory, $\lambda = (2^{\frac{1}{2}}\rho\sigma)^{-1}$ with σ representing the collision cross section, which we take to be $n^4a_0^2$ (characteristic area of the excited atom). The mean speed of the excited atom is

$$\bar{v} = \left(\frac{8}{\pi}\frac{kT}{M}\right)^{\frac{1}{2}}$$

with M being the mass of the H atom. Then the mean free time between collisions is $\tau = \lambda/\bar{v}$. Thus the collision broadening of the energy is $\delta E = \hbar/\tau$ or

$$\delta E = 4\hbar n^4 a_0^2 \rho (kT/\pi M)^{\frac{1}{2}}. \tag{35}$$

To compare Eqs. (34) and (35), we go over to atomic units for which $e = \hbar = a_0 = \mu = 1$. Then ρ is in units of (Bohr radii)$^{-3}$, $\mathcal{E}$ in units of 5.142×10^9 V/cm, and energy in units of 27.2 eV. Thus $\mathcal{E} \ll 1$, $M = 1836$, $kT \leq 1$ ($kT = 1$ eV for 12,000°K), and $\rho \ll 1$ in all possible cases. By choosing specific values for kT and n, it is easy to show that δE given by Eq. (35) is usually much larger than that given by Eq. (34). For example, for $T = 12{,}000$°K and $n = 5$, $n_1 = 0$, $n_2 = 4$, and $m = 0$ (the state from which the line π 18 originates, which exists only for $\mathcal{E} < 10^6$ V/cm),[10] we find that for the highest possible field the collision broadening is 5.95ρ while the apparent broadening is $0.04\rho + 2.38\rho^{\frac{4}{3}}$. Since ρ has to be very small compared to 1 (the total Stark shift of this state is less than 0.02 a.u., and the broadening must be much smaller than this), we see that the apparent broadening is small compared to the collision broadening. The same applies for other states.

For highly excited states, the bound electron may be moving so slowly that the adiabatic approximation is no longer valid, i.e., the Stark field may be changing appreciably during one atomic period. For such states the Stark effect will tend to average out and give a smaller shift and less broadening than calculated here. Thus, our final conclusion is that only actual collisions cause observable line broadening in the gas considered here.

* Supported in part by National Aeronautics and Space Agency Grant NsG 589 and by National Science Foundation Summer Undergraduate Research Grant No. GY-5727.

† Supported by a National Science Foundation Graduate Traineeship.

‡ Present address: Physics Department, Princeton University, Princeton, New Jersey.

[1] L. B. Mendelsohn, Phys. Rev. **176**, 90 (1968).

[2] M. H. Alexander, Phys. Rev. **178**, 34 (1969).

[3] J. D. Bekenstein and J. B. Krieger, Phys. Rev. **188**, 130 (1969).

[4] Atomic units (a.u.) are those for which the charge and mass of the electron and $\hbar$ are 1. The unit of length is the Bohr radius $\hbar^2/me^2 = 5.3 \times 10^{-9}$ cm and the unit of field is the field 1 Bohr radius away from a unit charge, i.e., $-e^5m^2/\hbar^4 = 5.142 \times 10^9$ V/cm. The unit of energy is twice the ground state energy of hydrogen or $me^4/\hbar^2 = 27.2$ eV.

[5] W. Pauli, *Handbuck der Physik, Vol. V: Principles of Quantum Theory*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1958), Part I, p. 83.

[6] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1966), p. 128.

[7] J. L. Dunham, Phys. Rev. **41**, 713 (1932).

[8] J. D. Bekenstein, M.S. thesis, Polytechnic Institute of Brooklyn, 1969 (unpublished).

[9] A. G. G. Mitchell and M. W. Zemansky, *Resonance Radiation and Excited Atoms* (Cambridge U.P., Cambridge, 1961), p. 98.

[10] R. Gebauer and H. R. v. Trabenberg, Z. Phys. **62**, 289 (1930).

$SL(2, C)$ Symmetry of the Gravitational Field Dynamical Variables

MOSHE CARMELI
General Physics Research Laboratory, Aerospace Research Laboratories
Wright-Patterson AFB, Ohio 45433

(Received 23 March 1970)

We represent the spin coefficients and the Riemann tensor in the form of linear combinations of the infinitesimal generators of the group $SL(2, C)$. This representation is similar to the way Yang and Mills write their dynamical variables in terms of the Pauli spin matrices. The spin coefficients take the role of the Yang-Mills-like potentials, whereas the Riemann tensor takes the role of the fields.

1. INTRODUCTION

The gravitational field dynamical variables of general relativity can be divided into three sets: (1) the Riemann tensor, decomposed into its irreducible components (the Weyl tensor, the trace-free parts of the Ricci tensor, and the Ricci scalar); (2) the spin coefficients; and (3) a tetrad system of vectors (from which one obtains the metric tensor). They are connected by three sets of first-order partial differential equations most conveniently given by Newman and Penrose.[1]

In this paper, we give a simple representation for the spin coefficients and for the components of the Riemann tensor in the form of linear combinations of the infinitesimal generators of the group $SL(2, C)$. This representation is very similar to the way Yang and Mills write their dynamical variables in terms of the Pauli spin matrices.[2] The spin coefficients take the role of the Yang-Mills-like potentials, whereas the Riemann tensor components take the role of the fields.

There is an essential difference, however, between our work and that of Yang and Mills. The group underlying the symmetry in our case is $SL(2, C)$, the group of all 2×2 complex matrices with determinant unity, whereas in the Yang-Mills case it is $SU(2)$.

The group $SL(2, C)$ seems to fit in with general relativity in a remarkable and natural way, just as 2-component spinors do. This is not an unexpected result, since spinors describe the finite-dimensional representation of $SL(2, C)$.[3,4]

With our representation of the field functions of gravitation, the Newman-Penrose field equations also have a simple and attractive representation. This latter result, however, is discussed elsewhere.

In Sec. 2, we give a very brief review of spinor calculus and its applications in general relativity. For details, see Ref. 1.

In Sec. 3, we obtain a set of four 2×2 complex matrices that have the form of the Yang-Mills potential and show how they represent the spin coefficients of general relativity. In Sec. 4, we define, following Yang and Mills, a set of six matrices to describe the gravitational field. It is then shown, in Sec. 5, that these six matrices describe the various components of the Riemann tensor.

In the Appendix, we give a brief discussion on the group $SL(2, C)$ and derive its infinitesimal generators.

2. SPIN FRAME

We review some standard techniques used in general relativity, thus establishing our notation. For details, see Ref. 1.

The correspondence between tensors and spinors is obtained by means of mixed quantities which are four 2×2 Hermitian matrices $\sigma^{\mu}{}_{AB'}$. Greek letters are used for tensor indices running over 0, 1, 2, 3 and Roman capitals for spinor indices taking the values 0, 1. Prime indices refer to the complex conjugate. The four matrices σ^{μ} satisfy the relation

$$g_{\mu\nu}\sigma^{\mu}{}_{AB'}\sigma^{\nu}{}_{CD'} = \epsilon_{AC}\epsilon_{B'D'}, \tag{2.1}$$

where $g_{\mu\nu}$ is the metric tensor and ϵ_{AC} and $\epsilon_{B'D'}$ along with ϵ^{AC} and $\epsilon^{B'D'}$ are the skew-symmetric Levi-Civita symbols given by

$$\epsilon_{AB} = \epsilon^{AB} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{2.2}$$

Raising or lowering a spinor index is done by means of the above symbol, with the following conventions:

$$\xi^A = \epsilon^{AB}\xi_B, \quad \xi_A = \xi^B\epsilon_{BA},$$
$$\eta^{A'} = \epsilon^{A'B'}\eta_{B'}, \quad \eta_{A'} = \eta^{B'}\epsilon_{B'A'}. \tag{2.3}$$

The spinor equivalent of a tensor is a quantity which has an unprimed and a primed spinor index for each tensor index. The spinor representing the tensor $T^{\alpha\beta}{}_{\gamma}$, for example, is

$$T^{AB'CD'}{}_{EF'} = \sigma_{\alpha}{}^{AB'}\sigma_{\beta}{}^{CD'}\sigma^{\gamma}{}_{EF'}T^{\alpha\beta}{}_{\gamma}. \tag{2.4}$$

The tensor representing the spinor $S^{AB'}{}_{CD'}$ is then

$$S^{\alpha}{}_{\beta} = \sigma^{\alpha}{}_{AB'}\sigma_{\beta}{}^{CD'}S^{AB'}{}_{CD'}. \tag{2.5}$$

Greek indices are raised and lowered, as usual, by the metric tensor $g^{\alpha\beta}$ and $g_{\alpha\beta}$, whose spinor expressions are given by

$$g^{AB'CD'} = \epsilon^{AC}\epsilon^{B'D'}, \quad g_{AB'CD'} = \epsilon_{AC}\epsilon_{B'D'}. \tag{2.6}$$

When taking the complex conjugate of a spinor, unprimed indices become primed, and primed indices become unprimed. The complex conjugate of $S^{AB'}$, for example, is $\bar{S}^{A'B}$, and, therefore, the condition for the vector S^α to be real is that its spinor be Hermitian:

$$S^{AB'} = \bar{S}^{B'A}. \tag{2.7}$$

The covariant derivative ∇_μ of a spinor ξ_A is

$$\nabla_\mu \xi_A = \partial_\mu \xi_A - \Gamma^B{}_{A\mu}\xi_B, \tag{2.8}$$

where $\Gamma^B{}_{A\mu}$ is the spinor affine connection. The choice of $\Gamma^B{}_{A\mu}$ is fixed by the requirement that the covariant derivatives of $\sigma^\mu{}_{AB'}$, ϵ_{AB}, and $\epsilon_{A'B'}$ shall all vanish:

$$\begin{aligned} \nabla_\alpha \sigma^\mu{}_{AB'} &= 0, \\ \nabla_\alpha \epsilon_{AB} &= 0, \\ \nabla_\alpha \epsilon_{A'B'} &= 0. \end{aligned} \tag{2.9}$$

At each point of space–time, two spinors $\zeta_a{}^A$, where $a = 0, 1$, are introduced to define a spin frame. These two spinors are supposed to satisfy the normalization condition

$$\zeta_a{}^B \epsilon_{BA} \zeta_b{}^A = \zeta_{aA}\zeta_b{}^A = \epsilon_{ab}. \tag{2.10}$$

An arbitrary spinor $S^{AB'}$ can then be written in their terms:

$$S^{AB'} = S^{ab'}\zeta_a{}^A \bar{\zeta}_{b'}{}^{B'}, \tag{2.11}$$

where

$$S_{ab'} = S_{AB'}\zeta_a{}^A \bar{\zeta}_{b'}{}^{B'} \tag{2.12}$$

are called the dyad components of $S_{AB'}$. By the same token, the quantity $\nabla_\mu \xi^A$, obtained by taking the covariant derivative of a spinor ξ^A, can be written as

$$\nabla_\mu \xi^A = B^b{}_\mu \zeta_b{}^A, \tag{2.13}$$

where the $B^b{}_\mu$, with $b = 0, 1$, are some vectors. In particular, Eq. (2.13) applies for the spinors $\zeta_a{}^A$. This gives

$$\nabla_\mu \zeta_a{}^A = B_a{}^b{}_\mu \zeta_b{}^A, \tag{2.14}$$

where again, $B_a{}^b{}_\mu$, with $a, b = 0, 1$, are some vectors. Using matrix notation, we see that Eq. (2,14) has the form

$$\nabla_\mu \zeta = B_\mu \zeta. \tag{2.15}$$

Here, B_μ and ζ are 2×2 complex matrices whose elements are $B_a{}^b{}_\mu$ and $\zeta_a{}^A$, respectively.

Quantities with lower-case indices behave the same way algebraically as the same quantities with capital spinor indices. But when covariant differentiation is applied, no term involving an affine connection appears for the lower-case indices.

The quantities $\sigma^\mu{}_{AB'}$ are not vectors. The expression

$$\sigma^\mu{}_{ab'} = \zeta_a{}^A \sigma^\mu{}_{AB'}\bar{\zeta}_{b'}{}^{B'}, \tag{2.16}$$

however, does define a null tetrad of vectors when ab' take the values $00'$, $01'$, $10'$, and $11'$. They satisfy the orthogonality relation

$$\sigma^\mu{}_{ab'}\sigma_{\mu cd'} = \epsilon_{ac}\epsilon_{b'd'}. \tag{2.17}$$

3. POTENTIALS AND SPIN COEFFICIENTS

In Sec. 2, we obtained the formula

$$\nabla_\mu \zeta = B_\mu \zeta, \tag{3.1}$$

where

$$B_\mu = (\nabla_\mu \zeta)\zeta^{-1} \tag{3.2}$$

are four 2×2 complex matrices whose elements are $B_a{}^b{}_\mu$ and ζ is the complex matrix whose elements are $\zeta_a{}^A$.

Moreover, by Eq. (2.10),

$$\det \zeta = 1. \tag{3.3}$$

Accordingly, $\zeta(x)$ is an element of the group $SL(2, C)$.

It is convenient to introduce another set of four matrices $\tilde{B}_\mu$, connected to B_μ by a similarity transformation

$$\zeta \tilde{B}_\mu = B_\mu \zeta. \tag{3.4}$$

The new set of matrices then satisfy

$$\nabla_\mu \zeta = \zeta \tilde{B}_\mu, \tag{3.5}$$

$$\tilde{B}_\mu = \zeta^{-1}\nabla_\mu \zeta. \tag{3.6}$$

The matrix elements of $\tilde{B}_\mu$ and B_μ are related as follows. If $B_a{}^b{}_\mu$ are the elements of B_μ, then $B_A{}^B{}_\mu$ will be those of $\tilde{B}_\mu$. This fact can easily be seen by writing the matrix elements of both sides of Eq. (3.4). The left-hand side gives

$$(\zeta \tilde{B}_\mu)_e{}^F = \zeta_e{}^D B_D{}^F{}_\mu, \tag{3.7}$$

whereas the right-hand side gives

$$(B_\mu \zeta)_e{}^F = B_e{}^d{}_\mu \zeta_d{}^F. \tag{3.8}$$

Using Eqs. (2.11) and (2.12), we see that both of these expressions are equal to $B_e{}^F{}_\mu$.

Now the matrices B_μ and $\tilde{B}_\mu$ are traceless. This is a result of the fact that ζ is unimodular. Hence, both B_μ and $\tilde{B}_\mu$ can be written as linear combinations of the infinitesimal matrices of the group $SL(2, C)$ similar to the way the Yang–Mills matrix potential is written in terms of the Pauli spin matrices. Denoting the infinitesimal generators by

$$\boldsymbol{\lambda} = (\lambda_+, \lambda_0, \lambda_-), \tag{3.9}$$

where (see Appendix)

$$\lambda_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \lambda_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \lambda_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{3.10}$$

we have

$$B_\mu = \mathbf{b}_\mu \cdot \boldsymbol{\lambda}. \tag{3.11}$$

Here, $\mathbf{b}_\mu \cdot \boldsymbol{\lambda}$ is a scalar product in the complex 3-dimensional $SL(2, C)$ space:

$$\mathbf{b}_\mu \cdot \boldsymbol{\lambda} = b^+{}_\mu \lambda_+ + b^0{}_\mu \lambda_0 + b^-{}_\mu \lambda_- . \tag{3.12}$$

[Boldface quantities denote complex 3-component vectors in $SL(2, C)$ space.] A similar equation holds for $\tilde{B}_\mu$.

From the matrices B_μ, one can define another set of four matrices

$$B_{cd'} = \sigma^\mu{}_{cd'} B_\mu . \tag{3.13}$$

Again, we can write these latter matrices as linear combinations of $\boldsymbol{\lambda}$:

$$B_{cd'} = \mathbf{b}_{cd'} \cdot \boldsymbol{\lambda}. \tag{3.14}$$

The four vectors

$$\mathbf{b}_{cd'} = (b^+{}_{cd'}, b^0{}_{cd'}, b^-{}_{cd'})$$

in the complex $SL(2, C)$ space are denoted by

$$\begin{aligned} \mathbf{b}_{00'} &= (-\kappa, \epsilon, \pi), \quad \mathbf{b}_{01'} = (-\sigma, \beta, \mu), \\ \mathbf{b}_{10'} &= (-\rho, \alpha, \lambda), \quad \mathbf{b}_{11'} = (-\tau, \gamma, \nu). \end{aligned} \tag{3.15}$$

$B_{ab'}$ then has the form

$$\begin{aligned} B_{00'} &= -\kappa\lambda_+ + \epsilon\lambda_0 + \pi\lambda_- , \\ B_{01'} &= -\sigma\lambda_+ + \beta\lambda_0 + \mu\lambda_- , \\ B_{10'} &= -\rho\lambda_+ + \alpha\lambda_0 + \lambda\lambda_- , \\ B_{11'} &= -\tau\lambda_+ + \gamma\lambda_0 + \nu\lambda_- , \end{aligned} \tag{3.16}$$

or, in matrix form,

$$\begin{aligned} B_{00'} &= \begin{pmatrix} \epsilon & -\kappa \\ \pi & -\epsilon \end{pmatrix}, \quad B_{01'} = \begin{pmatrix} \beta & -\sigma \\ \mu & -\beta \end{pmatrix}, \\ B_{10'} &= \begin{pmatrix} \alpha & -\rho \\ \lambda & -\alpha \end{pmatrix}, \quad B_{11'} = \begin{pmatrix} \gamma & -\tau \\ \nu & -\gamma \end{pmatrix}, \end{aligned} \tag{3.17}$$

when the representation (3.10) is used for $\boldsymbol{\lambda}$.

The twelve complex functions ϵ, κ, π, etc., were first introduced by Newman and Penrose[1] and are known in general relativity as spin coefficients. From the point of view of Yang–Mills field theory, these same quantities are potentials, the field of which is introduced in the next section.

4. THE $F_{\mu\nu}$ FIELD

In analogy to the procedure of obtaining the Yang–Mills field, we define a set of six traceless matrices[5]

$$F_{\mu\nu} = \nabla_\nu B_\mu - \nabla_\mu B_\nu + [B_\mu, B_\nu], \tag{4.1}$$

where

$$[B_\mu, B_\nu] = B_\mu B_\nu - B_\nu B_\mu .$$

The field $F_{\mu\nu}$ appears naturally if one applies the commutator $(\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu)$ on ζ:

$$(\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu)\zeta = F_{\mu\nu}\zeta. \tag{4.2}$$

By a similarity transformation, we define another set of six traceless matrices $\tilde{F}_{\mu\nu}$

$$\zeta\tilde{F}_{\mu\nu} = F_{\mu\nu}\zeta, \tag{4.3}$$

which satisfy

$$(\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu)\zeta = \zeta\tilde{F}_{\mu\nu} \tag{4.4}$$

and whose explicit expression is given by

$$\tilde{F}_{\mu\nu} = \nabla_\nu\tilde{B}_\mu - \nabla_\mu\tilde{B}_\nu - [\tilde{B}_\mu, \tilde{B}_\nu]. \tag{4.5}$$

By using the quantities $\sigma^\mu{}_{AB'}$ and $\sigma^\mu{}_{ab'}$, one obtains, from the matrices $F_{\mu\nu}$,

$$F_{AB'CD'} = \sigma^\mu{}_{AB'}\sigma^\nu{}_{CD'}F_{\mu\nu} \tag{4.6}$$

and

$$F_{ab'cd'} = \sigma^\mu{}_{ab'}\sigma^\nu{}_{cd'}F_{\mu\nu} . \tag{4.7}$$

Analogous expressions from $\tilde{F}_{\mu\nu}$ can be obtained.

Similar to the potential matrices, the matrix elements of $F_{\mu\nu}$ and $\tilde{F}_{\mu\nu}$ will be $F_a{}^b{}_{\mu\nu}$ and $F_A{}^B{}_{\mu\nu}$, respectively.

$F_{\mu\nu}$ defines a field with 18 complex functions. This is equivalent to the 20 (real) components of the Riemann tensor plus the 16 (real) components of the tetrad $\sigma^\mu{}_{ab'}$.[6] Again, one can write $F_{\mu\nu}$ as linear combinations of $\boldsymbol{\lambda}$,

$$F_{\mu\nu} = \mathbf{f}_{\mu\nu} \cdot \boldsymbol{\lambda}, \tag{4.8}$$

where the right-hand side of (4.8) is a scalar product in the complex $SL(2, C)$ space of the vector

$$\mathbf{f}_{\mu\nu} = (f^+{}_{\mu\nu}, f^0{}_{\mu\nu}, f^-{}_{\mu\nu}) \tag{4.9}$$

and $\boldsymbol{\lambda}$.

In the next section, we decompose the matrices $F_{ab'cd'}$ into their irreducible parts.

5. SYMMETRY OF $F_{ab'cd'}$

To find the $SL(2, C)$ structure of $F_{ab'cd'}$, we proceed as follows.

Let ξ^P be an arbitrary spinor. Then

$$\begin{aligned} (\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu)\xi^P &= (\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu)\xi^g\zeta_g{}^P \\ &= \xi^g(\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu)\zeta_g{}^P . \end{aligned} \tag{5.1}$$

Now, using Eq. (4.4), we obtain

$$(\nabla_\nu \nabla_\mu - \nabla_\mu \nabla_\nu)\xi^P = \xi^G \zeta_G{}^H F_H{}^P{}_{\mu\nu}. \tag{5.2}$$

Hence, we have

$$(\nabla_\nu \nabla_\mu - \nabla_\mu \nabla_\nu)\xi_Q = F_{PQ\mu\nu}\xi^P, \tag{5.3}$$

or, equivalently,

$$(\nabla_{AC'}\nabla_{BD'} - \nabla_{BD'}\nabla_{AC'})\xi_Q = F_{PQBD'AC'}\xi^P. \tag{5.4}$$

By decomposing the commutator of differentiation on the left-hand side of Eq. (5.4), we obtain[7]

$$\tfrac{1}{2}\epsilon_{C'D'}(\nabla_{AF'}\nabla_B{}^{F'} + \nabla_{BF'}\nabla_A{}^{F'})\xi_Q + \tfrac{1}{2}\epsilon_{AB}(\nabla_{EC'}\nabla^E{}_{D'} + \nabla_{ED'}\nabla^E{}_{C'})\xi_Q = F_{PQBD'AC'}\xi^P. \tag{5.5}$$

But the left-hand side of Eq. (5.5) is equal to[1]

$$\epsilon_{C'D'}[\Psi_{ABQP} - \Lambda(\epsilon_{PA}\epsilon_{BQ} + \epsilon_{PB}\epsilon_{AQ})]\xi^P + \epsilon_{AB}\Phi_{QPC'D'}\xi^P. \tag{5.6}$$

Here, Ψ_{ABCD} is a totally symmetric spinor which represents the Weyl spinor, and $\Phi_{QPC'D'}$ represents the trace-free part of the Ricci spinor having the symmetry

$$\Phi_{QPC'D'} = \Phi_{PQC'D'} = \Phi_{QPD'C'} = \bar{\Phi}_{C'D'QP}, \tag{5.7}$$

and

$$\Lambda = \tfrac{1}{24}R, \tag{5.8}$$

where R is the Ricci scalar.

Accordingly, we obtain

$$F_{PQBD'AC'} = \epsilon_{C'D'}[\Psi_{ABQP} - \Lambda(\epsilon_{PA}\epsilon_{BQ} + \epsilon_{PB}\epsilon_{AQ})] + \epsilon_{AB}\Phi_{QPC'D'}. \tag{5.9}$$

The same relation holds for lower-case indices:

$$F_p{}^q{}_{bd'ac'} = \epsilon_{c'd'}[\Psi_p{}^q{}_{ab} - \Lambda(\epsilon_{pa}\delta_b{}^q + \epsilon_{pb}\delta_a{}^q)] + \epsilon_{ab}\Phi_p{}^q{}_{c'd'}. \tag{5.10}$$

Using a standard notation

$$\Psi_{0000} = \Psi_0, \quad \Psi_{0001} = \Psi_1, \quad \Psi_{0011} = \Psi_2, \quad \Psi_{0111} = \Psi_3, \quad \Psi_{1111} = \Psi_4, \tag{5.11}$$

and

$$\begin{aligned} &\Phi_{000'0'} = \Phi_{00}, \quad \Phi_{010'1'} = \Phi_{11}, \quad \Phi_{000'1'} = \Phi_{01},\\ &\Phi_{011'1'} = \Phi_{12}, \quad \Phi_{010'0'} = \Phi_{10}, \quad \Phi_{110'1'} = \Phi_{21},\\ &\Phi_{001'1'} = \Phi_{02}, \quad \Phi_{111'1'} = \Phi_{22}, \quad \Phi_{110'0'} = \Phi_{20}, \end{aligned} \tag{5.12}$$

we finally obtain

$$\begin{aligned} F_{01'00'} &= \begin{pmatrix} \Psi_1 & -\Psi_0 \\ \Psi_2 & -\Psi_1 \end{pmatrix} + 2\begin{pmatrix} 0 & 0 \\ \Lambda & 0 \end{pmatrix},\\ F_{11'10'} &= \begin{pmatrix} \Psi_3 & -\Psi_2 \\ \Psi_4 & -\Psi_3 \end{pmatrix} - 2\begin{pmatrix} 0 & \Lambda \\ 0 & 0 \end{pmatrix},\\ F_{10'00'} &= \begin{pmatrix} \Phi_{10} & -\Phi_{00} \\ \Phi_{20} & -\Phi_{10} \end{pmatrix},\\ F_{11'01'} &= \begin{pmatrix} \Phi_{12} & -\Phi_{02} \\ \Phi_{22} & -\Phi_{12} \end{pmatrix},\\ F_{11'00'} &= \begin{pmatrix} \Psi_2 & -\Psi_1 \\ \Psi_3 & -\Psi_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11} & -\Phi_{01} \\ \Phi_{21} & -\Phi_{11} \end{pmatrix} - \begin{pmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{pmatrix},\\ F_{10'01'} &= -\begin{pmatrix} \Psi_2 & -\Psi_1 \\ \Psi_3 & -\Psi_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11} & -\Phi_{01} \\ \Phi_{21} & -\Phi_{11} \end{pmatrix} + \begin{pmatrix} \Lambda & 0 \\ 0 & -\Lambda \end{pmatrix}. \end{aligned} \tag{5.13}$$

Each of these matrices can be written as a linear combination of the $SL(2, C)$ infinitesimal generators λ_+, λ_0, and λ_-.

If we write

$$F_{ab'cd'} = \mathbf{f}_{ab'cd'} \cdot \boldsymbol{\lambda}, \tag{5.14}$$

where

$$\mathbf{f}_{ab'cd'} = (f^+{}_{ab'cd'}, f^0{}_{ab'cd'}, f^-{}_{ab'cd'}), \tag{5.15}$$

then we have

$$\begin{aligned} \mathbf{f}_{01'00'} &= (-\Psi_0, \Psi_1, \Psi_2 + 2\Lambda),\\ \mathbf{f}_{11'10'} &= (-\Psi_2 - 2\Lambda, \Psi_3, \Psi_4),\\ \mathbf{f}_{10'00'} &= (-\Phi_{00}, \Phi_{10}, \Phi_{20}),\\ \mathbf{f}_{11'01'} &= (-\Phi_{02}, \Phi_{12}, \Phi_{22}),\\ \mathbf{f}_{11'00'} &= (-\Psi_1 - \Phi_{01}, \Psi_2 + \Phi_{11} - \Lambda, \Psi_3 + \Phi_{21}),\\ \mathbf{f}_{10'01'} &= (\Psi_1 - \Phi_{01}, -\Psi_2 + \Phi_{11} + \Lambda, -\Psi_3 + \Phi_{21}). \end{aligned} \tag{5.16}$$

ACKNOWLEDGMENT

It is a pleasure to thank Dr. S. I. Fickler for conversations on gauge fields.

APPENDIX: INFINITESIMAL GENERATORS OF $SL(2, C)$

The group $SL(2, C)$ is the group of all 2×2 complex matrices with determinant unity. It is the covering group of the restricted Lorentz group describing homogeneous Lorentz transformations which are orthochroneous and proper. $SU(2)$ is, of course, a subgroup of $SL(2, C)$.

The group $SL(2, C)$ has been extensively discussed and represented in the classic book of Naimark.[3] Here, we only derive its infinitesimal generators using complex analysis. These generators play the same role that the Pauli matrices have with respect to $SU(2)$. To find the generators of $SL(2, C)$, we proceed as follows.

Let λ be a 2×2 complex matrix with determinant $+1$. Then λ can be expressed in terms of three independent complex parameters which we denote by θ^+, θ^0, and θ^-. It is then possible to show that every matrix

$$\lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \tag{A1}$$

of $SL(2, C)$ satisfying the condition $\lambda_{22} \neq 0$ can be represented uniquely in the form

$$\lambda = \lambda_+(\theta^+)\lambda_0(\theta^0)\lambda_-(\theta^-). \tag{A2}$$

The three matrices appearing on the right-hand side of Eq. (A2) are given by

$$\lambda_+(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \tag{A3}$$

$$\lambda_0(t) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \tag{A4}$$

$$\lambda_-(t) = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}, \tag{A5}$$

where t is a complex variable. They provide three 1-parameter subgroups of $SL(2, C)$ and satisfy

$$\begin{aligned} \lambda_+(t_1 + t_2) &= \lambda_+(t_1)\lambda_+(t_2), \\ \lambda_0(t_1 + t_2) &= \lambda_0(t_1)\lambda_0(t_2), \\ \lambda_-(t_1 + t_2) &= \lambda_-(t_1)\lambda_-(t_2). \end{aligned} \tag{A6}$$

The infinitesimal generators of $SL(2, C)$ are obtained from $\lambda_+(t)$, $\lambda_0(t)$, and $\lambda_-(t)$ by

$$\begin{aligned} \lambda_\pm &= [d\lambda_\pm(t)/dt]_{t=0}, \\ \lambda_0 &= [d\lambda_0(t)/dt]_{t=0}. \end{aligned} \tag{A7}$$

Hence, for the generators of $SL(2, C)$,[8] we have

$$\lambda_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \lambda_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \lambda_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{A8}$$

Conversely, the matrices $\lambda_+(t)$, $\lambda_0(t)$, and $\lambda_-(t)$ can be expressed in terms of the infinitesimal generators λ_+, λ_0, and λ_- by the formulas

$$\begin{aligned} \lambda_\pm(t) &= \exp(t\lambda_\pm), \\ \lambda_0(t) &= \exp(t\lambda_0), \end{aligned} \tag{A9}$$

as may be directly verified. Thus, for example,

$$\begin{aligned} \exp(t\lambda_0) &= I + t\lambda_0 + \frac{t^2}{2!}\lambda_0^2 + \frac{t^3}{3!}\lambda_0^3 + \cdots \\ &= \left(1 + \frac{t^2}{2!} + \cdots\right)I + \left(t + \frac{t^3}{3!} + \cdots\right)\lambda_0 \\ &= \cosh t\, I + \sinh t\, \lambda_0 \\ &= \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}. \end{aligned} \tag{A10}$$

We finally remark that the infinitesimal generators λ_+, λ_0, and λ_- satisfy the commutation relations

$$\begin{aligned} [\lambda_0, \lambda_\pm] &= \pm 2\lambda_\pm, \\ [\lambda_+, \lambda_-] &= \lambda_0. \end{aligned} \tag{A11}$$

[1] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[2] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[3] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964).
[4] For a generalization of spinors to infinite-dimensional representations of $SL(2, C)$, see M. Carmeli, J. Math. Phys. **11**, 1917 (1970).
[5] Yang and Mills define their field by using partial derivatives instead of the covariant derivatives in Eq. (4.1). The two expressions, however, are equal since $\nabla_\nu B_\mu - \nabla_\mu B_\nu = \partial_\nu B_\mu - \partial_\mu B_\nu$.
[6] I am indebted to Professor L. Witten for a discussion on this point.
[7] R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).
[8] Our infinitesimal generators are related to those of Gel'fand, Graev, and Vilenkin by

$$\lambda_\pm = a_\pm, \quad \lambda_0 = 2a_0.$$

See I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin *Integral Geometry and Representation Theory* (Academic, New York, 1966).

Radiative Transfer in a Rayleigh-Scattering Atmosphere with True Absorption*

T. W. SCHNATZ AND C. E. SIEWERT
Department of Nuclear Engineering, North Carolina State University, Raleigh, North Carolina 27607

(Received 4 March 1970)

The singular-eigenfunction-expansion technique is used to solve the equation of transfer for partially polarized light in a Rayleigh-scattering atmosphere with true absorption. The normal modes for the considered nonconservative vector equation of transfer are established; two discrete eigenvectors and two linearly independent continuum solutions are thus derived. Further, the necessary full-range completeness and orthogonality theorems are proved, so that all expansion coefficients can be determined explicitly, and, in order to illustrate the technique, an exact analytical solution for the infinite-medium Green's function is developed. Finally, a numerical tabulation of the required discrete eigenvalue, as a function of the single-scatter albedo, is given.

I. INTRODUCTION

In one of his classical papers on radiative transfer,[1] Chandrasekhar formulated explicitly the equations of transfer for the two components $I_l(\tau, \mu)$ and $I_r(\tau, \mu)$ of a polarized radiation field in a free-electron atmosphere; he also developed an approximate solution for the law of darkening appropriate to the considered Milne problem. This latter result was subsequently improved[2] as Chandrasekhar was able to observe the infinite limit of a discrete-ordinates procedure in order to establish a rigorous solution for the desired surface quantities.

More recently, as the study of neutron physics has developed, Wigner[3] has discussed a theory of neutron transport which takes into consideration the quantum mechanical effects of neutron polarization. The influence of neutron polarization on the scattering of fast neutrons by unpolarized nuclei has also been reported recently by Bell and Goad,[4] who used the P_l approximation to the transport solution.

Although most studies [for example, Refs. 5–8] of the scattering of polarized light have been based on Chandrasekhar's model,[9] the extension to include the effects of true absorption has been discussed briefly by Sobolev[10] and Simmons.[11] Further, Mullikin[12] recently extended his earlier work[13] on the conservative model and reported the results of a more general investigation, which accounted for true absorption by allowing the single-scatter albedo to be less than unity.

Since the principal interest relevant to many astrophysical studies of polarized light is in the evaluation of surface quantities, Chandrasekhar's invariance principles[9] have been widely used[5–8]; however, the singular-eigenfunction-expansion technique developed by Case[14] has been used to advantage by Siewert and Fraley[15] to construct rigorous analytical solutions, valid *anywhere* within the medium, to the Milne problem and other half-space problems. This latter method was also used by Mourad and Siewert[16] to establish full-range completeness and orthogonality theorems basic to the normal modes of a more general vector equation of transfer, also formulated by Chandrasekhar.[9] For this case, the mathematical model used to describe the scattering of polarized light by molecules was also shown to be appropriate for the theory of resonance line scattering.

Although Mourad and Siewert[16] did not find closed-form expressions for the more interesting half-range applications, the computational merits of their results have recently been confirmed for the half-space Milne problem.[17] The vector equation of transfer considered in Ref. 17 is inherently restricted to conservative media, but similar half-range methods should be available for the nonconservative model discussed here.

We consider, then, the equation of transfer

$$\mu \frac{\partial}{\partial \tau} \mathbf{I}(\tau, \mu) + \mathbf{I}(\tau, \mu) = \tfrac{1}{2} c \int_{-1}^{1} \mathbf{K}(\mu, \mu') \mathbf{I}(\tau, \mu')\, d\mu', \quad (1)$$

where the Rayleigh-scattering matrix is given by

$$\mathbf{K}(\mu, \mu') = \frac{3}{4} \begin{vmatrix} 2(1-\mu^2)(1-\mu'^2) + \mu'^2\mu^2 & \mu^2 \\ \mu'^2 & 1 \end{vmatrix}. \quad (2)$$

Here, τ is the optical variable, μ is the direction cosine (as measured from the *positive* τ axis) of the propagating radiation, and $c \in [0, 1]$ is the single-scatter albedo. Further, the desired intensities $I_l(\tau, \mu)$ and $I_r(\tau, \mu)$ are the two components of the vector $\mathbf{I}(\tau, \mu)$.

We prefer to make use of Sekera's[18] factorization

$$\mathbf{K}(\mu, \mu') = \tfrac{3}{4} \mathbf{Q}(\mu) \mathbf{Q}^{\mathrm{T}}(\mu'), \quad (3)$$

where the superscript "T" denotes the transpose operation, in order to write Eq. (1) in the form

$$\mu\frac{\partial}{\partial\tau}\mathbf{I}(\tau,\mu)+\mathbf{I}(\tau,\mu)=\tfrac{3}{8}c\mathbf{Q}(\mu)\int_{-1}^{1}\mathbf{Q}^{\mathrm{T}}(\mu')\mathbf{I}(\tau,\mu')\,d\mu', \tag{4a}$$

where

$$\mathbf{Q}(\mu)\overset{\text{DEF}}{=}\begin{vmatrix}\mu^2 & 2^{\frac{1}{2}}(1-\mu^2)\\ 1 & 0\end{vmatrix}. \tag{4b}$$

II. EIGENVALUE SPECTRUM AND EIGENVECTORS

Since the development of the normal modes of Eq. (1) follows previously reported analysis of vector equations of transfer,[15,16,19,20] we should like only to summarize our results here.

Proposing solutions of the form

$$\mathbf{I}(\tau,\mu)=e^{-\tau/\eta}\mathbf{\Phi}(\eta,\mu), \tag{5}$$

we note that the eigenvalues η and associated eigenvectors $\mathbf{\Phi}(\eta,\mu)$ are to be determined from the reduced equation

$$(\eta-\mu)\mathbf{\Phi}(\eta,\mu)=\tfrac{3}{8}c\eta\mathbf{Q}(\mu)\int_{-1}^{1}\mathbf{Q}^{\mathrm{T}}(\mu')\mathbf{\Phi}(\eta,\mu')\,d\mu'. \tag{6}$$

If we now introduce the normalization

$$\mathbf{M}(\eta)\overset{\text{DEF}}{=}\int_{-1}^{1}\mathbf{Q}^{\mathrm{T}}(\mu)\mathbf{\Phi}(\eta,\mu)\,d\mu, \tag{7}$$

then clearly the discrete eigenvectors are given by

$$\mathbf{\Phi}(\eta,\mu)=[3c\eta/8(\eta-\mu)]\mathbf{Q}(\mu)\mathbf{M}(\eta), \tag{8}$$

where

$$\left(\mathbf{I}-\tfrac{3}{8}c\eta\int_{-1}^{1}\mathbf{Q}^{\mathrm{T}}(\mu)\mathbf{Q}(\mu)\frac{d\mu}{\eta-\mu}\right)\mathbf{M}(\eta)=\mathbf{0}; \tag{9}$$

here, $\mathbf{I}$ denotes the identity matrix.

Thus, we obtain the discrete eigenvalue spectrum from the zeros of the dispersion function $\Lambda(z)$ defined as

$$\Lambda(z)\overset{\text{DEF}}{=}8\det\left(\mathbf{I}+\tfrac{3}{8}cz\int_{-1}^{1}\mathbf{Q}^{\mathrm{T}}(\mu)\mathbf{Q}(\mu)\frac{d\mu}{\mu-z}\right), \tag{10}$$

where the factor 8 has been included in order to obtain the more convenient form

$$\Lambda(z)=\Lambda_1(z)\Lambda_2(z)+12z^2(1-c)\Lambda_0(z). \tag{11}$$

Here,

$$\Lambda_\alpha(z)=(-1)^\alpha+3(1-z^2)\Lambda_0(z)-(-1)^\alpha 3z^2(1-c),\quad \alpha=1\text{ or }2, \tag{12a}$$

and

$$\Lambda_0(z)=1+\tfrac{1}{2}cz\int_{-1}^{1}\frac{d\mu}{\mu-z}. \tag{12b}$$

We note from Eq. (10) that $\Lambda(z)$ is a function analytic in the complex plane cut from -1 to 1 and, as we discuss in Sec. VI, the argument principle[21] can be used to show that $\Lambda(z)$ has only two zeros, which appear as a pair $z=\pm\eta_0$ of real eigenvalues. Thus, there are two discrete eigenvectors $\mathbf{\Phi}_\pm(\mu)$ which, after judiciously normalizing the vector $\mathbf{M}(\eta)$, we write in the tractable form

$$\mathbf{\Phi}_\pm(\mu)=\tfrac{3}{2}c\eta_0\frac{1}{\eta_0\mp\mu}\begin{vmatrix}\Lambda_2(\eta_0)(1-\mu^2)+2\eta_0^2(1-c)\\ 2\eta_0^2(1-c)\end{vmatrix}. \tag{13}$$

For the continuum $\eta\in(-1,1)$, we express the solution to Eq. (6) in the form

$$\mathbf{\Phi}(\eta,\mu)=\tfrac{3}{8}c\eta\left(\frac{P}{\eta-\mu}+\lambda(\eta)\delta(\eta-\mu)\right)\mathbf{Q}(\mu)\mathbf{M}(\eta), \tag{14}$$

where $\lambda(\eta)$ is as yet unspecified; further, the symbol P is used to denote that all ensuing integrals over η or μ are to be evaluated in the Cauchy principal-value sense, and $\delta(x)$ is the Dirac δ function. Multiplying Eq. (14) by $\mathbf{Q}^{\mathrm{T}}(\mu)$ and integrating over μ from -1 to 1 yields a homogeneous equation for $\mathbf{M}(\eta)$. The compatibility condition thus yields a quadratic equation in $\lambda(\eta)$. We find two independent solutions for $\lambda(\eta)$ and thus establish two linearly independent continuum eigenvectors. Choosing a convenient normalization for $\mathbf{M}(\eta)$, we write

$$\mathbf{\Phi}_1(\eta,\mu)=\begin{vmatrix}\tfrac{3}{2}c\eta(1-\eta^2)(1-\mu^2)P/(\eta-\mu)\\ \quad+[(1-\eta^2)\lambda_1(\eta)+2\eta^2(1-c)]\delta(\eta-\mu)\\ -2\eta^2(1-c)\delta(\eta-\mu)\end{vmatrix}, \tag{15a}$$

$$\mathbf{\Phi}_2(\eta,\mu)=\begin{vmatrix}\tfrac{3}{2}c\eta(1-\eta^2)P/(\eta-\mu)+\lambda_1(\eta)\delta(\eta-\mu)\\ \tfrac{3}{2}c\eta(1-\eta^2)P/(\eta-\mu)+\lambda_2(\eta)\delta(\eta-\mu)\end{vmatrix}, \tag{15b}$$

$$\lambda_\alpha(\eta)=(-1)^\alpha+3(1-\eta^2)\lambda_0(\eta)-(-1)^\alpha 3\eta^2(1-c), \tag{16a}$$

$$\lambda_0(\eta)=1-c\eta\tanh^{-1}(\eta). \tag{16b}$$

Having determined the eigenvectors, we can express the general solution of Eq. (1) as a linear sum of the independent solutions:

$$\mathbf{I}(\tau,\mu)=A_+e^{-\tau/\eta_0}\mathbf{\Phi}_+(\mu)+A_-e^{\tau/\eta_0}\mathbf{\Phi}_-(\mu)+\int_{-1}^{1}[A_1(\eta)\mathbf{\Phi}_1(\eta,\mu)+A_2(\eta)\mathbf{\Phi}_2(\eta,\mu)]e^{-\tau/\eta}\,d\eta, \tag{17}$$

where $A_{\pm}$, $A_1(\eta)$, and $A_2(\eta)$ are the arbitrary coefficients to be determined from the boundary conditions of a suitably defined physical problem.

We note that for the special case $c = 1$, the dispersion relation given by Eq. (11) and the established eigenvectors reduce to forms equivalent to those obtained by Siewert and Fraley.[15]

III. FULL-RANGE COMPLETENESS OF THE EIGENVECTORS

Theorem 1: The eigenvectors $\mathbf{\Phi}_{\pm}(\mu)$, $\mathbf{\Phi}_1(\eta, \mu)$, and $\mathbf{\Phi}_2(\eta, \mu)$ are complete on the full range, $\mu \in (-1, 1)$, in the sense that an arbitrary 2-component vector $\mathbf{\Psi}(\mu)$ satisfying the Hölder condition for $\mu \in (-1, 1)$ can be expanded in the form

$$\mathbf{\Psi}(\mu) = A_+\mathbf{\Phi}_+(\mu) + A_-\mathbf{\Phi}_-(\mu) + \int_{-1}^{1} A_1(\eta)\mathbf{\Phi}_1(\eta, \mu)\, d\eta + \int_{-1}^{1} A_2(\eta)\mathbf{\Phi}_2(\eta, \mu)\, d\eta, \quad \mu \in (-1, 1). \tag{18}$$

In order to prove the theorem, we use the methods of Muskhelishvili[22] to convert Eq. (18) to the equivalent Riemann–Hilbert problem

$$\mu(1 - \mu^2)\mathbf{\Psi}'(\mu) = \mathbf{\Lambda}^+(\mu)\mathbf{N}^+(\mu) - \mathbf{\Lambda}^-(\mu)\mathbf{N}^-(\mu), \quad \mu \in (-1, 1), \tag{19}$$

where

$$\mathbf{\Psi}'(\mu) = \mathbf{\Psi}(\mu) - A_+\mathbf{\Phi}_+(\mu) - A_-\mathbf{\Phi}_-(\mu) \tag{20}$$

and

$$\mathbf{N}(z) = \frac{1}{2\pi i}\int_{-1}^{1} \eta(1 - \eta^2)\mathbf{A}(\eta)\frac{d\eta}{\eta - z}; \tag{21}$$

here, $\mathbf{A}(\eta)$ is a vector with components $A_1(\eta)$ and $A_2(\eta)$. We note that $\mathbf{N}(z)$ is analytic in the complex plane cut from -1 to 1. Also, it vanishes at least as fast as $1/z$ when z increases without bound. Further, the boundary values of $\mathbf{N}(z)$ as z approaches the cut from above (+) and below (−) can be shown to satisfy the following relations deducible from the Plemelj formulas[22]:

$$\pi i[\mathbf{N}^+(\mu) + \mathbf{N}^-(\mu)] = \int_{-1}^{1} \eta(1 - \eta^2)\mathbf{A}(\eta)\frac{P}{\eta - \mu}\, d\eta \tag{22a}$$

and

$$\mathbf{N}^+(\mu) - \mathbf{N}^-(\mu) = \mu(1 - \mu^2)\mathbf{A}(\mu). \tag{22b}$$

In establishing Eq. (19), we introduced the matrix

$$\mathbf{\Lambda}(z) \overset{\text{DEF}}{=} \begin{vmatrix} (1 - z^2)\Lambda_1(z) + 2z^2(1 - c) & \Lambda_1(z) \\ -2z^2(1 - c) & \Lambda_2(z) \end{vmatrix}. \tag{23}$$

It can be seen from Eqs. (12) and Eq. (16) that the functions $\Lambda_\alpha(z)$, $\alpha = 1$ and 2, are also analytic in the complex plane cut from -1 to 1 and that the boundary values $\Lambda_\alpha^{\pm}(\mu)$ obey

$$\Lambda_\alpha^+(\mu) + \Lambda_\alpha^-(\mu) = 2\lambda_\alpha(\mu), \qquad \alpha = 1 \text{ or } 2, \tag{24a}$$

and

$$\Lambda_\alpha^+(\mu) - \Lambda_\alpha^-(\mu) = 3\pi i c\mu(1 - \mu^2), \quad \alpha = 1 \text{ or } 2. \tag{24b}$$

Defining a vector $\mathbf{P}(z)$ with components $P_1(z)$ and $P_2(z)$ as

$$\mathbf{P}(z) = \mathbf{\Lambda}(z)\mathbf{N}(z) - \frac{1}{2\pi i}\int_{-1}^{1} \mu(1 - \mu^2)\mathbf{\Psi}'(\mu)\frac{d\mu}{\mu - z}, \tag{25}$$

we note that $\mathbf{P}(z)$ is analytic in the complex plane cut from -1 to 1. It follows from Eq. (19) that $\mathbf{P}(z)$ is continuous across the cut and thus is an entire function. If we consider the behavior of the functions $\Lambda_\alpha(z)$, $\alpha = 0$, 1, and 2, as z tends to infinity, we observe that

$$\Lambda_0(z) \sim (1 - c) - c/3z^2 - c/5z^4, \quad z \to \infty, \tag{26a}$$

$$\Lambda_1(z) \sim 2(1 - c) - 2c/5z^2, \qquad z \to \infty, \tag{26b}$$

and

$$\Lambda_2(z) \sim 2(2 - c) - 6z^2(1 - c), \qquad z \to \infty. \tag{26c}$$

The above results may be employed to deduce the behavior of $\mathbf{\Lambda}(z)$ for large z:

$$\mathbf{\Lambda}(z) \sim \begin{vmatrix} \frac{2}{5}(5 - 4c) & 2(1 - c) \\ -2z^2(1 - c) & -6z^2(1 - c) \end{vmatrix}, \quad z \to \infty, \tag{27}$$

and thus, from Eqs. (25) and (21), we conclude that, as z tends to infinity, $P_1(z)$ vanishes while $P_2(z)$ has a first-order pole. Liouville's theorem[23] then requires that $P_2(z)$ be a first-order polynomial, whereas $P_1(z)$ must be identically zero. Thus, we find

$$\mathbf{P}(z) = \begin{vmatrix} 0 \\ a + bz \end{vmatrix}, \tag{28}$$

where a and b are arbitrary constants.

Equation (25) can now be solved for $\mathbf{N}(z)$ to yield

$$\mathbf{N}(z) = \mathbf{\Lambda}^{-1}(z)\left(\frac{1}{2\pi i}\int_{-1}^{1} \mu(1 - \mu^2)\mathbf{\Psi}'(\mu)\frac{d\mu}{\mu - z} + \mathbf{P}(z)\right), \tag{29}$$

where the inverse of $\mathbf{\Lambda}(z)$ is given by

$$\mathbf{\Lambda}^{-1}(z) = \frac{1}{(1-z^2)\Lambda(z)} \times \begin{vmatrix} \Lambda_2(z) & -\Lambda_1(z) \\ 2z^2(1-c) & (1-z^2)\Lambda_1(z) + 2z^2(1-c) \end{vmatrix}. \quad (30)$$

Since $\Lambda(z)$ has zeros at $z = \pm\eta_0$, we note that $\mathbf{N}(z)$ is not a holomorphic function in the complex plane cut from -1 to 1 unless we impose on $\mathbf{\Psi}'(\mu)$ the two constraints

$$\begin{vmatrix} \Lambda_2(\eta_0) & -\Lambda_1(\eta_0) \\ 2\eta_0^2(1-c) & (1-\eta_0^2)\Lambda_1(\eta_0) + 2\eta_0^2(1-c) \end{vmatrix} \times \left(\frac{1}{2\pi i}\int_{-1}^{1}\mu(1-\mu^2)\mathbf{\Psi}'(\mu)\frac{d\mu}{\mu \mp \eta_0} + \mathbf{P}(\pm\eta_0)\right) = \mathbf{0}. \quad (31)$$

In addition, we observe that $\mathbf{\Lambda}(z)$ is singular at the branch points $z = \pm 1$, so that we must carefully investigate the end-point[22] behavior of $\mathbf{N}(z)$. Observing the limits as z tends to ± 1 in Eq. (29), we obtain

$$\lim_{z\to\pm1}\mathbf{N}(z) \sim \{(1-z^2)[-(2-3c)^2 + 12(1-c)\Lambda_0(z)]\}^{-1} \times \begin{vmatrix} -(2-3c) & -(2-3c) \\ 2(1-c) & 2(1-c) \end{vmatrix} \times \left(\frac{1}{2\pi i}\int_{-1}^{1}\mu(1-\mu^2)\mathbf{\Psi}'(\mu)\frac{d\mu}{\mu \mp 1} + \mathbf{P}(\pm 1)\right). \quad (32)$$

The singularities at $z = \pm 1$ introduced by the factor $(1-z^2)^{-1}$ in the above equation are termed special end-points by Muskhelishvili,[22] and, in order for $\mathbf{N}(z)$ to have the proper end-point behavior, we impose the additional constraints

$$\begin{vmatrix} 1 \\ 1 \end{vmatrix}^{\mathrm{T}} \left[\frac{1}{2\pi i}\int_{-1}^{1}\mu(1\pm\mu)\mathbf{\Psi}'(\mu)\,d\mu - \begin{vmatrix} 0 \\ \pm a + b \end{vmatrix}\right] = 0. \quad (33)$$

Equation (33) can be solved immediately for a and b to yield

$$a = \frac{1}{2\pi i}\int_{-1}^{1}\mu^2 \begin{vmatrix} 1 \\ 1 \end{vmatrix}^{\mathrm{T}} \mathbf{\Psi}'(\mu)\,d\mu \quad (34a)$$

and

$$b = \frac{1}{2\pi i}\int_{-1}^{1}\mu \begin{vmatrix} 1 \\ 1 \end{vmatrix}^{\mathrm{T}} \mathbf{\Psi}'(\mu)\,d\mu. \quad (34b)$$

Having determined $\mathbf{P}(z)$, we investigate more thoroughly the original constraints on $\mathbf{\Psi}'(\mu)$ as given by Eq. (31). Rewriting Eq. (31) as two separate equations, we observe that

$$\begin{vmatrix} \Lambda_2(\eta_0) \\ -\Lambda_1(\eta_0) \end{vmatrix}^{\mathrm{T}} \times \left(\int_{-1}^{1}\mu(1-\mu^2)\mathbf{\Psi}'(\mu)\frac{d\mu}{\mu \mp \eta_0} + 2\pi i\mathbf{P}(\pm\eta_0)\right) = 0 \quad (35a)$$

and

$$\begin{vmatrix} 2\eta_0^2(1-c) \\ (1-\eta_0^2)\Lambda_1(\eta_0) + 2\eta_0^2(1-c) \end{vmatrix}^{\mathrm{T}} \times \left(\int_{-1}^{1}\mu(1-\mu^2)\mathbf{\Psi}'(\mu)\frac{d\mu}{\mu \mp \eta_0} + 2\pi i\,\mathbf{P}(\pm\eta_0)\right) = 0. \quad (35b)$$

The above expressions can be rearranged to yield

$$\int_{-1}^{1}\mu\mathbf{\Phi}_{\pm}^{\mathrm{T}}(\mu)\mathbf{\Psi}'(\mu)\,d\mu + \Delta\int_{-1}^{1}\mu(\mu\pm\eta_0)\begin{vmatrix} 1 \\ 1 \end{vmatrix}^{\mathrm{T}}\mathbf{\Psi}'(\mu)\,d\mu - 2\pi i\Delta\begin{vmatrix} 0 \\ 1 \end{vmatrix}^{\mathrm{T}}\mathbf{P}(\pm\eta_0) = 0, \quad (36)$$

where

$$\Delta = 3c\eta_0^3(1-c)/(1-\eta_0^2). \quad (37)$$

The results given by Eqs. (34) for a and b can now be introduced into Eq. (36), thus reducing the constraints on $\mathbf{\Psi}'(\mu)$ to the explicit form

$$\int_{-1}^{1}\mu\mathbf{\Phi}_{\pm}^{\mathrm{T}}(\mu)\mathbf{\Psi}'(\mu)\,d\mu = 0. \quad (38)$$

In general, this condition is not met; however, noting $\mathbf{\Psi}'(\mu)$ as given by Eq. (20), we can determine A_+ and A_- such that Eq. (38) is satisfied. With A_+ and A_- so established and a and b given by Eqs. (34), the result for $\mathbf{N}(z)$ expressed by Eq. (29) exhibits the proper analytic properties; the completeness theorem is thus proved.

Although the proof of Theorem 1 can be pursued to yield explicit expressions for all expansion coefficients $A_\pm$, $A_1(\eta)$, and $A_2(\eta)$, we prefer to use the alternative full-range orthogonality theorem developed in the next section to establish these results.

IV. ORTHOGONALITY, NORMALIZATION INTEGRALS, AND ADJOINT FUNCTIONS

Theorem 2: The eigenvectors $\mathbf{\Phi}_\pm(\mu)$, $\mathbf{\Phi}_1(\eta,\mu)$, and $\mathbf{\Phi}_2(\eta,\mu)$ are orthogonal on the full range, with respect to weight function μ, i.e.,

$$\int_{-1}^{1}\mu\mathbf{\Phi}_i^{\mathrm{T}}(\eta',\mu)\mathbf{\Phi}_j(\eta,\mu)\,d\mu = 0, \quad \eta \neq \eta', \quad i,j = +,-,1, \text{ or } 2. \quad (39)$$

We begin the proof by premultiplying Eq. (6) by $\mathbf{\Phi}^{\mathrm{T}}(\eta', \mu)/\eta$. Equation (6) with η replaced by η' is then transposed and postmultiplied by $\mathbf{\Phi}(\eta, \mu)/\eta'$. The resulting equations are integrated over μ from -1 to I and subtracted to yield

$$\left(\frac{1}{\eta'} - \frac{1}{\eta}\right)\int_{-1}^{1} \mu \mathbf{\Phi}^{\mathrm{T}}(\eta', \mu)\mathbf{\Phi}(\eta, \mu)\, d\mu = 0. \quad (40)$$

Thus, the proof is established. However, some difficulty still remains, since the continuum eigenvectors are degenerate. A vanishing scalar product between the continuum eigenvectors is not guaranteed by the theorem. In fact, if we define

$$\delta(\eta - \eta')\eta M_{ij}(\eta) \overset{\mathrm{DEF}}{=} \int_{-1}^{1} \mu \mathbf{\Phi}_i^{\mathrm{T}}(\eta', \mu)\mathbf{\Phi}_j(\eta, \mu)\, d\mu,$$
$$\eta \text{ and } \eta' \in (-1, 1), \quad i, j = 1 \text{ or } 2, \quad (41)$$

we find upon evaluating the above integrals that

$$M_{12}(\eta) = M_{21}(\eta) = (1 - \eta^2)\Lambda_1^+(\eta)\Lambda_1^-(\eta) - 4\eta^2(1 - c)[1 - 3\eta^2(1 - c)], \quad (42a)$$
$$M_{11}(\eta) = (1 - \eta^2)^2\Lambda_1^+(\eta)\Lambda_1^-(\eta) + 4\eta^2(1 - c) \times [(1 - \eta^2)\lambda_1(\eta) + 2\eta^2(1 - c)], \quad (42b)$$

and

$$M_{22}(\eta) = \Lambda_1^+(\eta)\Lambda_1^-(\eta) + \Lambda_2^+(\eta)\Lambda_2^-(\eta). \quad (42c)$$

A Schmidt-type procedure can now be used to develop a set of adjoint eigenvectors such that, if we define the scalar product as

$$\langle i \mid j\rangle \overset{\mathrm{DEF}}{=} \int_{-1}^{1} \mu \mathbf{\Phi}_i^{\mathrm{T}\dagger}(\eta, \mu)\mathbf{\Phi}_j(\eta, \mu)\, d\mu,$$
$$i, j = +, -, 1, \text{ and } 2, \quad (43)$$

where the adjoint vectors are defined as

$$\mathbf{\Phi}_\pm^\dagger(\mu) = \mathbf{\Phi}_\pm(\mu), \quad (44a)$$
$$\mathbf{\Phi}_1^\dagger(\eta, \mu) = M_{22}(\eta)\mathbf{\Phi}_1(\eta, \mu) - M_{12}(\eta)\mathbf{\Phi}_2(\eta, \mu), \quad (44b)$$

and

$$\mathbf{\Phi}_2^\dagger(\eta, \mu) = M_{11}(\eta)\mathbf{\Phi}_2(\eta, \mu) - M_{21}(\eta)\mathbf{\Phi}_1(\eta, \mu), \quad (44c)$$

then the desired orthogonality property is established, viz.,

$$\langle i \mid j\rangle = 0, \quad i \neq j. \quad (45)$$

With the definitions given by Eqs. (44), the necessary normalization integrals can be evaluated straightforwardly. We find

$$\langle \pm \mid \pm\rangle = M_\pm \quad (46a)$$

and

$$\langle 1 \mid 1\rangle = \langle 2 \mid 2\rangle = M(\eta)\delta(\eta - \eta'), \quad (46b)$$

where

$$M_\pm = \pm\tfrac{3}{8}c\eta_0^3\mathbf{M}^{\mathrm{T}}(\eta_0) \times \left(\tfrac{3}{8}c\int_{-1}^{1} \mathbf{Q}^{\mathrm{T}}(\mu)\mathbf{Q}(\mu)\frac{d\mu}{(\mu - \eta_0)^2} - \frac{1}{\eta_0^2}\mathbf{I}\right)\mathbf{M}(\eta_0) \quad (47a)$$

and

$$M(\eta) = \eta(1 - \eta^2)^2\Lambda^+(\eta)\Lambda^-(\eta). \quad (47b)$$

We now proceed to determine explicit expressions for $A_\pm$, $A_1(\eta)$, and $A_2(\eta)$ in the full-range expansion given by Eq. (18). If we multiply Eq. (18) by $\mu\mathbf{\Phi}_\pm^{\mathrm{T}\dagger}(\mu)$ and integrate over μ from -1 to 1, we find

$$A_\pm M_\pm = \int_{-1}^{1} \mu \mathbf{\Phi}_\pm^{\mathrm{T}\dagger}(\mu)\mathbf{\Psi}(\mu)\, d\mu, \quad (48a)$$

where we have utilized the results given by Eqs. (45) and (46). Similarly, we take scalar products of Eq. (18) with the adjoint vectors $\mathbf{\Phi}_1^\dagger(\eta, \mu)$ and $\mathbf{\Phi}_2^\dagger(\eta, \mu)$ to find

$$A_i(\eta)M(\eta) = \int_{-1}^{1} \mu \mathbf{\Phi}_i^{\mathrm{T}\dagger}(\eta, \mu)\mathbf{\Psi}(\mu)\, d\mu, \quad i = 1 \text{ or } 2. \quad (48b)$$

We note that the above results are identical with those which follow from the completeness proof given in Sec. III.

V. THE INFINITE-MEDIUM GREEN'S FUNCTION

We now illustrate the utility of the above theory by constructing a solution for the infinite-medium Green's function. We seek a bounded solution to the equation

$$\mu\frac{\partial}{\partial\tau}\mathbf{I}(\tau, \mu) + \mathbf{I}(\tau, \mu) = \tfrac{3}{8}c\mathbf{Q}(\mu)\int_{-1}^{1} \mathbf{Q}^{\mathrm{T}}(\mu')\mathbf{I}(\tau, \mu')\, d\mu' + \mathbf{S}(\tau, \mu), \quad (49)$$

where $\mathbf{S}(\tau, \mu)$ is defined as

$$\mathbf{S}(\tau, \mu) = \delta(\tau)\begin{vmatrix} s_l\delta(\mu - \mu_l) \\ s_r\delta(\mu - \mu_r) \end{vmatrix},$$
$$\mu, \quad \mu_l, \quad \text{and} \quad \mu_r \in (-1, 1). \quad (50)$$

Here, s_l and s_r represent the source strengths of each of the two polarization states. In the usual manner,[24] we neglect the source term in Eq. (49) and require the solutions of the resulting homogeneous equation to satisfy the "jump" boundary condition

$$\mu[\mathbf{I}(0, \mu_l, \mu_r; 0^+, \mu) - \mathbf{I}(0, \mu_l, \mu_r; 0^-, \mu)] = \begin{vmatrix} s_l\delta(\mu - \mu_l) \\ s_r\delta(\mu - \mu_r) \end{vmatrix}, \quad (51)$$

where the argument list has been extended to include the location of the source as well as the parameters μ_l and μ_r.

The solution is separated into two parts which are respectively bounded in the two half-spaces $\tau \gtrless 0$. Thus, we write

$$\mathbf{I}(0, \mu_l, \mu_r; \tau, \mu) = A_+\mathbf{\Phi}_+(\mu)e^{-\tau/\eta_0} + \int_0^1 [A_1(\eta)\mathbf{\Phi}_1(\eta, \mu) + A_2(\eta)\mathbf{\Phi}_2(\eta, \mu)]e^{-\tau/\eta}\, d\eta, \quad \tau > 0, \tag{52a}$$

$$\mathbf{I}(0, \mu_l, \mu_r; \tau, \mu) = -A_-\mathbf{\Phi}_-(\mu)e^{\tau/\eta_0} - \int_{-1}^0 [A_1(\eta)\mathbf{\Phi}_1(\eta, \mu) + A_2(\eta)\mathbf{\Phi}_2(\eta, \mu)]e^{-\tau/\eta}\, d\eta, \quad \tau < 0, \tag{52b}$$

where the negative signs appearing in Eq. (52b) were included for convenience. Applying the "jump" boundary condition given by Eq. (51), we obtain the full-range expansion

$$\frac{1}{\mu}\begin{vmatrix} s_l\delta(\mu - \mu_l) \\ s_r\delta(\mu - \mu_r)\end{vmatrix} = A_+\mathbf{\Phi}_+(\mu) + A_-\mathbf{\Phi}_-(\mu) + \int_{-1}^1 A_1(\eta)\mathbf{\Phi}_1(\eta, \mu)\, d\eta + \int_{-1}^1 A_2(\eta)\mathbf{\Phi}_2(\eta, \mu)\, d\eta, \quad \mu,\ \mu_l,\ \text{and}\ \mu_r \in (-1, 1). \tag{53}$$

It is clear that all required expansion coefficients can now be obtained simply by taking scalar products of Eq. (53) with the appropriate adjoint vectors. Thus,

$$A_\pm = \frac{1}{M_\pm}\int_{-1}^1 \mathbf{\Phi}_\pm^{T\dagger}(\mu)\begin{vmatrix} s_l\delta(\mu - \mu_l) \\ s_r\delta(\mu - \mu_r)\end{vmatrix} d\mu \tag{54a}$$

and

$$A_i(\eta) = \frac{1}{M(\eta)}\int_{-1}^1 \mathbf{\Phi}_i^{T\dagger}(\eta, \mu)\begin{vmatrix} s_l\delta(\mu - \mu_l) \\ s_r\delta(\mu - \mu_r)\end{vmatrix} d\mu, \quad i = 1 \text{ or } 2. \tag{54b}$$

Since the integrals above are elementary, we write the expanded form only for the discrete coefficient:

$$A_\pm = \frac{3c\eta_0}{2M_\pm}\left(s_l\frac{[\Lambda_2(\eta_0)(1 - \mu_l^2) + 2\eta_0^2(1 - c)]}{\eta_0 \mp \mu_l} + s_r\frac{2\eta_0^2(1 - c)}{\eta_0 \mp \mu_r}\right). \tag{55}$$

Having established all of the unknown expansion coefficients, we consider the construction of the Green's function to be completed.

VI. DISCRETE EIGENVALUES

One of the most important parameters in the above formalism is the discrete eigenvalue η_0. As pointed out in Sec. II, η_0 is a real number greater than unity and is the positive zero of the dispersion function $\Lambda(z)$ given by Eq. (11). Sobolev[10] has calculated η_0 for several values of c. In Table I below, we present values for the discrete eigenvalue as a function of the single-scatter albedo c; for convenience, we include those values reported by Sobolev.[10] The calculation was performed on the IBM 360 Model 75 digital computer, using Newton's iteration technique,[25] and the results are believed to be accurate to within $\pm 5 \times 10^{-7}$.

According to the argument principle,[21] the change in the argument as z traverses some closed contour of a function analytic inside that contour is 2π times the number of enclosed zeros. The dispersion function given by Eq. (11) is analytic in the entire plane cut from -1 to 1, and it can be shown that

$$\lim_{z\to\infty} \Lambda(z) = 8(1 - c)(1 - \tfrac{7}{10}c). \tag{56}$$

We choose as our contour one part designated γ which encompasses, but is arbitrarily close to, the cut, and a second part termed R at infinity. Noting Eq. (56), we see that the change in the argument of $\Lambda(z)$ as z traverses R is zero. On γ, we must consider the boundary values of $\Lambda(z)$, namely,

$$\Lambda^\pm(\mu) = -1 + 9(1 - \mu^2)^2[\lambda_0^2(\mu) - \tfrac{1}{4}\pi^2c^2\mu^2] + 3\mu^2(1 - c)[4\lambda_0(\mu) - 3\mu^2(1 - c) + 2] \pm i\pi c\mu[9(1 - \mu^2)^2\lambda_0(\mu) + 6\mu^2(1 - c)]. \tag{57}$$

Since $\Lambda(-z) = \Lambda(z)$ and the complex conjugate of $\Lambda^+(\mu)$ equals $\Lambda^-(\mu)$, it is sufficient to determine the

TABLE I. Table of η_0.

c	η_0	
	Sobolev[10]	Present work
0.1		1.000001
0.2		1.000709
0.3		1.007230
0.4		1.025904
0.5		1.062363
0.581	1.11	1.110624
0.6		1.125231
0.7		1.232743
0.712	1.25	1.250329
0.798	1.43	1.427842
0.8		1.433478
0.861	1.67	1.665949
0.9		1.924622

argument change of $\Lambda^{+}(\mu)$ along that part of γ for which $\mu \in (0, 1)$; the total change will then be four times the calculated value. An investigation of both the imaginary and real parts of $\Lambda^{+}(\mu)$, $\mu \in (0, 1)$, reveals a change in the argument of π. Thus, the total change will be 4π and the number of enclosed zeros two. Further, since $\Lambda^*(z) = \Lambda(z^*)$ and $\Lambda(z) = \Lambda(-z)$, it follows that the zeros occur as a $\pm$ pair, which upon closer inspection can be shown to be real for $c \in (0, 1)$ and to coalesce at infinity for $c = 1$. Here the symbol * denotes complex conjugate.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Dr. E. E. Burniston, Department of Mathematics, North Carolina State University, and Dr. P. F. Zweifel, Department of Physics, Virginia Polytechnic Institute, for offering several helpful suggestions during the course of this investigation.

* This work was supported in part by the National Science Foundation through Grant GK-11935.

[1] S. Chandrasekhar, Astrophys. J. **103**, 351 (1946).
[2] S. Chandrasekhar, Astrophys. J. **105**, 164 (1947).
[3] E. P. Wigner, "Problems of Nuclear Reactor Theory," *Proceedings of the Symposium on Applied Mathematics* (American Mathematical Society, Providence, R.I., 1961), Vol. XI.
[4] G. I. Bell and W. B. Goad, Nucl. Sci. Eng. **23**, 380 (1965).
[5] K. L. Coulson, J. V. Dave, and Z. Sekera, *Tables Related to Radiation Emerging from a Planetary Atmosphere with Rayleigh Scattering* (U. of California P., Berkeley, Cal., 1960).
[6] J. V. Dave and W. H. Walker, Astrophys. J. **144**, 798 (1966).
[7] J. V. Dave, J. Atmospheric Sci. **22**, 273 (1965).
[8] P. S. Shieh and C. E. Siewert, Astrophys. J. **155**, 265 (1969).
[9] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[10] V. V. Sobolev, *A Treatise on Radiative Transfer* (Van Nostrand, Princeton, N.J., 1963).
[11] G. M. Simmons, thesis, Stanford University, 1966.
[12] T. W. Mullikin, "Multiple Scattering of Partially Polarized Light," in *SIAM-AMS Proceedings* (American Mathematical Society, Providence, R.I., 1969), Vol. I.
[13] T. W. Mullikin, Astrophys. J. **145**, 886 (1966).
[14] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[15] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[16] S. A. Mourad and C. E. Siewert, Astrophys. J. **155**, 555 (1969).
[17] S. A. Mourad, Ph.D. thesis, North Carolina State University, 1970.
[18] Z. Sekera, The RAND Corporation, Technical Report R-413-PR, 1963.
[19] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[20] C. E. Siewert and P. F. Zweifel, J. Math. Phys. **7**, 2092 (1966).
[21] R. V. Churchill, *Complex Variables and Applications* (McGraw-Hill, New York, 1960).
[22] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[23] F. D. Gakhov, *Boundary Value Problems*, I. N. Sneddon, Ed. (Addison-Wesley, Reading, Mass., 1966).
[24] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[25] F. B. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill, New York, 1956).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Proper Orientation of Space-Time

M. Enosh
Department of Physics and Astronomy, Tel-Aviv University
Tel-Aviv, Israel

(Received 23 December 1969)

An observer determining space-time around his world line by various measurements may encounter the problem of not being able to define a unique time order and a spatial orientation along world lines of other objects. The problem is discussed and resolved with the aid of geometrical considerations.

1. INTRODUCTION

Suppose that an observer is familiar with space-time[1] according to the general theory of relativity and knows his own world line in it. Suppose also that he knows the time order on his world line as well as the spatial orientation on it; that is, given any three ordered independent vectors orthogonal on his world line at a certain event, he knows whether or not they form a right-handed triad. If he now observes and determines the world line of a second observer far away in space-time, he (the first observer) can also compute the separation (proper time of the second observer) between any two events in the history of the second observer, using the well-known formula[2]

$$ds^2 = c^2\, d\tau^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu.$$

However, in spite of the first observer's complete knowledge of space-time, it does not seem that he can always define a time order and a spatial orientation on the world line of the second observer, if he wants to preserve a certain intuitional argument change of $\Lambda^+(\mu)$ along that part of γ for which $\mu \in (0, 1)$; the total change will then be four times the calculated value. An investigation of both the imaginary and real parts of $\Lambda^+(\mu)$, $\mu \in (0, 1)$, reveals a change in the argument of π. Thus, the total change will be 4π and the number of enclosed zeros two. Further, since $\Lambda^*(z) = \Lambda(z^*)$ and $\Lambda(z) = \Lambda(-z)$, it follows that the zeros occur as a $\pm$ pair, which upon closer inspection can be shown to be real for $c \in (0, 1)$ and to coalesce at infinity for $c = 1$. Here the symbol * denotes complex conjugate.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Dr. E. E. Burniston, Department of Mathematics, North Carolina State University, and Dr. P. F. Zweifel, Department of Physics, Virginia Polytechnic Institute, for offering several helpful suggestions during the course of this investigation.

* This work was supported in part by the National Science Foundation through Grant GK-11935.

[1] S. Chandrasekhar, Astrophys. J. **103**, 351 (1946).
[2] S. Chandrasekhar, Astrophys. J. **105**, 164 (1947).
[3] E. P. Wigner, "Problems of Nuclear Reactor Theory," *Proceedings of the Symposium on Applied Mathematics* (American Mathematical Society, Providence, R.I., 1961), Vol. XI.
[4] G. I. Bell and W. B. Goad, Nucl. Sci. Eng. **23**, 380 (1965).
[5] K. L. Coulson, J. V. Dave, and Z. Sekera, *Tables Related to Radiation Emerging from a Planetary Atmosphere with Rayleigh Scattering* (U. of California P., Berkeley, Cal., 1960).
[6] J. V. Dave and W. H. Walker, Astrophys. J. **144**, 798 (1966).
[7] J. V. Dave, J. Atmospheric Sci. **22**, 273 (1965).
[8] P. S. Shieh and C. E. Siewert, Astrophys. J. **155**, 265 (1969).
[9] S. Chandrasekhar, *Radiative Transfer* (Oxford U.P., London, 1950).
[10] V. V. Sobolev, *A Treatise on Radiative Transfer* (Van Nostrand, Princeton, N.J., 1963).
[11] G. M. Simmons, thesis, Stanford University, 1966.
[12] T. W. Mullikin, "Multiple Scattering of Partially Polarized Light," in *SIAM-AMS Proceedings* (American Mathematical Society, Providence, R.I., 1969), Vol. I.
[13] T. W. Mullikin, Astrophys. J. **145**, 886 (1966).
[14] K. M. Case, Ann. Phys. (N.Y.) **9**, 1 (1960).
[15] C. E. Siewert and S. K. Fraley, Ann. Phys. (N.Y.) **43**, 338 (1967).
[16] S. A. Mourad and C. E. Siewert, Astrophys. J. **155**, 555 (1969).
[17] S. A. Mourad, Ph.D. thesis, North Carolina State University, 1970.
[18] Z. Sekera, The RAND Corporation, Technical Report R-413-PR, 1963.
[19] C. E. Siewert and P. F. Zweifel, Ann. Phys. (N.Y.) **36**, 61 (1966).
[20] C. E. Siewert and P. F. Zweifel, J. Math. Phys. **7**, 2092 (1966).
[21] R. V. Churchill, *Complex Variables and Applications* (McGraw-Hill, New York, 1960).
[22] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, The Netherlands, 1953).
[23] F. D. Gakhov, *Boundary Value Problems*, I. N. Sneddon, Ed. (Addison-Wesley, Reading, Mass., 1966).
[24] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967).
[25] F. B. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill, New York, 1956).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Proper Orientation of Space-Time

M. Enosh
Department of Physics and Astronomy, Tel-Aviv University
Tel-Aviv, Israel

(Received 23 December 1969)

An observer determining space-time around his world line by various measurements may encounter the problem of not being able to define a unique time order and a spatial orientation along world lines of other objects. The problem is discussed and resolved with the aid of geometrical considerations.

1. INTRODUCTION

Suppose that an observer is familiar with space-time[1] according to the general theory of relativity and knows his own world line in it. Suppose also that he knows the time order on his world line as well as the spatial orientation on it; that is, given any three ordered independent vectors orthogonal on his world line at a certain event, he knows whether or not they form a right-handed triad. If he now observes and determines the world line of a second observer far away in space-time, he (the first observer) can also compute the separation (proper time of the second observer) between any two events in the history of the second observer, using the well-known formula[2]

$$ds^2 = c^2\, d\tau^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu.$$

However, in spite of the first observer's complete knowledge of space-time, it does not seem that he can always define a time order and a spatial orientation on the world line of the second observer, if he wants to preserve a certain intuitional

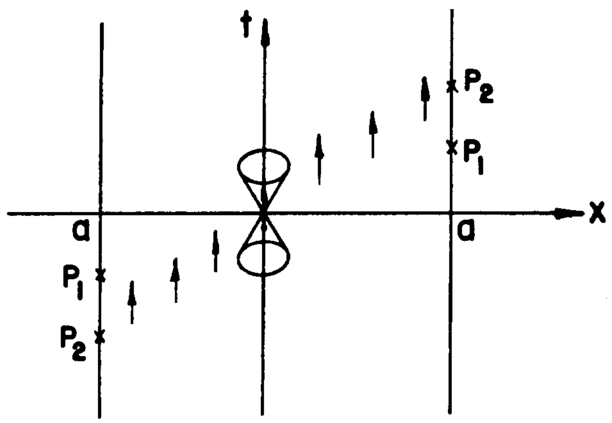


FIG. 1.

continuity (to be defined exactly in the following two paragraphs).

This is easily seen in the following:

Example (*Fig. 1*): Space-time is flat and can be covered by a single system of coordinates (which are not admissible everywhere)

$$-\infty < t < \infty, \quad -a < x < a,$$
$$-\infty < y < \infty, \quad -\infty < z < \infty,$$

and we identify the two hypersurfaces $x = \pm a$ in the following way: (t, a, y, z) and $(-t, -a, y, z)$ represent the same event. The metric is

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2.$$

The world line of the first observer is $x = y = z = 0$, and the world line of the second observer is $x = a$, $y = z = 0$.

In Fig. 1 we see a trial to fix a time order on the world line of the second observer, assuming that a time order is given on the first observer's world line. This time order is represented by the arrow which is drawn at the origin of the figure. The other drawn arrows should point to the future. However, this trial fails. We see a pair of events P_1 and P_2 in the history of the second observer. According to the identification of the hypersurfaces $x = \pm a$, this pair appears twice in the figure. On the right we see that P_2 is after P_1; however, on the left P_2 is before P_1.

Below we shall see that the above-mentioned problem is connected with limitations (due to physical reasons) on space-time, even before knowing anything about the field equations. The following discussion should be of essential importance for a physicist who believes that the theory ought to fix a time order and a spatial orientation on every timelike world line; it may be of informative significance for a physicist who does not.

2. THE BASIC ASSUMPTION: PROPERLY ORIENTABLE SPACE-TIME

In special relativity, an orientation in space and time is determined by a set of ordered tetrads, which is an equivalence class of the proper Lorentz group (Lp). Two ordered tetrads are "properly equivalent" if they are connected by proper Lorentz transformation. The timelike member of every such tetrad points to the future, and the remaining ordered triad is right handed.

We shall accept this assumption also in general relativity. We define (remembering that tetrad is a concept which is defined by using the metric components only) the following:

Definition 1: "A properly oriented (PO) event" is an event with an equivalence class of Lp given at it.

Every event may be provided with four different proper orientations. It is easy to prove the following lemma.

Lemma 1: Let $P(u)$ be a continuous curve [possibly with $P(u_1) = P(u_2)$ for $u_1 \neq u_2$], and let ${}_{(\alpha)}\xi^\mu(u)$ ${}_{(\alpha)}\zeta^\mu(u)$ be two oriented tetrad fields continuous in u at $P(u)$. Then ${}_{(\alpha)}\xi^\mu(u)$ and ${}_{(\alpha)}\zeta^\mu(u)$ are properly equivalent for every u *if and only if* they are properly equivalent at some $u = u_0$. Moreover, the Lorentz transformations connecting ${}_{(\alpha)}\xi^\mu(u)$ and ${}_{(\alpha)}\zeta^\mu(u)$ for different values of u are properly equivalent, that is, they differ from each other by a proper Lorentz transformation, at most.

Definition 2: "A properly oriented (PO) curve" $P(u)$ is a continuous curve such that, for every u, a proper orientation is given at $P(u)$ [but possibly $u_1 \neq u_2$, $P(u_1) = P(u_2)$, and the proper orientations given at u_1 and u_2 may be different], and there exists a continuous tetrad field ${}_{(\alpha)}\xi^\mu(u)$ belonging to the equivalence class of Lp at u.

From Lemma 1 we have that the proper orientation of a PO curve $P(u)$ is determined by the proper orientation at an arbitrary $u = u_0$. For a closed curve, $P(u_1) = P(u_2)$, we have also that the proper orientations at u_1 and u_2 are always (for any proper orientation of the curve) the same or always different. This enables us to define the following:

Definition 3: "A properly orientable (POA) space-time" is one in which every closed curve, $P(u_1) = P(u_2)$, possesses the property that for every proper orientation given on it, the proper orientations at u_1 and u_2 are equal.

Everyday experience and our knowledge of special relativity motivate us to assume the following:

The basic assumption: In space-time which describes the physical world according to general relativity, the events are PO events, and the curves equipped with these proper orientations are PO curves. We shall call such space-times "properly oriented."

In a PO space-time there exist natural time order and space orientation along every timelike curve.

It is easily seen that every PO space-time is POA and, conversely, every POA space-time can be provided with four different proper orientations which are uniquely determined by fixing a proper orientation at an arbitrary event. Therefore, if we accept the basic assumption, we have to confine ourselves to POA space-times, and we have to check every proposed model of space-time according to this limitation.

In the next paragraph we discuss a characteristic property of POA space-times. Some relevant classes of space-time are discussed in the last paragraph.

3. TIME-ORIENTABLE SPACE-TIME

In special relativity an orientation in time is determined by the set of vectors pointing to the future. We define the following in general relativity:

Definition 4: "A time-oriented (TO) event" is an event with an equivalence class of the vectors pointing to the future (see below).

The future-pointing vectors and the past-pointing vectors are the two time-equivalence classes. The time-equivalence relation between timelike or null vectors is actually an equivalence relation, which is defined by terms of the metric components only. Every event may be provided by two different time orientations. The following lemma is easily seen to be correct:

Lemma 2: $P(u)$ is a continuous curve. $\xi^\mu(u)$ and $\zeta^\mu(u)$ are two continuous vector fields having the properties[3]

$$\xi^\mu(u) \neq 0, \qquad \zeta^\mu(u) \neq 0,$$
$$\xi^\mu(u)\xi_\mu(u) \geq 0, \quad \zeta^\mu(u)\zeta_\mu(u) \geq 0.$$

Then $\xi^\mu(u)$ and $\zeta^\mu(u)$ are time equivalent for every u *if and only if* they are time equivalent at some $u = u_0$.

This enables us to define the following:

Definition 5: "A time-orientable (TOA) space-time" is one in which, for every closed curve, $P(u_1) = P(u_2)$, and for every continuous vector field $\xi^\mu(u)$ defined on it, $\xi^\mu(u_1)$ and $\xi^\mu(u_2)$ are time equivalent, provided that $\xi^\mu(u) \neq 0$ and $g_{\mu\nu}(P(u))\xi^\mu(u)\xi^\nu(u) \geq 0$ for every u.

(According to Lemma 2, it is sufficient to find along every closed curve only one vector field $\xi^\mu(u)$ having the above-mentioned properties for proving that space-time is TOA.) The determinant of every proper Lorentz transformation matrix is positive. This immediately leads to the following important theorem:

Theorem 1: A space-time is POA if and only if it is TOA and its differential manifold is orientable.

For the meaning of orientable differential manifold see, for example, Ref. 4 (p. 43). (The space-time of the example in the introduction is not TOA and possesses also unorientable manifold.)

4. SPACE-TIME CLASSES OF PRACTICAL IMPORTANCE

It is well known[5] that every simply connected differential manifold is necessarily orientable.

Every space-time is locally properly orientable. Using this, we can prove (using the usual pure-mathematical methods):

Theorem 2: A simply connected space-time is POA.

Among the multiply connected space-times, we can find examples with the following properties:

(1) TOA space-time with unorientable differential manifold,

(2) non-TOA space-time with unorientable differential manifold (in the Introduction),

(3) non-TOA space-time with orientable differential manifold.

For checking space-times with orientable differential manifolds the following discussion will sometimes be of use (recalling Theorem 1).

The tangent vector on a C^1 curve is a continuous vector field along this curve. Therefore, in a TOA space-time, a curve $x(u)$, $0 \leq u \leq 1$, having the following properties does not exist:

(1) $x(u)$ is a closed curve, $x(0) = x(1)$;

(2) $x(u)$ is piecewise C^1;

(3) $\dfrac{dx}{du} \neq 0$; $\quad g_{\mu\nu}\dfrac{dx^\mu}{du}\dfrac{dx^\mu}{du} \geq 0$;

(4) at an *odd* number of values of u, including $u = 0$, left and right derivatives exist and are not time equivalent. (Such a curve is drawn in Fig. 2.)

The nonexistence of such curves is usually also a sufficient condition for space-time to be TOA, but

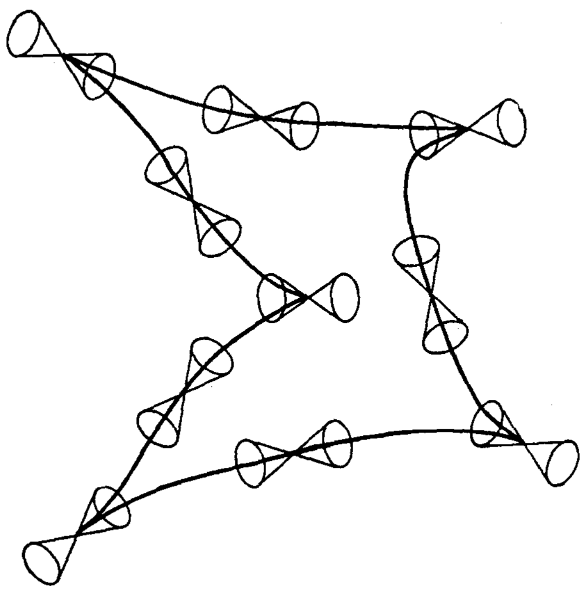


FIG. 2.

it is not an applicable condition, because in a given space–time we can usually approximate all closed curves by curves of the above-mentioned type, except for the restriction (4). In the following theorem we confine ourselves to less general space–times and find more applicable sufficient condition for time orientability of these space–times. (This still includes the example given in the Introduction.)

Theorem 3: Suppose that a space–time has the following properties:

(a) There exists a timelike line Γ equipped with a time-order.

(b) No event horizon with respect to Γ exists.

(c) A continuity assumption: For every curve $X(u)$, a continuous function $X(u, v)$ exists such that $X(u, 0) = X(u)$, $0 \leq v \leq 1$ and $X(u, 1) \in \Gamma$, and $\tilde{X}_u(v) = X(u, v)$ is a curve which may be the history of information passed from $X(u)$ to Γ. [Therefore, $\partial X(u, 1)/\partial v$ points to the future according to the time order given on Γ.] Then space–time is TOA *if and only if* a piecewise C^1 closed curve, $x(0) = x(1)$, having the following properties does not exist:

(1) $0 \neq \dfrac{dx}{du}$, $\quad g_{\mu\nu}\dfrac{dx^\mu}{du}\dfrac{dx^\nu}{du} \geq 0$;

(2) left and right derivatives of $x(u)$ are time equivalent (except at $u = 0$, $u = 1$);

(3) $dx(0)/du$ and $dx(1)/du$ are time inequivalent. (Such a curve is drawn in Fig. 3.)

Remark: In particular, the condition is fulfilled if closed timelike or null curves do not exist at all.

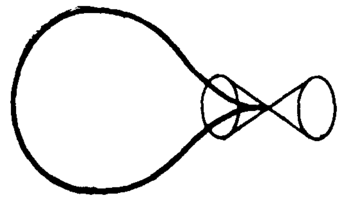


FIG. 3.

Proof: The condition is obviously necessary. For proving that the condition is sufficient, it is sufficient to define a time orientation at every event so that a TO space–time is constructed. Using (a) and (b), we define a time orientation at each event in space–time. The definition is consistent because of the continuity of the tangent vector of the information line and because of Lemma 2 and the sufficient condition given in this direction of the proof. Space–time is now TO because of (c). The proof is completed.

From Theorem 3 it follows that the space–time of the example in the Introduction must contain a closed time like curve. It is an easy exercise to find such curves.

We shall end this article with two remarks:

(1) A space–time in which a global time-coordinate exists is TOA, provided that every two hypersurfaces $t = t_1$ and $t = t_2$ ($t_1 \neq t_2$) have no events in common (no topological identifications are made).

(2) It can be shown easily that, if S' is a hypersurface in a flat odd-dimensional space X, possessing the properties

(a) $x \in S' \Leftrightarrow -x \in S'$ (the x^A are Euclidean coordinates of X),

(b) S' is connected, and

(c) S' is the envelope of a (bounded or unbounded) domain in X,

and if S is constructed by a topological identification of x and $-x$ in S', then S is an unorientable differential manifold.

Therefore, our basic assumption and Theorem 1 rule out some polar de Sitter universes mentioned in Ref. 4 (pp. 257, 263).

ACKNOWLEDGMENT

I would like to thank Dr. A. Kovetz for his encouragement and helpful criticism.

[1] Space–time is a 4-dimensional Riemannian space of a normal hyperbolic type.
[2] Greek indices run from 0 to 3.
[3] The signature is taken to be $(+1, -1, -1, -1)$.
[4] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1964),
[5] M. Greenberg, *Lectures on Algebraic Topology* (Benjamin, New York, 1966), p. 116.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Physical-Region Discontinuity Equation*

JOSEPH COSTER† AND HENRY P. STAPP
Lawrence Radiation Laboratory, University of California, Berkeley, California 94720

(Received 18 March 1970)

A Cutkosky-type formula for the discontinuity around an arbitrary physical-region singularity is derived from precisely formulated S-matrix principles.

1. INTRODUCTION

We shall derive the following result: The discontinuity of the S matrix S around any physical-region singularity surface is given by a Cutkosky-type formula obtained by replacing each vertex of the corresponding diagram D by the associated (physical-region) S matrix, replacing the set of lines α joining each pair of vertices of D by a function S_α^{-1}, and integrating over all the (topologically inequivalent) mass-shell values of the variables corresponding to the intermediate lines. The function S_α^{-1} is defined by $S_\alpha S_\alpha^{-1} = I_\alpha$, where S_α is the restriction of S to the space corresponding to the set of lines α and I_α is the corresponding restriction of unity.

This rule gives the discontinuity for S itself. The result for the connected part is obtained by retaining only connected graphs. Then the disconnected contributions to the S's associated with the various vertices will vanish, in general, due to the constraints imposed by the extra conservation laws. But where these disconnected contributions do not vanish they must be included, in order to obtain the complete discontinuity around the surface in question. The discontinuity formula described here holds at all points where the full discontinuity is the sum of contributions associated with the diagram D and diagrams that can be reduced to D by contraction of lines. Thus, it continues to hold at points where Landau surfaces corresponding to diagrams that contract to D intersect the Landau surface corresponding to D. This makes the form particularly useful for substituting into other expressions, since one does not have to take special account of these particular points where related Landau surfaces touch.

The discontinuity formula stated above is similar to the one obtained by Cutkosky.[1] However, his formula was incomplete because important questions concerning the sheet structure were not answered.[2] Also, his derivation depended on perturbation theory. The present results are derived within the mass-shell S-matrix framework and give the discontinuity in terms of the actual physical-region scattering functions. This confirms earlier indications[3,4] that the physical-region discontinuities are completely determined by general S-matrix principles.[5] They do not depend on the special properties (such as locality) exhibited by the terms of the perturbation theory.

In Sec. 2 the results needed from earlier works are summarized. The discontinuity formula is derived in Sec. 3 by using an infinite-series (mass-shell) expansion for S. Some properties of S_α^{-1} are discussed in Sec. 4.

A derivation not based on the infinite series for S is given in Sec. 5, for the case of "leading singularities." A leading singularity is one such that the set of particles corresponding to the set of lines α joining any pair of vertices of D is a "leading set." A leading set of particles is a set that cannot make a transition to a set having a lower sum of rest masses. We hope to give later a derivation for the case of nonleading singularities that is not based on the infinite series for S.

In the final section our work is compared with other works in the field.

2. BASIC TOOLS

A. Cluster Decomposition

The S matrix is the transition matrix from "in" to "out." Linearity ensures that the transition matrix from "out" to "in" is S^{-1}. We do not use unitarity ($S^{-1} = S^\dagger$). (All that is used in S-matrix derivations of discontinuity equations are the cluster properties and $i\epsilon$ rules of S and S^{-1}: it is not important that S^{-1} be $S^\dagger$.)

The cluster decompositions of S and S^{-1} are conveniently represented by a diagram notation[3]: A box with a plus (minus) sign inside represents $S(S^{-1})$; a bubble (i.e., circle) with a plus (minus) sign inside represents the connected part of $S(S^{-1})$. The left side of each box or bubble is the origin of a set of leftward-directed lines, and the right side is the terminus of such a set. Each line j is associated with a physical-particle variable, which is a set (p_j, μ_j, t_j) consisting of a particle-type index t_j, a spin (magnetic) quantum number μ_j, and a real positive-energy mass-shell 4-vector p_j.

The cluster decomposition of $S(S^{-1})$ is represented by writing each plus (minus) box as a sum of columns of plus (minus) bubbles, the sum being over all topologically distinct ways that the lines originating and terminating on the box can be partitioned among bubbles of a column, with each bubble having at least one incoming and one outgoing line.

The connected parts of S and S^{-1} divided by the over-all conservation δ function are the *scattering functions* S_c and S_c^-, respectively.

B. Bubble Diagram Functions

The cluster decompositions of S and S^{-1} induce corresponding decompositions of quantities like SS^{-1}, $SS^{-1}S$, etc. The rule for computing such a product is to first draw all topologically distinct bubble diagrams B composed of the appropriate number of columns of the appropriately signed bubbles. The lines originating on the bubbles of one column terminate on those of the column standing to its left, if there is one. For each such B, one constructs a corresponding function F^B by summing over all physical values of the variables (p_i, μ_i, t_i) for each internal line i, subject to the constraint that topologically equivalent contributions be counted only once. The function being calculated is precisely the sum of the functions F^B defined in this way.[3] (For fermions some signs must be considered.)

Two contributions to F^B are topologically equivalent if and only if the corresponding diagrams, with each line j identified by a corresponding variable (p_j, u_j, t_j), can be continuously distorted into each other with the external end points of the external lines held fixed. Each bubble is identified as to its column, and the distortions must leave each bubble in its own column. (Alternatively, one must keep all the "trivial" bubbles having only one incoming and one outgoing line. These bubbles are often omitted because they do not affect the value of the integral, except in this matter of counting.)

C. Macrocausality

Macroscopic particle phenomena have a characteristic space-time structure. If effects of long-range interactions and massless particles are ignored, then particles move along straight space-time trajectories except when they come close to other particles. A quantitative description of the phenomena is provided by the Newton-Einstein laws of motion. These laws assign to each particle j a momentum-energy vector p_j that is directed along its space-time trajectory and that satisfies the mass-shell constraint $p_j^2 = m_j^2$. Momentum-energy is conserved and is exchanged between particles only when they are close to each other; one imagines momentum-energy to be transmitted by a short-range interaction.

If one requires this space-time structure of macroscopic phenomena to emerge from S-matrix theory, in appropriate classical, macroscopic limits, and demands also that classical estimates based on short-range interactions should become valid in these limits, at least to order of magnitude, then certain physical-region analyticity properties follow. These include the cluster decomposition property described above, and also the properties described in the following two sections.

D. The Positive-α Rule

The first important consequence of the macrocausality condition is that the physical-region singularities of the scattering functions $S_c^{\pm}$ are confined to positive-α Landau surfaces[6] associated with connected diagrams.[7]

Landau surfaces are associated with Landau diagrams. A Landau diagram D is a diagram that represents a classical multiple-scattering process with point interactions. It consists of a set of leftward directed line segments j that meet at vertices v. Each line j is associated with a real momentum-energy vector p_j that satisfies the mass-shell constraint

$$p_j^2 - m_j^2 = 0, \quad p_j^0 > 0, \tag{2.1a}$$

where m_j is the mass of the particle associated with line j. Momentum-energy is conserved at each vertex v:

$$\sum_{\text{into } v} p_j - \sum_{\text{out of } v} p_j = 0. \tag{2.1b}$$

The vector Δ_i from the space-time origin of internal line i of D to its space-time terminus must be directed along its momentum-energy, i.e., for some scalar α_i, one has

$$\Delta_i = \alpha_i p_i \,. \tag{2.1c}$$

Finally, the sum of the space-time displacements Δ_i around any closed loop of internal lines of D must add to zero:

$$\sum_l \pm\Delta_i \equiv \sum_l \pm\alpha_i p_i = 0. \tag{2.1d}$$

Here, the $\pm$ sign is plus if the loop l is directed along Δ_i and minus otherwise.

These equations express the constraints on the multiple-scattering diagram D imposed by classical relativistic particle mechanics. They are called the *Landau equations*. The Landau surface $L(D)$ is the set of external $P \equiv (p_1, \cdots, p_n)$ that are compatible with the Landau equations associated with diagram D. The trivial solution with all $\alpha_i = 0$ is not accepted.

Physical particles carry positive energy forward in time. The α_i must therefore be positive:

$$\alpha_i \geq 0. \tag{2.2}$$

The subset of $L(D)$ that allows a solution of the Landau equations (2.1) subject to the positive-α condition (2.2) is denoted by $L^+(D)$, and is called a positive-α Landau surface. The positive-α rule says that the scattering functions $S_c^{\pm}(P)$ are analytic at all physical points not lying on the union of positive-α surfaces

$$L^+ \equiv \bigcup L^+(D). \tag{2.3}$$

The scattering functions $S_c^{\pm}$ are defined only on the mass shell $\mathcal{M}$, which is defined by the mass-shell constraints (2.1a) and the over-all momentum–energy conservation law. Thus the ordinary definition of analyticity does not apply. The appropriate definition is given in Refs. 5, 7, and 8.

Certain general properties of the set L^+ are used in formulating the $i\epsilon$ rule. These are now described.

A given surface $L^+(D)$ generally coincides with the surfaces $L^+(\bar{D})$ of an infinite set of other diagrams $\bar{D}$. These arise in a trivial way: If a set of internal lines of D all originate at the same vertex v' and all terminate at the same vertex v'', then the Landau equation requires them all to be moving along together, relatively at rest. Thus, they can undergo trivial forward scatterings upon each other without affecting the kinematic relations. Any number of these trivial forward scatterings can occur. This leads to an infinite set of diagrams $\bar{D}$ such that $L^+(\bar{D}) = L^+(D)$.

It is convenient to introduce diagrams that do not have these trivial forward scattering vertices. A *basic diagram* D_β is a Landau diagram that has no part that (i) is connected to the rest of the diagram at only two vertices, (ii) contains more than two vertices, and (iii) contains no external lines. Every $L^+(D)$ is confined to the $L^+(D_\beta)$ of some corresponding basic diagram D_β. Thus one can write

$$L^+ \equiv \bigcup L^+(D) = \bigcup L^+(D_\beta). \tag{2.3$'$}$$

Only a finite number of D_β have $L^+(D_\beta)$ that enter any bounded portion of the physical region.[9]

The representation of L^+ is further simplified by introducing "basic surfaces," defined as follows: Let $\mathcal{M}_0$ represent the part of the mass shell where two or more initial momentum–energy vectors p_j are parallel or two or more final p_j are parallel. Then for any Landau diagram D the set $L_0^+(D)$ is that part of $L^+(D) - \mathcal{M}_0$ such that the Landau equations for $L^+(D)$ have no solution with any $\alpha_i = 0$.

It is clear that any point on $L^+(D) - \mathcal{M}_0$ that is not on $L_0^+(D)$ must lie on the $L_0^+(D')$ of a *contraction* D' of D, constructed by contracting to points and removing from D the lines corresponding to $\alpha_i = 0$. Thus, L^+ can be written as

$$L^+ = \bigcup L_0^+(D_B) + \mathcal{M}_0. \tag{2.3$''$}$$

The importance of this representation lies in the fact that $L_0^+(D_\beta)$ is a real codimension-1 analytic submanifold of the mass shell $\mathcal{M}$.[8] That is, each point $\bar{P}$ of $L_0^+(D_\beta)$ has a mass-shell neighborhood $N(\bar{P})$ such that inside $N(\bar{P})$ the set $L_0^+(D_\beta)$ coincides with the set $f = 0$, where f is a real analytic function of the local real analytic coordinates of the mass shell at $\bar{P}$ (see, e.g., Refs. 7 or 8), and grad $f \equiv \nabla f$ is nonzero in $N(\bar{P})$.

The representation (2.3$''$) shows that $L^+ - \mathcal{M}_0$ is the union of a set of codimension-1 real analytic submanifolds of $\mathcal{M}$, only a finite number of which enter any bounded portion of the physical region. Since $\mathcal{M}_0$ has codimension 3, the set L^+ has codimension 1. (The codimension of $\mathcal{S}$ plus the dimension of $\mathcal{S}$ is the dimension of imbedding space, here $3n - 4$.)

The positive-α rule says, therefore, that $S_c(P)$ is analytic at almost all physical points and that the remaining set L^+ has, apart from the small set $\mathcal{M}_0$, a local representation as the zeros of a finite set of real analytic functions f_i, each having nonzero gradient ∇f_i.

E. The $i\epsilon$ Rule

Macrocausality implies also that the scattering function S_c near any $\bar{P}$ of $L^+ - \mathcal{M}_0$ can be represented as the limit from any direction in the intersection of the upper half-planes Im $f_i > 0$ of the (unique) analytic continuation into this intersection of the function $S_c(P)$ defined on $L^+ - \mathcal{M}_0$. The functions f_i are the functions that define L^+ near $\bar{P}$, and their signs are fixed by the requirement that a formal increase of the masses associated with the internal lines of D by a common scale factor shifts $L_0^+(D)$ in the plus f direction. This sign is known to be independent of the particular diagram D that defines $L_0^+(D)$: All locally coincident surfaces $L_0^+(D)$ can be defined by the same function f (Theorem 7 of Ref. 8).

This $i\epsilon$ rule for S_c is known as the plus $i\epsilon$ rule. The function S_c^- obeys the minus $i\epsilon$ rule, which is the same rule except that the upper half-planes Im $f_i > 0$ are replaced by lower half-planes Im $f_i < 0$.

These rules have content only at those points $\bar{P}$ of $L^+ - \mathcal{M}_0$ for which the appropriate half-planes have a nonempty intersection that contains $\bar{P}$ on its boundary. This property is obviously satisfied for any $\bar{P}$ that lies on only one $L_0^+(D_\beta)$ [or only on several $L_0^+(D_\beta)$ that all locally coincide with one single one]. Such points comprise almost all of $L^+ - \mathcal{M}_0$, since the rest have codimension 2. Thus the $i\epsilon$ rules have content at almost all points of $L^+ - \mathcal{M}_0$.

It is important that the $i\epsilon$ rules have content also at a certain of the remaining points of $L^+ - \mathcal{M}_0$. It is known (Theorem 13, Ref. 8) that the intersection of the upper half-planes corresponding to $\bar{P}$ (on $L^+ - \mathcal{M}_0$) is nonempty, and contains $\bar{P}$ on its boundary, whenever all the D_β with $\bar{P} \in L_0^+(D_\beta)$ are contractions of some single D.

There are, however, some points $\bar{P}$ of $L^+ - \mathcal{M}_0$ such that the intersections of the various upper half-planes associated with $\bar{P}$ are empty near $\bar{P}$. The scattering function S_c cannot be represented near such a $\bar{P}$ as the limit of a single analytic function. To cope with such points, we shall introduce in the next section an independence property, which says, in effect, that singularities associated with unrelated diagrams are independent. This will allow the $i\epsilon$ rule to be applied at all points of $L^+ - \mathcal{M}_0$.

Full technical details concerning the $i\epsilon$ rules are given in Refs. 7 and 8. The intersection of the upper half-planes at $\bar{P}$ is defined, in effect, as the set of mass-shell variations δ that satisfy

$$\operatorname{Im} \delta \cdot \nabla f_i(\bar{G}) > 0,$$

where G is a set of local real analytic coordinates at $\bar{P}$ and $\bar{G} = G(\bar{P})$. (See also Ref. 10.)

The basic tool in the analysis of physical-region singularities is a theorem that extends the positive-α and $i\epsilon$ rules to all bubble diagram functions. This theorem is described next.

F. Fundamental Theorem[11,12]

1. Assumptions of Theorem

(a) *Positive-α Rule.* The physical-region singularities of the scattering functions S_c and S_c^- are confined to the union L^+ of positive-α Landau surfaces.

(b) *Independence Property.* Each point $\bar{P}$ of $L^+ - \mathcal{M}_0$ has a real mass-shell neighborhood $N(\bar{P})$ such that $S_c^\pm(P)$ in $N(\bar{P}) - L^+$ decomposes into a finite sum of terms, one for each basic diagram D_β for which $L^+(D_\beta)$ contains $\bar{P}$. The singularities of the term of $S_c^\pm$ associated with D_β are confined to

$$\hat{L}^\pm(D_\beta) \equiv L^\pm(D_\beta) \cup [\bigcup L^\pm(D_\beta')], \qquad (2.4)$$

where D_β' is any contraction of D_β. Each term obeys a corresponding $i\epsilon$ rule, as is described next. (The justification of the independence property is given in Sec. 2G.)

(c) *The iε Rules.* The individual terms of S_c and S_c^- described in the independence property obey the plus and minus $i\epsilon$ rules, respectively. The upper and lower half-planes for each term are specified by the singularity surfaces occurring in that term alone.

(d) *Technical Assumption.* The singularities at $\mathcal{M}_0$ are not too pathological. (This assumption is discussed in Sec. 2F3.)

2. Conclusions of Theorem

Let B be any connected bubble diagram. Let F^B be the corresponding bubble diagram function. Define

$$F_c^B(P) \equiv F^B(P)/\delta^4\left(\sum_{\text{in}} p - \sum_{\text{out}} p\right). \qquad (2.5)$$

Then the following properties hold:

(a) *Generalized Positive-α Rule.* The physical-region singularities of F_c^B are confined to the union of the Landau surfaces $L^\sigma(D_B)$. A D_B is a Landau diagram constructed by inserting a connected basic Landau diagram D_b for each bubble b of B, with the incoming and outgoing lines of D_b identified in a 1-to-1 fashion with the incoming and outgoing lines of b, respectively. The surface $L^\sigma(D_B)$ is the part of $L(D_B)$ that is compatible with the Landau equations of $L(D_B)$, subject to the constraint that each line i of D_B that is an *internal* line of some D_b must have an α_i that satisfies

$$\alpha_i \sigma_b \geq 0, \qquad (2.6)$$

where σ_b is the sign of b. The (original) lines of B itself, which are external lines of various D_b, have no sign constraint.

(b) *Generalized Independence Property.* Each point $\bar{P}$ of $\bigcup L^\sigma(D_B) - \mathcal{M}_0$ has a real mass-shell neighborhood $N(\bar{P})$ such that F_c^B decomposes on $N(\bar{P}) - \bigcup L^\sigma(D_B)$ into a finite sum of terms, one for each D_B for which $L^\sigma(D_B)$ contains $\bar{P}$. The singularities of the term associated with a given D_B are confined to

$$\hat{L}^\sigma(D_B) = L^\sigma(D_B) \cup [\bigcup L^\sigma(D_B')], \qquad (2.7)$$

where the D_β' are contractions of lines of D_B that are internal lines of some D_b.

(c) *Generalized iε Rule.* The functions $F_c^B(P)$ obey a rule that is completely analogous to the plus $i\epsilon$ rule, except that the upper half-planes at $\bar{P}$ are now defined by using, instead of $f = f(P)$, the functions

$$\sigma_{\bar{P}}(P) \equiv \sum \alpha_i(\bar{P})[p_i(P) - p_i(\bar{P})] \cdot p_i(\bar{P}). \qquad (2.8)$$

There is one such function for each solution at $\bar{P}$ of the Landau equations of $L^\sigma(D_B)$. The $\alpha_i(\bar{P})$ and $p_i(\bar{P})$ are the parameters of the internal lines of D_B corresponding to the solution at $\bar{P}$. The $p_i(P)$ is any set of internal p_i satisfying the conservation law constraints of D_B at P. [The function $\sigma_P(P)$ will not depend on the particular choice of the $p_i(P)$, because of the Landau loop equation.]

The ordinary $i\epsilon$ rules connect the physical-region scattering functions in different sectors of $\mathcal{M} - L^+$. Similarly, the generalized $i\epsilon$ rules connect the "physical-region" functions F_c^B in different sectors of $\mathcal{M} - \bigcup L^\sigma(D_B)$. The *physical-region* functions F^B are defined as integrals over the physical-region scattering functions. These are the functions F^B that occur in the decomposition of the functions SS^{-1}, $SS^{-1}S$, etc.

It might be possible to continue F_c^B from some given sector of $\mathcal{M} - \bigcup L^\sigma(D_B)$ by following, alternatively, different alternative paths around some $L^\sigma(D_B)' - \mathcal{M}_0$. The generalized $i\epsilon$ rule asserts that it definitely is possible to continue through the intersection of the upper planes defined by (2.8), provided the intersection of these upper half-planes is nonempty arbitrarily close to $\bar{P}$. Moreover, the function arrived at on the other side of $L^\sigma(D_B) - \mathcal{M}_0$ will then be precisely the physical-region function F_c^B. Also, an integral over the physical-region function F_c^B can be represented by an integral over a contour distorted infinitesimally away from $\bar{P} \in \bigcup L^\sigma(D_B)$ and into the intersection of the upper half-planes at $\bar{P}$.

By F_c^B we shall, unless otherwise stated, always mean the physical region F_c^B, not some analytic continuation of it; the only continuations considered are the infinitesimal ones specified by the general $i\epsilon$ rules, unless otherwise stated.

The generalized $i\epsilon$ rule has content at $\bar{P}$ of $L^\sigma(D_B) - \mathcal{M}_0$ only if the various upper half-planes at $\bar{P}$ have a nonempty intersection at $\bar{P}$ [i.e., only if there is a $(3n - 4)$-dimensional variation δ in $\mathcal{M}$ satisfying $\operatorname{Im} \delta \cdot \nabla\sigma_P(\bar{P}) > 0$ for all $\sigma_P(P)$ associated with $L^\sigma(D_B)$]. If this intersection is empty at $\bar{P}$, then no continuation past $L^\sigma(D_B)$ is assured at $\bar{P}$.

There are some important points $\bar{P}$ of $L^\sigma(D_B)$ for which the intersection of the upper half-planes is obviously empty. In particular, every point of $L^\sigma(D(B))$ has this property. The diagram $D(B)$ is the particular D_B obtained by replacing each bubble b of B by a point vertex. Since no line of $D(B)$ comes from inside any bubble, there are no constraints on the signs of the $\alpha_i(\bar{P})$. Thus the reversal of all these signs will give another solution. This solution will have the signs of all the functions $\sigma_P(P)$ reversed. Thus the positions of all upper half-planes will be reversed. Thus the intersection of the upper half-planes at $\bar{P}$ will be empty, and the $i\epsilon$ rule will be without content there.

This failure of the $i\epsilon$ analyticity property at points of $L^\sigma(D(B))$ plays a crucial role in what follows. It is related to the breakdown of the definition of F^B at these points. The function F^B is defined as an integral that contains, in effect, a conservation-law δ function for each bubble b of B and a mass-shell δ function for each internal line i of B. A product of δ functions under an integral sign is defined as follows: One transforms to a new set of variables that contains the argument g_j of each δ function as an independent variable, and then omits the integrations on these variables. This definition fails at $\bar{P}$ (i.e., the Jacobian becomes infinite) if the gradients ∇g_j are linearly dependent at $\bar{P}$.

These linear dependence relations turn out to be precisely the Landau loop equations corresponding to $D(B)$. Since the mass-shell and conservation-law constraints are also satisfied, the equations that define the points where F^B is ill-defined are just the Landau equations for $D(B)$, and the corresponding set of points P is the Landau surface $L(D(B)) \equiv L^\sigma(D(B))$.

The function F^B generally does not continue into itself around points of $L(D(B))$. That is, F^B, in different sectors of $\mathcal{M} - L(D(B))$ near $\bar{P}$ of $L(D(B))$, are generally not parts of a single analytic function. In fact, the function F^B is obviously identically zero at points of $\mathcal{M}$ where it is not possible to satisfy simultaneously the various mass-shell and conservation law constraints associated with B. The boundary of this region lies in $L(D(B))$. Furthermore, every point of $L^-(D(B))$ lies on this boundary. Thus F^B can never continue into itself around $L^+(D(B))$, unless it is identically zero.

The portion of $\mathcal{M}$ where it is possible to satisfy all the mass-shell and conservation-law constraints of B is called the *physical region* of B. According to the above remarks, the function F^B is nonzero only in the physical region of B. Moreover, $L^+(D(B))$ lies on the boundary of this region. The sign conventions on the functions f_i are such that the physical region of B near $\bar{P}$ of $L_0^+(D(B))$ is either confined to $L_0^+(D(B))$ or lies on the positive-f side of it.[10] That is, F^B is identically zero on the negative-f side of $L_0^+(D(B))$.

The above-mentioned fact is important in the derivation of the discontinuity formula. It ensures that *all* the terms in the discontinuity formula vanish on the negative-f side of the singularity surface $L_0^+(D_\beta)$ in question. The "principal term" of the discontinuity formula, which is the one such that each vertex v of D_β corresponds to the connected part of the corresponding S, will have its physical region bounded by $L_0^+(D_\beta)$. Generally speaking, the physical regions of the nonprincipal terms will not extend to $L_0^+(D_\beta)$ because of the extra constraints imposed by the extra conservation laws. Thus the nonprincipal terms will generally not contribute to the discontinuity around $L_0^+(D_\beta)$. But if the physical region of some nonprincipal term does reach $L_0^+(D_\beta)$, then this term will contribute to the discontinuity around $L_0^+(D_\beta)$.

3. *The Technical Assumption*

The macrocausality condition does not rule out singularities at $\mathcal{M}_0$. The proof of the theorem requires, however, that the singularities at $\mathcal{M}_0$ be not too pathological. It is known from the boundedness property $S_c[\phi_1, \cdots, \phi_n] \leq \|\phi_1\| \cdots \|\phi\|$, which follows from linearity and the probability interpretation, that the integrals defining F^B do not diverge at $\mathcal{M}_0$. An additional requirement is that the integrals defining the derivative of F^B also be well defined at $\mathcal{M}_0$.

G. Maximal Analyticity

This principle is that $S^{\pm 1}(P)$ has only those singularities that are required by general principles. The full content of this principle, as it applies to physical-region points, is the *independence property* (b): Singularities violating this property are not required to be present; hence they are required to be absent.

The point is this. The positive-α rule and the $i\epsilon$ rules impose certain constraints on the *allowed* singularities, but they do not *require* any singularity actually to be present in S_c or S_c^-. On the other hand, the cluster properties of S and S^{-1}, by themselves, actually require the scattering functions to have singularities.

These arise as follows. Suppose one expresses identities such as $SS^{-1} = I$, $S = SS^{-1}S$, or $S = SS^{-1}SS^{-1}S$, etc., in the form of bubble diagram equations

$$\sum_{B\in\mathcal{B}'} F^B = \sum_{B\in\mathcal{B}''} F^B, \tag{2.9}$$

where $\mathcal{B}'$ and $\mathcal{B}''$ are classes of bubble diagrams. Then the assumption that the S_c and S_c^- are all singularity free gives contradictions: Certain terms of (2.9) will have explicit singularities that cannot be cancelled by any other singularities. Thus the cluster properties of S and S^{-1} definitely require some of the scattering functions to have singularities.

The above argument does not show precisely which singularities are required in S_c and S_c^-. However, it can be extended to do just that. In particular, the various identities (2.9), which follow simply from the cluster properties of S and S^{-1}, supplemented by the conclusions of the fundamental theorem, permit the derivation of a formula for the discontinuity around each physical-region singularity allowed by the positive-α rule. This formula shows that each allowed singularity is also required; i.e., it has a nonzero discontinuity. These required singularities are apparently compatible with the independence property. Thus we have an apparently self-consistent singularity structure that has no singularities that violate the independence property. Thus no singularity that violates this property is required. Then maximal analyticity says none is allowed. Hence, the independence property must hold.

We can now turn to the derivation of the discontinuity formula. It will be convenient to assign to each internal line i of each Landau diagram D a sign σ_i that determines the sign of α_i in the corresponding Landau equations:

$$\alpha_i\sigma_i \geq 0.$$

A diagram that has all $\sigma_i = +1$ is called a positive-α diagram and is denoted by D^+. Thus

$$L(D^+) \equiv L^+(D).$$

3. ITERATIVE SOLUTION

A. Expansion of S

Introducing $R^\pm \equiv S^{\pm 1} - 1$, we obtain

$$R^+ + R^- + R^+R^- = 0. \tag{3.1}$$

The formal iterative solution for R^+ gives

$$R^+ = \sum_{n=1}^{\infty}(-1)^n(R^-)^n. \tag{3.2}$$

Each factor R^- is represented by a sum of columns of minus bubbles, the sum being over all topologically different ways of joining a column of bubbles to the external lines. However, at least one bubble of each column must be nontrivial. (Trivial bubbles are those with just one incoming line and just one outgoing line.)

In the assessment of topological equivalence one considers the bubbles to be confined to particular columns.[3] This means that the three terms shown in Fig. 1 must all be counted.

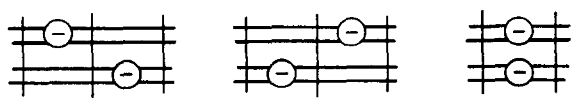


FIG. 1. Three contributions to the expansion of a 4-line S. The vertical lines show the separation into factors R^-. Trivial bubbles have been omitted, since they do not alter the function.

The first two factors have coefficients $(-1)^2 = 1$ in (3.2), whereas the last has coefficient (-1). Thus there is a cancellation and only one term survives.

This result is general: In the expansion (3.2) one needs to count only one of any set of topologically equivalent contributions, where in the assessment of topological equivalence one now disregards both trivial bubbles and the separation of bubbles into columns. The sign of the single surviving term is $(-1)^n$, where n is the number of (nontrivial) minus bubbles of the term.

The bubbles b of the original B are partially ordered by the ordering of the columns in which they lie. If the column identification of the bubbles is removed, then the bubbles are partially ordered only by the requirement that all lines be directed from right to left. For each such partially ordered B^- there remains, after the cancellations, precisely one term F^{B^-}. Thus if the unit contribution is added back to give $S = 1 + R^+$, one obtains[13]

$$S = \sum_{B^-} (-1)^n F^{B^-}. \qquad (3.2')$$

The sum is over all topologically different partially ordered bubble-diagrams B^- having only nontrivial minus bubbles, and n is the number of bubbles of B^-.[13]

The expansion (3.2′) contains in an implicit form an expression for the discontinuities. As one moves across a positive-α threshold, new terms appear in (3.2′). If mixed-α singularities (i.e., singularities corresponding to solutions of Landau equations that require α's of both signs) can be ignored (see Sec. 6 below) and if only one positive-α surface is relevant, then the discontinuity is just the sum of these new terms. This is because any term in (3.2′) that is present below the threshold will, by virtue of the fundamental theorem, continue around any singularity at threshold via the minus $i\epsilon$ rule. This leaves the new terms as the discontinuity. The problem of calculating the discontinuity is then to identify the infinite number of terms that appear in (3.2′) as one crosses the threshold and to combine them into a useful form. The following sections are, in effect, devoted to that end.

B. A Fundamental Identity

Let α be some set of incoming lines of S. A minus bubble in the expansion (3.2′) of S will be called an α bubble if and only if all the incoming lines of that bubble belong to the set α. We define S^α to be the subset of the expansion (3.2′) consisting of all terms having no α bubble. Thus for each term of S^α each line in the set α either ends at a minus bubble that has some incoming line not belonging to α, or it touches no minus bubble at all and is therefore an "unscattered" line (i.e., it is both incoming and outgoing).

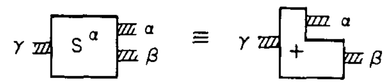
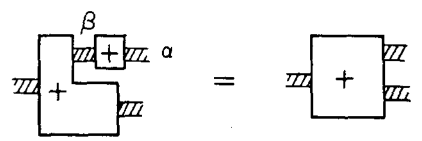


FIG. 2. Diagrammatic representation of S^α. The shaded strips represent arbitrary sets of external lines.

It is convenient to represent S^α by the diagram shown in Fig. 2.

The diagram on the right of Fig. 2 is to be regarded as a representation of a partial sum of terms of the expansion (3.2′). The missing section indicates the absence of all terms having an α bubble.

With this notation a fundamental identity is



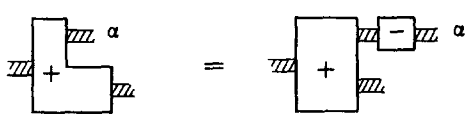
(3.3)

This equation expresses the fact that, if one attaches to S^β the set obtained from the expansion of the small plus box and sums over β, then one obtains the full expansion (3.2′) of S. In particular, all the terms with α bubbles are reinstated, and each one only once.

To prove (3.3), the concept of a *cut* is useful. The lines of the $D(B^-)$ corresponding to any B^- are drawn running from right to left. A *flow line* is a continuous path in D that runs from the extreme right to the extreme left. It consists of an ordered sequence of line segments L_j of D all of which point in the direction of the path. A *cut* is a set of lines that contains at most one line L_j of any flow line. The set of flow lines *defined* by a cut is the set of all flow lines that contain a line contained in the cut. *Equivalent cuts* are cuts that define identical sets of flow lines. A line l_1 *lies left* of l_2 if and only if l_1 lies left of l_2 on some flow line. A cut C_1 *lies left* of a cut C_2 if and only if C_1 is equivalent to C_2, at least one line of C_1 lies left of some line of C_2, and no line of C_2 lies left of any line of C_1. A *leftmost* cut is a cut such that no cut lies left of it.[13,14]

In (3.3) the cut β is the leftmost cut equivalent to α. That no cut lies left of it follows from the definition of S^β. For each fixed β the terms of (3.2′) give, independently, all terms of S^β on the left of β and all terms of the small plus box $S_{\beta\alpha}$ on the right.

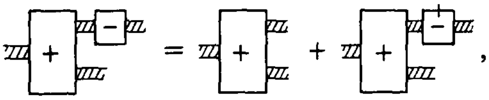
Multiplication of (3.3) by a small minus box on the right gives



(3.3′)

The fact that the combination on the right is equivalent to a sum of bubble diagram functions F^B corresponding to B's having no α bubbles was shown earlier in Ref. 15. There only finite operations were used and the sum was over a finite number of terms. (Both plus and minus bubbles occurred in the B's representing the terms of that finite expression.)

The validity of (3.3′) can be seen directly from the expansion (3.2′). If this expansion is substituted into both terms of the right side of



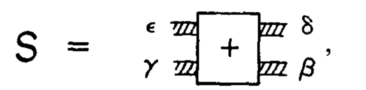
(3.4)

where the slashed box is R^-, one finds an exact cancellation of all terms having an α bubble: Each bubble diagram B^- that has precisely one α bubble appears precisely twice on the right, and these two terms have opposite signs. Each term having precisely two α bubbles appears four times, twice with a plus sign and twice with a minus sign. Each term having precisely $n > 0$ α bubbles appears 2^n times, half with plus and half with minus signs. However, each term with no α bubbles appears only once, in the first term. This confirms (3.3′) and gives an independent confirmation of (3.3).

C. Leading Normal Threshold Formula

Using the identity just obtained, one easily derives the normal threshold formula obtained earlier[15] without using infinite series.

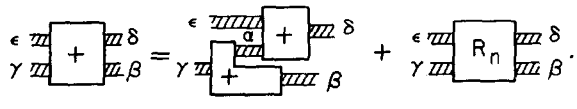
In the expansion (3.2′) of



(3.5)

some terms will have a cut C such that all the flow lines through this cut begin in δ and end in γ and such that the removal of the lines of this cut separates S into two disjoint parts, one containing ϵ and δ, the other containing γ and β. Let the sum of terms having no such (empty or nonempty) cut C be called R_n.

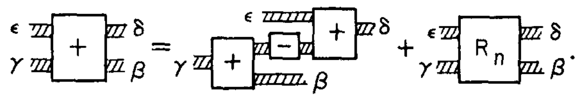
A term having such a cut C may have several. All these must be equivalent, since each defines precisely the set of all flow lines that begin at δ and end at γ. Let the leftmost of these cuts be labeled α. Then the separation of the terms of the expansion of (3.5) into terms having, or not having, a cut C gives



(3.6)

Each term in the expansion of the left side either has no cut C and hence belongs to R_n, or has a left-most cut α and appears precisely once in the first term on the right of (3.6).

Insertion of (3.3′) into (3.6) gives



(3.6′)

This formula is essentially the same as that derived (laboriously) in Ref. 15, by means of finite methods. There the plus boxes were the actual S matrices (rather than their infinite-series expansion) and R_n was a certain *finite* sum of bubble diagram functions F^B having just the property that defines R_n; i.e., no B corresponding to a term of the sum R_n has a D_B having point vertices for all minus bubbles that supports a cut C of the kind described.

The important property of R_n is that it contains no B having a D_B that contracts to any positive-α normal threshold diagram D_n^+ of the form indicated in Fig. 3. (D_B is defined in Sec. 2F.)

The first term on the right of (3.6′) vanishes below the leading normal threshold associated with diagrams of the form D_n^+. The second term on the right has, by construction, no positive-α singularity corresponding to any diagram that contracts to any diagram of the form D_n^+. If mixed-α singularities (i.e., singularities associated with solutions of Landau equations that involve α_i of both signs) can be ignored (see Sec. 6) and if the only diagrams D^+ giving surfaces $L(D^+)$ through a point $\bar{P}$ are those that contract to a diagram of the form D_n^+, then the only singularities of R_n at $\bar{P}$ are those associated with diagrams that contract to D_n^-. The function R_n must then, by virtue of the fundamental theorem, continue into itself via a *minus* $i\epsilon$ rule around the threshold. It is consequently the continuation of S from the region just below threshold to the region underneath the cut starting at threshold. The first term on the

$D_n^+ \equiv$

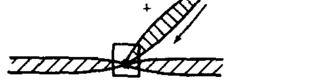


FIG. 3. The positive-α normal threshold diagram D_n^+. The + sign indicates that the σ_i of all lines of the set of lines between the two vertices are plus one. The arrow indicates that all lines have the direction indicated. D_n^- is defined by the same diagram with minus in place of plus. The boxes around the vertices indicate that it is not necessary that the vertices within them be *single* points; a point within a box can represent several disconnected point vertices. The "leading" normal threshold associated with a diagram of the form D_n^+ is the one corresponding to the lowest sum of rest masses of the exchanged particles.

right of (3.6′) is thus just the discontinuity around the normal threshold.

D. A Generalized Identity

The function S^α is the set of terms of (3.2′) such that no cut lies left of the cut α.

Let the mass M_α of a set of lines α be the sum of rest masses of the lines α. Let α' denote a cut that lies left of α and also satisfies $M_{\alpha'} \geq M_\alpha$. Let $S^{\alpha'}$ be the subset of (3.2′) that has no α'.

Let P_α be the projection function that is zero or one according to whether the set of lines β on which it acts satisfies $M_\beta < M_\alpha$ or $M_\beta \geq M_\alpha$. Let $S_\alpha = P_\alpha S P_\alpha$. That is, S_α is S if both incoming and outgoing lines have mass $\geq M_\alpha$, but it is zero otherwise. Then near the α threshold one obtains the following generalization of (3.3): For any S with a (sub) set of incoming lines α,

$$S^{\alpha'} S_\alpha = S, \tag{3.7}$$

where, in complete analogy to (3.3), S_α acts between the sets α and α'. [The proof is essentially the same as for (3.3); the nearness to threshold ensures that the leftmost cut α' is unique.[13]]

From (3.7) one obtains, as the generalization of (3.3′),

$$S^{\alpha'} = S S_\alpha^{-1}, \tag{3.8}$$

where S_α^{-1} is the inverse of S_α:

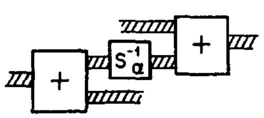
$$S_\alpha S_\alpha^{-1} = P_\alpha. \tag{3.9}$$

(This definition of S_α^{-1} is slightly more general than the one given in the Introduction; it covers also the special case when two different sets of communicating particles have the same sum of rest masses.)

E. General Normal Threshold Formula

Consider the expansion (3.2′) of S of (3.5). Let α be a cut of the type described below (3.5) with the additional condition that M_α be equal to or greater than some fixed sum of rest masses.

The arguments leading to (3.6) are now repeated, but now with R_α containing the terms having no cut α. One then obtains, for the discontinuity around the α normal threshold, the formula

$T_\alpha =$  . (3.10)

This result is the same as that obtained by finite methods in Ref. 15, except that there M_α was required to be less than the lowest communicating 4-particle threshold. This limitation is here removed.

F. General Physical-Region Discontinuity Formula

Essentially the same argument gives the general discontinuity formula described in the introduction.

Consider some basic positive-α diagram D_β^+. Let α label the sets of lines connecting the various pairs of vertices of D_β^+. Let the *mass* of a set of lines be the sum of the rest masses of these lines, and let M_α be the mass of α.

A bubble diagram B is said to *contain* D_β^+ if and only if $D(B)$ contains D_β^+. [$D(B)$ is the diagram obtained by shrinking the bubbles of B to points.] A D contains D_β^+ if and only if it has a set of mutually disjoint cuts C_α, one corresponding to each of the sets α of D_β^+. The cut C_α corresponding to the set α must be a cut that consists of positively signed lines having mass M_α. Moreover, the cutting of all the lines of all these sets C_α must divide D into a set of N mutually disjoint parts, one corresponding to each of the N vertices of D_β^+. The part of D corresponding to the nth vertex of D_β^+ must contain the appropriate end points (leading or trailing) of the appropriate lines of the appropriate sets, as prescribed by $\epsilon_{\alpha n}$. ($\epsilon_{\alpha n}$ is the common sign of the ϵ_{in} of D_β^+ for i in α.) The connectedness of the part n of D is irrelevant; as in Fig. 3, it can be either connected or disconnected.

A B *excludes* D_β^+ if and only if no D_B contains D_β^+. (D_B is defined in Sec. 2F. Notice that "contain" and "exclude" are opposites provided that all the bubbles of B are minus bubbles.)

The important properties of these two classes are these: First, any sum T of F^{B}'s over B's that contain D_β^+ must vanish outside the physical region of D_β^+, and hence on the negative-f side of $L(D_\beta^+)$ (see Sec. 2F). Second, any sum R of F^{B}'s over B's that exclude D_β^+ must, by virtue of the fundamental theorem, have a minus-$i\epsilon$ continuation into itself past $\bar{P}$ of $L(D_\beta^+)$, provided $\bar{P}$ lies on no $L(D^+)$ except those such that D^+ contains D_β^+ and provided R has no mixed-α singularities at $\bar{P}$. It follows that a separation of S in two terms T and R that contain and exclude D_β^+, respectively, exhibits T as the discontinuity around any such $\bar{P}$ of $L(D_\beta^+)$.

Consider any B^- that contains D_β^+. Then $D(B^-)$, which is the diagram obtained by replacing each (minus) bubble of B^- by a point vertex, must have some set of cuts C_α corresponding to the sets α of D_β^+. A cut *strongly equivalent* to C_α is a cut that is equivalent to C_α and has the same mass. Any C_α may be replaced by any cut strongly equivalent to it without destroying its correspondence to α of D_β^+.

The result just stated is proved in Appendix C. It is assumed there, and in what follows, that the point $\bar{P}$

under consideration lies on $L(D_\beta^+)$ and lies on no $L(D^+)$ unless D^+ contains D_β^+.

The Landau equations for D_β^+ at $\bar{P}$ require the momentum-energy vectors of all the lines in a given set α of D_β^+ to have a common direction d_α. It also is assumed in Appendix C, and in what follows, that these directions d_α are all different, for the $\bar{P}$ under consideration.

Consider now the structure $\tilde{T}$ obtained by replacing each vertex of D_β^+ by the expansion (3.2′) of the S corresponding to that vertex. Delete from the expansion of each S all terms corresponding to diagrams having some cut that is strongly equivalent to, and stands left of, the cut corresponding to any set α of incoming lines of that S.

This structure T contains every term B^- in the expansion (3.2′) of S that contains D_β^+: For any such term there must be a set of cuts C_α that correspond to the various α of D_β^+. Consider the leftmost cuts C_α' strongly equivalent to these. These C_α' separate B^- into parts that correspond to the vertices of D_β^+. The part corresponding to the nth vertex will be some term in the expansion (3.2′) of the S corresponding to that vertex, and it will be one of the terms that is retained in the construction of T.

Thus any term in the expansion (3.2′) of S that contains D_β^+ will be some term in the structure T, and any term in the structure T evidently contains D_β^+ and is a term of (3.2′).

It remains to show that each term of (3.2′) that contains D_β^+ is contained precisely once in T. If this is true, then the remainder R will exclude D_β^+, and the desired separation of S will be achieved.

Each term in (3.2′) that contains D_β^+ will be contained precisely once in T, provided any B^- that contains a set of leftmost cuts C_α' corresponding to the α of D_β^+ contains precisely one such set: For every such set of cuts C_α' in B^- this term is contained precisely once in T.[13] Thus, we must show that each B^- that has a set of leftmost C_α' corresponding to the α of D_β^+ has precisely one such set.

Suppose for some B^- there are two sets of leftmost cuts C_α' that correspond to the α of D_β^+. The function F^{B^-} will vanish in an infinitesimal neighborhood of $\bar{P}$ unless the constraints of B^- allow the p_i' corresponding to the lines of each of these sets of C_α''s to assume the (unique) values $p_i(\bar{P})$ that solve the Landau equations of D_β^+ at $\bar{P}$.

Consider a reduced diagram $\bar{D}^+$ that contains only those lines of $D(B^-)$ that lie on one or the other of the two sets C_α'. Since the Landau equations at $\bar{P}$ must be satisfied for the lines coming from each of the sets C_α' separately, they must be satisfied for the whole diagram $\bar{D}^+$: $\bar{P}$ must lie on $L(\bar{D}^+)$ if B^- is to contribute near $\bar{P}$.

The conditions on D_β^+ for there to be a $\bar{D}^+$ that contains D_β^+ in two essentially different ways, as above, are very stringent. For example, the leading vertex of D_β^+ that expands into more than a single vertex of $\bar{D}^+$ must have a set of outgoing lines that represent particles that can decay into the particles represented by another set of outgoing lines of that vertex (see Fig. 7, Appendix B). This places strong conditions on the momenta p_j associated with these lines, and hence stringent conditions on $\bar{P}$ We call "redundancy conditions" these conditions on $\bar{P}$ that must be satisfied if D_β^+ is to be contained in several essential different ways in some D^+.

Our conclusion then is this: Suppose the following conditions are satisfied:

(i) $\bar{P}$ lies on $L(D_\beta^+)$ and on no $L(D^+)$ unless D^+ contains D_β^+.

(ii) The directions d_α of the p_j of the various sets of lines α of D_β, as defined by the Landau equations of D_β^+ at $\bar{P}$, are all different.

(iii) The redundancy conditions on D_α^+ are not satisfied at $\bar{P}$.

(iv) The remainder $R = S - T$ has no mixed-α singularities at $\bar{P}$ (see Sec. 6).

Then the discontinuity of S around $L(D_\beta^+)$ at $\bar{P}$ is given by the rules described at the beginning of the paper, where the diagram D is just D_β^+. Notice that condition (i) ensures that $\bar{P}$ lies on the codimension-1 surface $L_0(D_\beta^+)$ (see Sec. 2D).

The disconnected parts of S have, of course, conservation-law δ-function factors. The discontinuities associated with these parts are calculated in the natural way, by taking the discontinuity corresponding to a path that encircles the singularity surface $L_0(D_\beta^+)$, while remaining in the manifold defined by the appropriate conservation-law δ functions.

We believe the discontinuity formula for S itself, rather than its connected part, will be the more useful in practice, because in any applications based on unitarity (or on other physical conditions) it is the full S, rather than its connected part, that is relevant. One lesson we have learned from our work is that general results for multiparticle processes are hard to derive from unitarity if one separates out the disconnected parts before the final stage.

The derivation given in this section is based on the infinite-series expansion for S. However, all infinite series are eliminated from the final result. This suggests that the results should be derivable directly from the equation $SS^{-1} = I$ that generated the infinite series. This has been done in many special cases.[3,4,15] In Sec. 5 we derive the result for all "leading" singularities, without using infinite series.

The expansion of (3.2′) for S has an infinite number of terms, one for each diagram D^+. An interesting finite expression is obtained by grouping together the contributions corresponding to different structures s. A structure s corresponds to the *class* of basic diagrams D_β^+ that differ only by the masses associated with the various sets of lines α. That is, the masses of the particles that pass between the two vertices specified by α are not restricted; they are allowed to be anything.

This grouping of terms gives

$$S = \sum_s S_s. \tag{3.11}$$

The expression for S_s is obtained by replacing each vertex of the structure diagram by a minus bubble and each set of lines α by the entire S matrix acting between the two corresponding minus bubbles.

This expansion (3.11) for S is something like a Feynman expansion, but with the following important differences:

(a) It is strictly mass shell and physical region.

(b) Only a finite number of terms contribute at any finite energy.

(c) Each propagator is the entire physical S matrix.

(d) Each vertex is a minus bubble.

This system of exact integral equations appears to be interesting, but their exploitation is not our present aim.

4. PROPERTIES OF S_α^{-1}

The function S_α^{-1} is the inverse of $S_\alpha = P_\alpha S P_\alpha$, where P_α is the projection on configurations of communicating particles having a sum of rest masses greater than or equal to the mass M_α associated with the lines α of some Landau diagram. The equation for S_α^{-1} has a formally Fredholm structure. In the case that M_α lies below the lowest 4-particle threshold (for communicating particles) the equation for S_α^{-1} has been converted to strict Fredholm form.[3] This has not yet been done in the general case.

The function S_α^{-1} can be expressed in terms of S and S^{-1} and their continuations. To obtain these expressions, first introduce the definitions

$$R_\alpha^- = S_\alpha^{-1} - I_\alpha \tag{4.1}$$

and

$$R_\alpha = S_\alpha - I_\alpha. \tag{4.2}$$

These satisfy

$$R_\alpha^- + R_\alpha + R_\alpha^- R_\alpha = 0. \tag{4.3}$$

Both R_α and R_α^- are restricted to the space allowed by $P_\alpha \equiv I_\alpha$. The function R_α^- is the restriction to this space of the $\bar{R}_\alpha^-$ defined by

$$\bar{R}_\alpha^- + R + \bar{R}_\alpha^- P_\alpha R = 0. \tag{4.4}$$

[The projection of (4.4) on α is just (4.3).]

Define the quantity $\bar{R}_\alpha^+$ by

$$\bar{R}_\alpha^+ + R^- + R^- Q_\alpha \bar{R}_\alpha^+ = 0, \tag{4.5}$$

where $Q_\alpha + P_\alpha = I$ and $R^{-1} = S^{-1} - I$. The restriction of $\bar{R}_\alpha^+$ to the space allowed by Q_α is called R_α^+:

$$R_\alpha^+ \equiv Q_\alpha \bar{R}_\alpha^+ Q_\alpha. \tag{4.6}$$

It satisfies

$$R_\alpha^+ + Q_\alpha R^- Q_\alpha + Q_\alpha R^- R_\alpha^+ = 0. \tag{4.5′}$$

Below the α threshold the Q_α are irrelevant and R_α^+ can be identified with $Q_\alpha R Q_\alpha$. We showed in Ref. 3 that R_α^+ evaluated just above the α threshold coincides with the continuation of $Q_\alpha R Q_\alpha$ from the physical region lying just below the α threshold, the continuation being via the minus $i\epsilon$ rule. We also established a number of interesting relationships between $\bar{R}_\alpha^+$ and $\bar{R}_\alpha^-$, such as

$$\bar{R}_\alpha^+ = -\bar{R}_\alpha^- \tag{4.7}$$

and

$$S_\alpha^{-1} = P_\alpha S^{-1} P_\alpha - P_\alpha S^{-1} Q_\alpha S^{-1} P_\alpha - P_\alpha S^{-1} R_\alpha^+ S^{-1} P_\alpha. \tag{4.8}$$

This latter equation [Eq. (C12) of Ref. 3] allows S_α^{-1} to be expressed in terms of S^{-1} and the continuation of $Q_\alpha R Q_\alpha$ to underneath the α cut.

In Ref 3. the results just described were derived only for energies lying below the lowest 4-particle threshold of the channel in question. However, they hold also in general, at least in our iterative framework. To see this, one can first consider R_α^+ to be defined to be the sum of all terms of the expansion (3.2) that contain no direct-channel α cut. That is, R_α^+ is the sum of all terms of expansion (3.2) that exclude the direct-channel normal-threshold structure diagram D_α^+, where α specifies a certain sum of rest masses. In this case, our general expansion of S according to D_α^+ gives [see (3.10)]

$$S = S S_\alpha^{-1} S + R_\alpha^+ + Q_\alpha. \tag{4.9}$$

Multiplication on the left by S^{-1} gives

$$I = S_\alpha^{-1} S + S^{-1} R_\alpha^+ + S^{-1} Q_\alpha. \tag{4.10}$$

Recalling that

$$S_\alpha^{-1} = P_\alpha S_\alpha^{-1} P_\alpha \tag{4.11}$$

and noting that

$$R_\alpha^+ = Q_\alpha R_\alpha^+ Q_\alpha, \tag{4.12}$$

we obtain, by left multiplication of (4.11) by Q_α, the original definition of (4.5′) of R_α^+.

Left and right multiplication of (4.11) by P_α gives the defining equation for S_α^{-1}. Left multiplication of (4.11) by P_α and right multiplication by $S^{-1}P_\alpha$ gives (4.9). Equation (4.8) can be derived in the same way as in Ref. 3. [See (5.18) and Appendix C of Ref. 3.]

The above argument shows that the quantity R_α^+ defined by (4.5′) is equal to the sum of all terms (but Q_α) of the expansion (3.2′) of S that exclude D_α^+ and that it is, accordingly, the continuation of $Q_\alpha R Q_\alpha$ to underneath the cut starting at the α threshold.

It is surprising that the R_α^+ defined by (4.5′) is the continuation of $Q_\alpha R Q_\alpha$ to underneath the α cut. For many terms of the iterative solution to (4.5′) do contain D_α^+. However, a detailed examination shows that each such term of $Q_\alpha R^- Q_\alpha + Q_\alpha R^- R_\alpha^+$ is cancelled by an identical term with opposite sign.

This cancellation allows the results of Ref. 15 to be extended without essential change to the regions above the lowest 4-particle channel threshold, except that the justification of some steps by Fredholm theory is no longer supplied. We expect it could be supplied by the same sort of arguments that were given in Ref. 3 for the 2- and 3-particle intermediate states.

5. INDUCTIVE SOLUTION

This section contains an alternative derivation of the discontinuity around "leading" singularities. This derivation does not rely on the infinite-series expansion for S, but is based instead on the results of Ref. 15. The point $\bar{P}$ is as above.

The principal results of Ref. 15 are these: (i) Over any bounded domain, S can be converted, by a finite number of applications of $SS^{-1} = I$, to the form $T[D_n^+] + R[D_n^+]$, where $T[D_n^+]$ is the first term on the right of (3.6′) and $R[D_n^+]$ is a certain finite sum of bubble diagram functions F^B, each corresponding to a B that excludes the normal-threshold diagram D_n^+ of Fig. 3. (ii) The quantity Σ on the right of (3.3′) can be similarly converted to a finite sum Σ′ of F^B's, each corresponding to a B that has no cut $\alpha' \not\equiv \alpha$ that is equivalent to α.

The discontinuity around any leading singularity can be derived by repeated application of these two results. To do this, first select a leading vertex V of D_β^+ (i.e., all incoming lines of V are incoming lines of D_β^+). Let $D_n^+(V)$ be the D_n^+ obtained by contracting all internal lines of D_β^+ but those that are outgoing lines of V. Then any B that excludes $D_n^+(V)$ will exclude also D_β^+. Thus, the second term on the right of

$$S = T[D_n^+(V)] + R[D_n^+(V)] \tag{5.1}$$

consists of terms that exclude D_β^+.

The first term on the right of (5.1) has the form of the first term on the right of (3.6′). The part Σ of this term that is the right-hand side of (3.3′) can be converted by means of a property (ii) to a sum Σ′ of F^B's, each corresponding to a B that has no $\alpha' \not\equiv \alpha$ equivalent to α. This gives the alternative form

$$S = T' + R[D_n^+(V)]. \tag{5.2}$$

Let D' be any $D_{T'}$ that contains D_β^+, with $\bar{P}$ on $L(D')$.

Let C_V be the sum of the leftmost cuts C_α' of D' that correspond to the sets α that begin at V of D_β^+. Property (ii), together with the requirement that the sets α be leading sets, entails that any C_V in D' consist precisely of the set of lines Γ of T' that run out of the right-hand plus box and into Σ′. That is, property (ii) requires any C_V to lie to the right of Σ′, and the condition that the various sets α be leading sets rules out the possibility that C_V lies inside the plus box (i.e., the kinematic constraints at $\bar{P}$ do not allow the particles in different leading sets α to come together again after leaving V; see Appendix C.)

Thus any C_V in D' must consist of precisely the lines Γ. Let $\{p_i(\bar{P})\}$ be the $\{p_i\}$ of the unique[8] solution of the Landau equations of D_β^+ at $\bar{P}$. Then, the only part of the integral over the lines of Γ that contributes to the singularity at $\bar{P}$ associated with D_β^+ comes from the region near the points where the p_i of Γ assume the values $p_i(\bar{P})$: The other parts of the integral do not allow the Landau equations of D_β^+ to be satisfied at $\bar{P}$.

Let the lines of Γ be divided into sets Γ_α, one for each of the sets C_α' of C_V, such that, near the point $p_i = p_i(\bar{P})$, the set Γ_α contains the lines contained in C_α'. Then $T \equiv T[D_n^+]$ can be separated into three terms:

$$T = T^a + T^b + T^c. \tag{5.3}$$

The term T^a consists of those terms of T such that some minus bubble of T connects lines from different sets Γ_α. The remaining terms have no minus bubble connecting these sets, and the separation into sets Γ_α of the set Γ induces a corresponding separation into sets Γ_α' of the set of lines Γ′ that emerge from the minus box and enter the left-hand plus box. Let this plus box be written as $T[\hat{D}_\beta^+] + R[\hat{D}_\beta^+]$, where $\hat{D}_\beta^+$ is the diagram obtained by removing V from D_β^+. The two corresponding terms of T are called T^b and T^c, respectively. Then T^b is the desired $T[D_\beta^+]$.

We proceed by induction on the number of vertices of D_β^+. Thus, $T[\hat{D}_\beta^+]$ is assumed to have the form described in the Introduction, and $R[\hat{D}_\beta^+]$ is assumed to have no singularities corresponding to diagrams D^+ that contain $\hat{D}_\beta^+$. The analogous property must then be derived for D_β^+.

In this section we shall accept an *extended independence property* that asserts that in any equation $G = 0$, derived from unitarity (or $SS^{-1} = I$), the *net* singularity corresponding to any basic diagram D_β^+ is zero. That is, the various singularities corresponding to any one D_β^+ cancel among themselves. This is what one would naturally expect; the singularities corresponding to different basic diagrams should generally have different analytic characters and would not be expected to cancel against each other, even if they could coincide.

This assumption simplifies the present proof, but is not actually necessary, as is discussed in Sec. 6.

The work of Ref. 15 that gives property (ii) can be extended to show that $T^b \equiv T[D_\beta^+]$ can be converted to a form $T^{b'}$ that has the same property as T': Any cut C_V must lie in Γ.

Consider, then, the identity

$$T' - T^{b'} = T^a + T^b. \tag{5.4}$$

Multiplication on the right by the inverse of the right-hand plus box gives

$$F' = F. \tag{5.5}$$

The equality of the two sides of this equation is a consequence of unitarity (or $SS^{-1} = I$).

The function F' has the property of Σ': Any cut C_V must lie in Γ. The function F has the opposite property: No cut C_V can lie in Γ. We conclude that F' has no net singularity corresponding to C_V in Γ. But then $T' - T^{b'} = T - T^b$ can have no singularity corresponding to D_β^+. This property holds true also for $S - T$ [see (5.1)]. Thus, it must hold for their sum

$$S - T^b \equiv R^b = R(D_\beta^+).$$

This completes the induction proof.

6. DISCUSSION OF ASSUMPTIONS

The assumptions used in our derivation of the discontinuity formula are these: First, there are some general assumptions embodied in the cluster decomposition principle, the positive-α rule (which says that the singularities of S_c and S_c^- are confined to positive-α Landau surfaces) and the $i\epsilon$ rule. These general assumptions are consequences of the macrocausality requirement, as was discussed in Sec. 2. Second, there are the independence property and the technical assumption, which are needed for the fundamental theorem. The independence property is the full content in this work of maximal analyticity. We plan to discuss the technical assumption elsewhere.

A third set of assumptions are special conditions on the point $\bar{P}$. In the first place, $\bar{P}$ is required to lie on $L(D_\beta^+)$, but on no $L(D^+)$ unless D^+ contains D_β^+. Second, the directions d_α of the momentum–energy vectors corresponding to different sets α of internal lines of D_β^+ at $\bar{P}$ are required to be all different. And third, $\bar{P}$ is required to be such that at $\bar{P}$ no $\bar{D}^+$ contains D_β^+ in two essentially different ways. These conditions on $\bar{P}$ are to ensure that positive-α singularities associated with diagrams other than D_β^+ do not contribute at $\bar{P}$ and that those associated with D_β^+ contribute precisely once.

The discontinuities at points $\bar{P}$ where these conditions on $\bar{P}$ fail can be calculated by making use of the independence property. Suppose, for example, that $\bar{P}$ lies on $L(D^+)$ for some D^+ that does not contain D_β^+. The diagram D^+ can be assumed to be basic. Then, $\bar{P}$ must lie also on $L(\bar{D}_\beta^+)$, where the basic diagram $\bar{D}_\beta^+$ is a contraction of D^+. (One contracts out the lines of D^+ that correspond to $\alpha_i = 0$.) The independence property then ensures that the singularities at $\bar{P}$ associated with the D_β^+ and $\bar{D}_\beta^+$ are independent (i.e., additive) unless there is some $\hat{D}_\beta^+$ that contains both D_β^+ and $\bar{D}_\beta^+$, with $\bar{P}$ on $L(\hat{D}_\beta^+)$. Since the Landau equations for $L(D_\beta^+)$ and $L(\bar{D}_\beta^+)$ are both satisfied at $\bar{P}$, this point must lie also on $L(\hat{D}_\beta^+)$. If $\bar{P}$ lies on $L(D^+)$ for no other basic diagram D^+, then one can classify all basic diagrams $\tilde{D}_B^+$ such that $\bar{P}$ lies on $L(\tilde{D}_\beta^+)$ according to whether $\tilde{D}_\beta^+$ contains just D_β^+, just $\bar{D}_\beta^+$, or both (and hence also $\hat{D}_\beta^+$). The terms corresponding to the last case would be counted in both $T[D_\beta^+]$ and $T[\bar{D}_\beta^+]$. But they are also the terms included in $T[\hat{D}_\beta^+]$. Thus the discontinuity is $T[D_\beta^+] + T[\bar{D}_\beta^+] - T[\hat{D}_\beta^+]$.

In this case $\bar{P}$ lies on both $L(D_\beta^+)$ and $L(\bar{D}_\beta^+)$, and the above discontinuity is the difference between the function in the physical region of D_β^+ and its continuation around both $L(D_\beta^+)$ and $L(\bar{D}_\beta^+)$, where the continuation moves first through the plus $i\epsilon$ region associated with $\hat{D}_\beta^+$ and then through the corresponding minus $i\epsilon$ region.

More general cases are treated similarly, by using the general principle of inclusion and exclusion (see Appendix D of Ref. 15). The same sort of considerations apply also to cases where one or both of the other two conditions on $\bar{P}$ fail: Again one uses the independence property together with the principle of inclusion and exclusion to isolate the relevant set of terms.

The final assumption is that $R = S - T$ has no mixed-α singularities at $\bar{P}$.

We now argue that the sum on the right of $S = R + T$ should have no net mixed-α singularities. Since the quantity S on the left has singularities only on positive-α Landau surfaces, the only possible net mixed-α singularities on the right are those that happen to lie exactly on top of positive-α surfaces.

It is conceivable that these particular mixed-α would not cancel out, as all the others must, but it seems unlikely. In the first place, the physical arguments (macrocausality) that imply that the singularities of S are confined to positive-α surfaces correlate these singularities to positive-α diagrams. Thus, it would be unnatural for them to arise mathematically from other diagrams, which just happen to give the same Landau surfaces.[16] In the second place, the mixed-α singularities that happen to lie on positive-α surfaces are intimately related via hierarchy effects to the mixed-α singularities that do not lie on positive-α surfaces. It seems unlikely that the latter could all vanish identically without the former vanishing also.

On the basis of these arguments, we shall accept the proposition that in any equation of the form $S = X$ derived from $SS^{-1} = I$ the mixed-α singularities of the bubble diagram functions that comprise the right-hand side exactly cancel out (in the physical region). This will be our basic assumption about mixed-α nonsingularities. It may be possible to derive it by some inductive argument, but we do not attempt this here.

On the basis of this assumption we can confirm the absence of the mixed-α singularities in $R = S - T$ by confirming it rather for T.

The only lines of T that can be minus lines are the lines of the cuts C_α. By virtue of energy conservation, the momenta of all these lines are fixed at precisely the value defined by the Landau equations of D_β^+ at $\bar{P}$. (The Landau equations define the unique way of achieving the boundary point of the physical region of D_β^+. See Sec. 2F.)

Any mixed-α D_T such that $\bar{P}$ lies on $L(D_T)$ is a member of a continuum of such D_T. This continuum is generated by adding to the solution of the Landau equations corresponding to $\bar{P}$ on $L(D_T)$ a real multiple of the solution corresponding to $\bar{P}$ on $L(D_\beta^+)$. If the real multiple is sufficiently large and positive, then the mixed-α D_T is converted to a D_T^+, because all the lines corresponding to the C_α are eventually made positive. Thus, any point $\bar{P}$ on $L(D_\beta^+)$ that lies on the $L(D_T)$ of a mixed-α D_T must lie also on $L(D_T^+)$ for a continuum of $D_T^+ \neq D_\beta^+$, where D_T^+ contains D_β^+.

This shows that T can have no mixed-α singularities at simple points of $L(D_\beta^+)$, which are points that correspond to just one D_β.

At the nonsimple points $\bar{P}$ of $L(D_\beta^+)$ that lie on $L(D_T^+)$ for the continuum of $D_T^+ \neq D_\beta^+$ the meaning of our assumption about mixed-α singularities must be clarified, for we have to consider diagrams that can be continuously shifted from mixed-α to positive-α status. The correspondence between singularities and diagrams then becomes ambiguous. At these points of $L(D_\beta^+)$, where these flexible diagrams could give mixed-α singularities to T, we interpret our assumption that all mixed-α singularities of $T + R$ cancel to mean that the only net mixed-α singularities of R are those associated with the same flexible diagrams that give the possible mixed-α singularities of T.

With this interpretation we can show that the mixed-α singularities of R that might occur at these special points would not, in any case, upset our proof. The point is that contributions to R associated with these flexible diagrams must have minus-$i\epsilon$ continuations past the surface $L(D_\beta^+)$. This is because the construction of R ensures that these contributions can occur only if the minus lines of the (flexible) diagram come from inside minus bubbles. But then the proof of the fundamental theorem shows that the continuation past the surface $L(D_\beta^+)$ will follow the minus $i\epsilon$ rule, due to the presence of these necessarily minus lines. But then the proof of the discontinuity formula would go through even at these very special points at which the flexible diagrams give singularities.

In Sec. 5 an extra assumption (extended independence) was used to simplify the argument. To avoid the assumption, one need modify the proof only slightly. First, the function $R[D_\beta^+]$ is considered to be decomposed (using the ordinary independence property) according to basic positive-α diagrams $\tilde{D}_\beta^+$ (this decomposition is unambiguous). Then the assumption of the induction argument is that all terms corresponding to diagrams $\tilde{D}_\beta^+$ that contain $\hat{D}_\beta^+$ vanish from $R[\hat{D}_\beta^+]$. The analogous property must then be proved for $R[D_\beta^+]$.

The proof proceeds as before, but one now decomposes also the two sides of $F' = F$ according to basic positive-α diagrams. Only the terms that can contribute to the final D_β^+ need be considered (see below). But the singularity surfaces bounding the supports of these terms are not the same on the two sides of $F' = F$. Thus these terms must vanish. But then $T - T^b$ has no terms corresponding to D_β^+. Nor does $S - T$. Thus neither does their sum $S - T^b = R^b$.

[The condition that $\bar{P}$ lies on no $L(D^+)$ for any D^+ not containing D_β^+ implies that one need consider only terms that contribute to the final D_β^+, for, if any other diagrams could exactly compensate for the missing term in F', then this term also would give an unallowed D^+.]

The argument given above, in effect, justifies the extended independence property, in the context in which it was used.

The present work generalized the results obtained earlier by ourselves[3,13] and by the Cambridge group.[4] We now contrast our methods and results with theirs.

Regarding final results, our discontinuity formula covers all physical-region singularities, whereas their general result covers only the case of simple diagrams. (In simple diagrams, each set α consists of just one line.) They have obtained results also for certain special nonsimple diagrams, and are working toward the general result.

Some theorems in the early part of their work are somewhat similar to our fundamental theorem. However, the treatment of technical details is considerably different in the two works.

Our basic procedure is quite different from that of the Cambridge group. Their approach is in a way more general, since they first derive general formulas for discontinuities of integrals in terms of the discontinuities of their integrands. Then they use these results to show that for singularities associated with simple diagrams the Cutkosky discontinuity formula is consistent with unitarity. Finally, they show, by means of an inductive procedure, that no other solution is possible: If the Cutkosky formula is valid for all simple diagrams up to a certain order of complexity, then it must hold also for diagrams of the next order of complexity, provided singularities corresponding to nonsimple diagrams can be ignored.

Their procedure, then, is first to make a detailed general analysis of discontinuity formulas and then to introduce these results into unitarity, which is used in only a limited way.

Our procedure is the reverse. The manipulations involved in our approach are purely topological and involve multiple applications of unitarity (or, more accurately, the cluster properties of S and S^{-1}). These topological manipulations give equations

$$S = R[D_\beta^+] + T[D_\beta^+],$$

where the topological characteristics of the terms on the right guarantee that $R[D_\beta^+]$ is the continuation of S around $L(D_\beta^+)$ via the minus $i\epsilon$ rule and hence that $T[D_\beta^+]$ is the discontinuity. Analyticity is used only at the last stage, and thus complications connected with distortions of contours are avoided.

This procedure is more special, in that it refers to the particular problem at hand, but it yields a variety of strict identities[15] that can be used in other contexts. These identities are consequences of the cluster properties alone and are purely topological in nature; analyticity is not involved.

The assumptions needed in the two approaches are, with one important exception, essentially the same. In particular, the independence and boundedness properties are needed in both methods,[11] and the considerations involving the special conditions on $\bar{P}$ are essentially the same.

The one important difference is that the Cambridge group does not assume that the singularities of S and S^{-1} are confined to positive-α surfaces: Their aim is to derive this result. On the other hand, they do assume the $i\epsilon$ rules, for positive-α points, and also certain similar rules at mixed-α points. Our viewpoint is that these strong $i\epsilon$ requirements should not be imposed ad hoc, but must be justified. We justify the $i\epsilon$ rules on the basis of macrocausality and get the positive-α rule at the same time. Alternatively, one might justify the $i\epsilon$ rules on the basis of self-consistency, but one should then also prove uniqueness.

APPENDIX A: THE INDEPENDENCE PROPERTY AND THE FUNDAMENTAL THEOREM

The fundamental theorem quoted in Sec. 2F has slightly weaker assumptions and slightly stronger conclusions than the theorems proved in Ref. 12. In this appendix we discuss these assumptions and show how the proof of Ref. 12 can be extended to give the theorem quoted in Sec. 2F.

One technical detail should be mentioned first. What is proved in Ref. 7 is that S_c (or S_c^-) *considered as a distribution* can be represented as the limit of the analytic function. That is, this representation is shown to be valid when one is calculating the average of S_c (or S_c^-) over a Schwartz test function. But what is needed to prove the structure theorems is something slightly different. One needs to evaluate products of different S_c's and S_c^-'s with one another.

In the proof of the structure theorems, each of these functions S_c and S_c^- was considered to be a limit of the analytic functions described above, and their products were *defined*, for certain fixed real values of the external (unintegrated) momenta, by performing the appropriate integration over internal momenta along a multidimensional contour that remains in the region of analyticity of all the relevant functions S_c and S_c^-. This contour is such that it can be shifted (staying in the analyticity domain) to a position arbitrarily close to the real physical region. By virtue of the (multidimensional) Cauchy theorem, such a shift does not alter the value of the integral.

For any fixed real value of the (external) variables K of $F^B(K)$, the integrations occurring in the definition of F^B were assumed to be given by the above rule, provided the relevant domains of analyticity of the various functions S_c and S_c^- overlap in such a way that the required contour through the intersection of the analyticity domains, infinitesimally removed from the real physical region, exists. The function $F^B(K)$ was shown to be analytic at such values of K, and the rule for continuing the thus defined function $F^B(K)$ around any singularity at real K was derived.

This rule defining the integrals in $F^B(K)$ was used to evaluate the terms of SS^{-1}, $SS^{-1}S$, etc. If one considers the S matrix to be defined basically in terms of limits of analytic functions, then this definition of the meaning of the SS^{-1}, $SS^{-1}S$, etc., is the reasonable one. However, if one starts with S and S^{-1} considered to be operators in a Hilbert space, then this rule for defining their products must be justified. The required justification is given at the end of this appendix.

It was asserted in Sec. 2F that the independence properties of S_c and S_c^- lead to analogous properties of the bubble diagram functions F^B. The point is that the proofs of the structure theorems show that the singularities of F^B corresponding to any basic diagram D_β^+ arise from singularities of the bubbles b of B that are associated with the parts $D_{\beta b}^+$ of D_β^+ that lie in b, when D_β^+ is regarded as a D_B^+. These parts $D_{\beta b}^+$ must be basic diagrams if D_β^+ is. Now, by virtue of the independence property of S_c, the singularities of b associated with different basic diagrams $D_{\beta b}^+$ are independent. If any one specific $D_{\beta b}^+$ is inserted into each b of B, then one specific D_B^+ is formed. This contracts to some unique basic $D_{B\beta}^+$. It thus follows that the singularities of F^B corresponding to different basic diagrams D_β^+ must arise from independent singularities of at least one b of B and must therefore be independent.

The independence property can, alternatively, be derived from macrocausality at almost all points of the surface of singularities L^+. However, there is then the problem of extending the property to those rare (Type II) points at which this argument breaks down.

The independence property is not included among the assumptions mentioned by the Cambridge group.[4] This omission is connected with their somewhat relaxed way of specifying the precise conditions under which their basic theorems are valid. If one wishes to formulate their theorems precisely, in forms strong enough to do the job, then the independence property or something similar seems required. Following their philosophy, one might try to justify the independence property by an inductive procedure: The independence property for complex basic diagrams might be shown to follow from that of the simpler ones. However, an inductive procedure for proving independence would involve an artificial assumption that the singularities can be "ordered" and that one can proceed by stages, completely ignoring "higher-order" singularities at each stage. But since the discontinuity associated with any D_β^+ is, in effect, a sum of contributions corresponding to diagrams that are *more complex* than D_β^+, a justification of independence based on "hierarchy" is subject to question. In the procedure we adopt, no ordering is invoked, and there is never any "temporary neglecting" of certain singularities. Also, the full content of maximal analyticity is explicitly stated.

The second and third structure theorems (Theorems 3.2 and 3.3) given in Ref. 12 are specifically restricted to simple points of the Landau surfaces $L(D_B)$. That is, it is assumed that the point $\bar{P}$ corresponds to a *unique* basic diagram. This assumption is needed because the arguments cover only the case where there is only one constraint (3.7) (of Ref. 12). Now suppose there are many such constraints. The question is whether there is a set of variations δh_f of the Feynman loop parameters that keeps all the $\delta p_j^2 = 0$ and all the $\delta\sigma > 0$. (Such a set of variations would shift the contour simultaneously into the domain of analyticity of all the bubble functions, while maintaining all the mass-shell and conservation-law constraints.)

To solve this problem, consider the following lemma:

Lemma 1: For any set of real numbers η_{ba}, the system of equations

$$\sigma_b = \sum_a \eta_{ba}\delta_a, \quad \sigma_b > 0, \tag{A1}$$

has a solution δ_a if and only if the system of equations

$$\sum \alpha_b \eta_{ba} = 0, \quad \alpha_b > 0, \tag{A2}$$

has no solution.

Proof: Suppose (A1) has a solution. Insertion of this solution into (A2) gives a contradiction. Thus (A2) can have no solution. Conversely, suppose (A2) has no solution. Then the space X spanned by positive linear combinations of the vectors $\vec{\eta}_b$ with components η_{ba} is convex. Then there exists some vector δ that has positive inner product with every vector of X. This vector solves (A1), and the lemma is proved.

A slight generalization is the following.

Lemma 1': For any sets of real numbers η_{ba} and λ_{ca}, the system of equations

$$\sigma_b = \sum_a \eta_{ba}\delta_a, \quad \sigma_b > 0, \tag{A3a}$$

$$0 = \sum_a \lambda_{ca}\delta_a \tag{A3b}$$

has a solution δ_a if and only if the system of equations

$$\sum_b \alpha_b \eta_{ba} + \sum_c \beta_c \lambda_{ca} = 0, \quad \alpha_b > 0, \tag{A4}$$

has no solution.

Proof: If (A3) has a solution, then (A4) can clearly have none. Conversely, if (A4) has no solution, then the space X of positive linear combinations of the $\bar{\eta}_b$ must be convex and must contain no vector in the linear space Y spanned by the $\bar{\lambda}_c$. Thus the orthogonal complement $X^\perp$ of X must have dimension at least that of Y. Moreover, $X^\perp$ cannot be contained in $Y^\perp$, for then X would contain vectors in Y. Thus, if Y is nonnull, there must be a nonzero vector that lies in $X^\perp$ but not in $Y^\perp$. The sum of a multiple of this vector with the vector in X satisfying (A3a) (found in Lemma 1) solves (A3), and the lemma is proved.

Lemma 1′ is precisely what is needed to extend the second and third structure theorems to nonsimple points.

It was mentioned at the beginning of this appendix that the integrations occurring in the definitions of the bubble-diagram functions $F^B(K)$ were defined to be along contours displaced infinitesimally from the physical region into the simultaneous analyticity domain of all the occurring functions S_c and S_c^-, provided the real K was such that such a contour exists. The proofs of the structure theorems show that such contours do exist for most real K, that the $F^B(K)$ is analytic at such points, and that $F^B(K)$ continues analytically around the remaining real points K via paths defined by certain rules.

It is reasonable to define the integrations in the way indicated. But if one begins with the idea that S and S^{-1} are operators in a Hilbert space,then this rule must be justified. The problem is that macrocausality gives the analytic representation for S_c and S_c^- considered as distributions, rather than as operators. It is not known whether this representation is valid for operators. However, we now show that the functions F^B considered as products of operators restricted to the space of Schwartz test functions can be defined by performing the integrations along the distorted contours described above.

Let H_p, H_q, and H_k be three Hilbert spaces of square-integrable functions of the multidimensional variables p, q, and k, respectively. Let $A: H_q \to H_p$ and $B: H_p \to H_k$ be two bounded operators. Let $\varphi(q)$, $\chi(p)$, and $\psi(k)$ be Schwartz test functions of compact support. Suppose for sufficiently small supports we know that

$$(\chi, A\varphi) = \lim_{\epsilon\to 0} \int dp\, dq \chi^*(p) A_\epsilon(p, q)\varphi(q)$$

and

$$(B\psi, \chi) = \lim_{\epsilon\to 0} \int dk\, dp \psi^*(k) B_\epsilon(k, p)\chi(p),$$

where $A_\epsilon(p, q) = A(p + i\epsilon_p, q + i\epsilon_q)$ and $\epsilon = (\epsilon_p, \epsilon_q)$ is a vector of fixed direction lying in a certain open convex cone (which can depend on the small supports of χ and φ), and similarly for $B_\epsilon(k, p)$. The function $A(p + i\epsilon_p, q + i\epsilon_q)$ is supposed to be analytic when p and q are in the supports of χ and φ, respectively, and ϵ is in the cone, and similarly for B.

[The functions A and B have certain energy–momentum δ functions as factors. The analyticity discussed above is for the factor that multiplies these δ functions, as described in detail in Refs. 7 and 8. We shall not explicitly write down the δ-function factors, but we will use the fact that the conservation laws entail that $A\varphi$ and $B\psi$ have compact supports if φ and ψ do. That is, the region of integration is a compact "cycle"—it has no boundaries. (See Ref. 12.)]

Consider fixed φ and ψ of small compact supports. Let χ_i be a finite set of Schwartz test functions such that $\sum \chi_i = I$ on the compact p space. Suppose the χ_i can be chosen so that the corresponding domains of analyticity of A and B overlap, in the sense that there is a contour $\mathcal{C}$ defined by $\epsilon(p)$ such that $A(p + i\epsilon(p), q)$ is in the domain of analyticity corresponding to χ_i and φ whenever p and q are in the supports of χ_i and φ, respectively, and similarly for B. We wish to show that

$$(B\psi, A\varphi) = \int dp\, dq\, dk \psi^*(k) B(k, p + i\epsilon(p)) \times A(p + i\epsilon(p), q)\varphi(q).$$

That is, we wish to show that the operator product $B^\dagger A$, acting between the Schwartz test functions ψ and φ, can be represented by an integral over the fixed contour $\mathcal{C}$. The contour $\mathcal{C}$ is displaced by a finite amount from the real axis, but the assumption is that it can be shifted to arbitrarily close to the real region, staying always in the cones of analyticity.

It is sufficient for our purposes to consider only a special class of functions χ_i. These will be functions formed by taking products of functions in the individual variables of p. Furthermore, the functions in each individual variable will be unity except at distance less than $\lambda > 0$ from the ends of its supports. The function in the support and at distance less than λ from the left end of the support will be given by the function

$$f_\lambda(x) = \exp(-x^{-\frac{1}{2}})\{\exp(-x^{-\frac{1}{2}}) + \exp[-(\lambda - x)^{-\frac{1}{2}}]\}^{-1}$$
$$= 1 - \exp[-(\lambda - x)^{-\frac{1}{2}}]\{\exp(-x^{-\frac{1}{2}}) + \exp[-(\lambda - x)^{-\frac{1}{2}}]\}^{-1}.$$

The right end will be given by the analogous function. The virtue of these functions is first that they are easily combined to give functions that add to unity and second that they are analytic except at zero and λ and approach their values at these points exponentially from any direction in the cut (along their support) plane.

Consider now the integral on the right of

$$(\chi_i, A\varphi) = \lim_{\epsilon \to 0} \int dp \chi_i(p) A(p + i\epsilon_p, q)\varphi(q).$$

Because of the analyticity properties of χ_i one can perform the limit $\epsilon \to 0$ by, instead of shifting the entire contour down to the real axis, merely extending the contour in the surfaces $\mathrm{Re}\, z = x = 0$ and $x = \lambda$ along the direction of ϵ into $\epsilon = 0$. This follows from a distortion of the multidimension contour.

Macrocausality guarantees that the functions corresponding to A and B grow no faster than some inverse power of $|\epsilon|$ as $\epsilon \to 0$ inside the cone of analyticity. The exponential falloff of χ_i at $\chi = 0$ then guarantees that the limit $\epsilon \to 0$ can actually be taken; one can extend the contour right down to the physical region. At $\chi = \lambda$ the contour also can be extended to $\epsilon = 0$, for the same reason, provided one combines the parts coming from the two sides of $x = \lambda$. [On one side one has the χ_i of the form of $f_\lambda(x)$, while on the other side one has $\chi_i = 1$. The difference falls off exponentially as $\epsilon \to 0$ on the surface $\mathrm{Re}\, z = \lambda$.]

One observes now that the contributions from these strips at $\mathrm{Re}\, z = x = 0$ and λ are exactly cancelled by the contributions from the neighboring χ_i. Thus, if one adds contributions from many different neighboring χ_i, the contour of integration is free to move about in the domain of analyticity except for the parts corresponding to the outer boundary strips associated with $\chi = 0$ and $\chi = \lambda$.

That is, in our original form, the ϵ were required to be constant over each domain χ_i (and generally a different constant for different χ_i), but we have now converted this to a single continuous contour $\mathcal{C}$ that varies smoothly over the union of the supports. In our case, where the union of the χ_i cover the entire compact cycle in p space, the contour $\mathcal{C}$ never descends to the real axis, but remains always in the domain of analyticity.

The above results apply equally if all the χ_i are replaced by $\chi_i e^{ipu}$. Thus, the Fourier transform

$$F(u) = (e^{ipu}, A\varphi)$$

is given by

$$F(u) = \int_{\mathcal{C}} dp\, dq e^{iup} A(p, q)\varphi(q).$$

Similarly, one has

$$G(-u) = (B\psi, e^{-ipu})$$
$$= \int_{\mathcal{C}'} dk\, dp \psi(k) B(k, p) e^{-ipu}.$$

Because A and B are bounded operators, these Fourier transforms are well defined, and one can write (up to factors of 2π)

$$(B\psi, A\varphi) = \int du G(-u) F(u).$$

The integrand in the expressions for F and G are analytic in p. That is, the integration region in p space can be divided into small regions in which local coordinates can be introduced. And, in each region, the variables corresponding to conserved energy–momentum are introduced as coordinates and then eliminated by the δ functions, leaving A and B analytic in the remaining (local) coordinates on the contour.

The function $G(-u)F(u)$ is infinitely differentiable (because of the compact supports in p space), and it falls off rapidly (faster than any polynomial) in all directions. The rapid falloff is due in part to the infinite differentiability of $\varphi(q)$ and $\psi(k)$ (which are brought in by the elimination of δ functions) and in part to the analyticity properties of A and B in the remaining (local) coordinates. The A and B are analytic in some common cone C in the local coordinates, and they grow no faster than some inverse power of $|\epsilon|$ on approach to the real physical region. Thus the argument of Chap. IVC.a of Ref. 7 shows that $F(u)$ and $G(u)$ fall off rapidly uniformly in the complement of the polar cone C^+. The boundedness of $F(u)$ and $G(-u)$ follows from the boundedness of A and B. Because of the different sign of the arguments of $F(u)$ and $G(-u)$, the intersection of the complements of the two effective polar cones C^+ is empty. Thus, $G(-u)F(u)$ falls off rapidly in all directions.

This rapid falloff implies that

$$(B\psi, A\varphi) = \lim_{\eta_i \to 0} \int du \exp(-\sum |u_i| \eta_i) G(-u) F(u),$$

where the right-hand side is analytic in η_i. Because of the compactness of the p-space region of integration, the order of the integrations can be inverted, for sufficiently large η_i. The u-integration then gives a sum of products of poles of the form $(p_i - p_i' \pm i\eta_i)^{-1}$.

Taking the limit $\eta_i \to 0$ then gives, after some algebra, the desired form. The main point is that, as one lets the $\eta_i \to 0$, certain poles cross the fixed contours $\mathcal{C}$ and/or $\mathcal{C}'$ and effectively reduce them to a single contour.

The methods used above can be extended to show the various other properties entailed by the assertion that the analytic representation extends in the natural way from distributions to products of bounded operators considered as distributions. In particular, the result described above carries over to products of many operators and to the case where the q and k must also be shifted. In this latter case, one wants to show that, if (for sufficiently small supports of φ and ψ) there is a cone C of analyticity in (q, k) such that for each point in this cone one can find a contour over the internal variables that remains always in the domain of analyticity [and hence that the product of the functions $B^+A = H$ is analytic in $(q, k) = z$], then $(\psi, H\varphi)$ can be represented as

$$\lim_{\eta\to 0} \int H(z + i\eta)\Omega(z)\, dz,$$

where $\Omega = \psi\varphi$ and $i\eta$ is in the cone C. The proof goes precisely as before, with H and Ω replacing $B^+\psi$ and $A\varphi$. The falloff of $\tilde{\Omega}(u)$ is now due to the infinite differentiability of $\Omega(z)$.

APPENDIX B: SUPPLEMENTARY NOTES

On Eq. (3.2′)

A proof of (3.2′) by induction is easy. Suppose each term of (3.2′) corresponding to a diagram B' having n nontrivial bubbles correctly gives the sum of the corresponding terms of (3.2). Let B be a diagram with $n + 1$ nontrivial bubbles. Select from among these a bubble b all incoming lines of which are also incoming lines of B. Let the removal of b from B give B'. Let α be the incoming lines of B' identified with the outgoing lines of b. Consider the various terms t' in (3.2) that sum to give the term of (3.2′) corresponding to B'. From each such t' we construct $2m + 1$ terms t of (3.2) that correspond to B, where m is the number of columns of t' lying to the right of the first nontrivial bubble b' of B' reached by the incoming lines α of B'. These $2m + 1$ terms are constructed by placing b either in one of m columns that lie to the right of b', or in a new column (containing only b) that stands just to the left of any of these columns, or in a new column (containing only b) that stands just to the right of the first column of t'. The $m + 1$ terms t involving a new column will all have one new minus sign, whereas the m terms not involving a new column will not have an extra minus sign. Aside from these signs all the terms are equal, and equal to the operator product of F^b with the $F^{B'}$ corresponding to the particular term t' of (3.2). Thus, the sum of the $2m + 1$ terms t is just minus one times the operator product of F^b with this $F^{B'}$. Summing over all terms t' of (3.2) corresponding to this B', one obtains all the terms t of (3.2) corresponding to B. Since the same operator $-F^b$ is applied to each term, one obtains by induction the term of (3.2′) corresponding to B.

An alternative proof of (3.2′) that makes use of (3.1) is as follows: Suppose (3.2′) has been shown to hold for terms corresponding to bubble diagrams having up to $n - 1$ nontrivial minus bubbles. Substitute (3.2′) into the second term of the right-hand side of the equation $R^+ = -R^- - R^+R^-$, and consider the contributions to the right-hand side corresponding to a bubble diagram B_n, where the subscript n indicates the number of nontrivial minus bubbles. The contributions to $-R^+R^-$ correspond to some B_n of the product form $B_j^+B_k^-$ (so that the outgoing lines of B_k^- are identical with the ingoing lines of B_j^+), where B_k^- consists of a column of k nontrivial minus bubbles and of unscattered lines and where $j + k = n$, with j and k no less than 1. Let i be the number of initial bubbles of B_n, where an initial bubble is a nontrivial bubble whose incoming lines are all external. All bubbles of B_k^- are initial bubbles.

Suppose at first that B_n does not consist of a single column of nontrivial minus bubbles and unscattered lines. Then all contributions to $-R^- - R^+R^-$ having n nontrivial minus bubbles come from $-R^+R^-$ only and must correspond to bubble diagrams $B_n = B_j^+B_k^-$, where $k = 1, 2, \cdots, i$ with $i < n$. There are $2^i - 1$ different ways of constructing B_n, all of which give contributions to $-R^+R^-$ having the value $\pm F^{B_n}$. These add up to

$$-F^{B_n}\left[(-1)^{n-1}\binom{i}{1} + (-1)^{n-2}\binom{i}{2} + \cdots + (-1)^{n-i}\binom{i}{i}\right] = (-1)^n F^{B_n}.$$

Suppose next that B_n does consist of a bubble diagram topologically equivalent to a column of n nontrivial minus bubbles and of unscattered lines so that $i = n$. Then the reasoning just given still applies, but now the last term in the above sum is missing because $k < n = i$; also $-R^-$ in $-R^- - R^+R^-$ now gives a contribution $-F^{B_n}$. Since

$$-F^{B_n} = -(-1)^{n-i}\binom{i}{i} F^{B_n}$$

when $i = n$, we get the same answer as before. Thus, expansion (3.2′) is verified.

On Topological Equivalence

As an example of the meaning of topological equivalence, consider the bubble diagram of Fig. 4. Certain contributions to F^B will correspond to the case where all the internal lines correspond to the same type of particle. If one simply integrated without

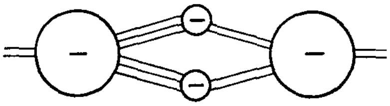


FIG. 4. A bubble diagram B.

respecting the requirement of topological independence, then one would get a contribution that would be too large by a factor of 2! 2! 2! 3! 3!. The two 3!'s come from the triples of lines on the left of the two intermediate bubbles. Two of the 2!'s come from the pairs of lines on the right of these bubbles. The other 2! comes from the topological equivalence of the upper and lower intermediate bubbles.

Leftmost Cuts

The definitions of equivalent cuts and of leftmost cuts are illustrated in Fig. 5.

Uniqueness of Leftmost Cut

The uniqueness, near the α threshold, of the leftmost cut equivalent to a cut C_α plays an important role in the arguments. At some finite distance above threshold, this uniqueness may fail, as Fig. 6 shows.

Leftmost Cuts C_α' in B^-

For any set of leftmost cuts C_α' in B^- corresponding to the sets α of D_β^+, there is a mapping Γ of $D(B^-)$ onto D_β^+. Each such Γ defines a set of parts $\Gamma^{-1}V$ of $D(B^-)$ (and hence of B^-) corresponding to the V of D_β^+. Each such Γ defines, in fact, precisely one way that B^- is realized as a term of T.

An example of a B^- that contains a D_β^+ in two distinct ways is shown in Figs. 7 and 8.

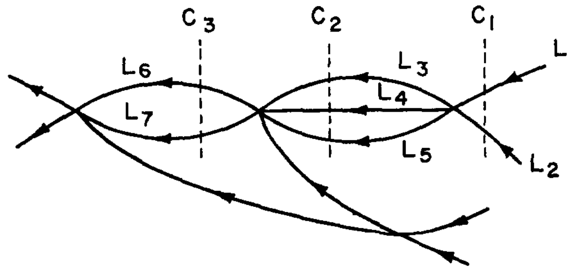


FIG. 5. The cuts $C_1 = (L_1, L_2)$ and $C_2 = (L_3, L_4, L_5)$ are equivalent. C_2 is a left-most cut. $C_3 = (L_6, L_7)$ is not equivalent to C_1 or C_2.

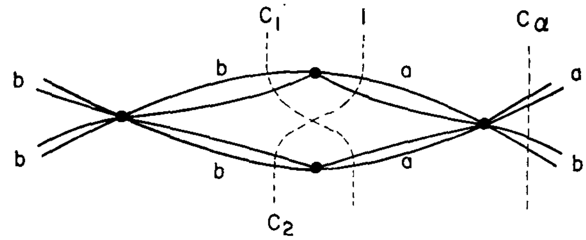


FIG. 6. A diagram with two leftmost cuts equivalent to C_α. We take $M_a < M_b$. Throughout this work it is assumed that the mass values of the stable particles have no accumulation points. It is then easy to see that the leftmost cut is unique in some finite neighborhood of the α threshold.

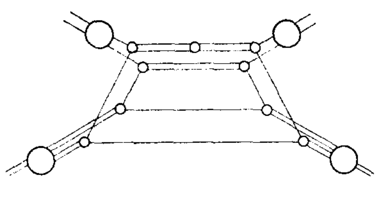


FIG. 7. A bubble diagram B that contains a certain D_β^+ in two essentially different ways. This D_β^+ is shown in Fig. 8.

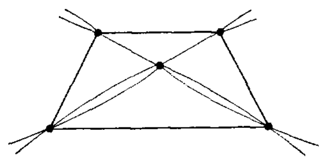


FIG. 8. A D_β^+ that is contained in two essentially different ways in the B of Fig. 7.

APPENDIX C: STRONGLY EQUIVALENT CUTS

In this appendix we show that any cut C_α corresponding to α of D_β^+ can be replaced by the leftmost cut C_α' strongly equivalent to it without destroying its correspondence to α of D_β^+.

The condition that B^- contain D_β^+ is equivalent to the condition that there is a continuous mapping $\Gamma: D(B^-) \to D_\beta^+$ that maps $D(B^-)$ onto D_β^+. The external lines of $D(B^-)$ must map onto the external lines of D_β^+ identified with them. The lines of the cuts $C_\alpha = \Gamma^{-1}\alpha$ are in 1-to-1 correspondence with the lines of α. The inverse image $\Gamma^{-1}V$ of vertex V of D_β^+ is the part of $D(B^-)$ that corresponds to V.

The point $\bar{P}$ is assumed to satisfy the following conditions:

(1) $\bar{P}$ lies on $L(D_\beta^+)$.

(2) $\bar{P}$ lies on $L(D^+)$ only if D^+ contains D_β^+.

(3) The solution of the Landau equation of D_β^+ at $\bar{P}$ defines momentum–energy vectors p_j such that no line of any set α of D_β^+ has its p_j parallel to that of any line of any other set α of D_β^+. As before, α runs over pairs of vertices of D_β^+ and specifies the set of lines L_j running between that pair of vertices.

We make use of one important kinematic result: If $D(B^-)$ contains D_β^+, then the equations of energy–momentum and mass constraint alone require that, if the external lines of $D(B^-)$ have the $\bar{p}_j$ defined by $\bar{P}$, then the unique values of the p_j of the lines of $C_\alpha = \Gamma^{-1}\alpha$, subject to the conservation-law and mass-shell constraints on these lines, are those defined by the Landau equations of D_β^+ at $\bar{P}$. This result is closely connected to the fact that $L(D_\beta^+)$ lies on the boundary of the physical region of D_β^+ and is proved in the same way.[10,3]

The arguments in the text are purely topological. In this appendix we make use also of the kinematic requirement just described. That is, we shall require that the contribution to the integral corresponding to B^- actually satisfy the energy–momentum conservation laws required at $\bar{P}$. By considering a sufficiently small neighborhood of $\bar{P}$, the internal p_j can be confined to an arbitrarily small neighborhood of the values required at $\bar{P}$. Thus, we can consider the p_j of the lines of the various sets C_α to be in a small neighborhood of the values defined by the Landau equations.

At $\bar{P}$ the momentum–energy vectors of the various lines corresponding to any single C_α are all parallel, by virtue of the Landau equations. In some particular Lorentz frame they are all at rest. Consider any C'_α strongly equivalent to C_α. Since C'_α and C_α define the same set of flow lines their total energy momentum is the same. Since the total rest masses are also equal, the lines of C'_α must also all correspond to particles at rest, in this particular frame.

We now prove the following result: If C'_α is strongly equivalent to C_α and lies left of it, then C'_α lies in $\Gamma^{-1}V$, where V is the vertex of D_β^+ upon which the set α terminates:

Let β label the various outgoing sets of lines of V and let $C_\beta = \Gamma\beta$. The momentum–energy vectors of the lines of C_β are, by assumption, not parallel to those of C_α. Thus, no line of C'_α can coincide with any line of any C_β. Thus, C'_α must either lie completely within $\Gamma^{-1}V$, or there is a part of $D(B^-)$ that consists of a set of paths that begin with certain lines of the sets C_β and end with certain lines of C'_α. Let this part of $D(B^-)$ be called Q. We wish to show that Q is necessarily empty, i.e., that C'_α lies in $\Gamma^{-1}V$.

Consider ΓQ, the image of Q in D_β^+. The energy–momentum conservation requirements at $\bar{P}$ can be satisfied only if the lines of ΓQ carry the momentum–energy prescribed by the Landau equations, as already noted. But, if the momentum–energy vectors are as prescribed by the Landau equations, then the vectors $\alpha_i p_i = \Delta x_i$ can be interpreted as space–time displacements: These displacements must fit together to give a classical multiple scattering process. But then the arguments of Ref. 9 immediately rule out the possibility that Q is nonempty; for the initial particles of ΓQ all start at the common vertex V, and they diverge from that point. It is then not possible that they transform by multiple scattering into a set of particles all relatively at rest, without allowing extra particles that come in from outside (i.e., that do not start at V). But interactions with extra incoming particles that do not start at V are incompatible with the condition that C'_α be strongly equivalent to C_α.

Thus ΓQ must be empty, and C'_α must therefore lie completely in $\Gamma^{-1}V$.

But, if C'_α lies completely in $\Gamma^{-1}V$, then it can be used in place of C_α in making the correspondence of $D(B^-)$ to D_β^+: The topological structure is not altered by replacing C_α by the leftmost cut C'_α that is strongly equivalent to it. This is the result that we need. A slight alternation of the argument shows that C_α can be replaced by any cut strongly equivalent to it without disrupting the correspondence to α of D_β^+.

* This work was done under the auspices of the U.S. Atomic Energy Commission.

† Present address: Western Illinois University, Macomb, Ill.

[1] R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960).
[2] R. E. Cutkosky, in *S-Matrix Theory of Strong Interactions*, G. Chew, Ed. (Benjamin, New York, 1961), p. 181.
[3] J. Coster and H. P. Stapp, J. Math. Phys. **10**, 371 (1969) (earlier works are cited here).
[4] M. J. Bloxham, D. I. Olive, and J. C. Polkinghorne, J. Math. Phys. **10**, 494, 545, 553 (1969) (earlier works are cited here).
[5] H. P. Stapp, Phys. Rev. **125**, 2139 (1962).
[6] L. D. Landau, Nucl. Phys. **13**, 181 (1959).
[7] D. Iagolnitzer and H. P. Stapp, Commun. Math. Phys. **14**, 15 (1969); D. Iagolnitzer, in *Lectures in Theoretical Physics* (Gordon and Breach, New York, 1969), Vol. XID, p. 221.
[8] C. Chandler and H. P. Stapp, J. Math. Phys. **10**, 826 (1969).
[9] H. P. Stapp, J. Math. Phys. **8**, 1606 (1967).
[10] F. Pham, Ann. Inst. Henri Poincaré **6A**, 89 (1967).
[11] See Appendix A for further details.
[12] H. P. Stapp, J. Math. Phys. **9**, 1548 (1968), and UCRL, Reprint No. 16816, 1966.
[13] See Appendix B for further details.
[14] See Appendix C for further details.
[15] J. Coster and H. P. Stapp, J. Math. Phys. **11**, 1441 (1970).
[16] The existence of such mixed-α diagrams was first pointed out by D. Branson, Nuovo Cimento **44A**, 1081 (1966); **54A**, 217 (1968).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Relationship between Geometrical-Optical and Full-Wave Solutions to the Problem of Propagation over a Nonparallel Stratified Medium

E. BAHAR
Electrical Engineering Department, University of Nebraska, Lincoln, Nebraska 68508

(Received 16 March 1970)

The problem of propagation over a wedge-shaped overburden has been analyzed using a full-wave-solution approach. The relationship between this solution and an earlier one employing the compensation theorem is discussed in detail. The compensation theorem method makes use of the simplifying concept of of the surface impedance which is determined by a geometrical-optical approach. The applicability and limitations of the latter approach are carefully studied for both dissipative and nondissipative overburdens.

1. INTRODUCTION

In earlier studies of mixed-path propagation,[1] the earth's surface has been characterized by a sectionally constant surface impedance. Thus, an abrupt change in the surface impedance along the path of propagation characterizes a land–sea boundary. This representation of the earth's surface has yielded considerable information on problems of mixed-path propagation, and these results have been borne out in controlled, idealized laboratory experiments.[2] The analysis of this problem using the compensation theorem[3,4] indicates rapid fluctuations of the electromagnetic fields at the land–sea boundary, which are attributed to the idealized surface-impedance profile. To overcome this problem, such that the results would resemble physical situations more closely, it may seem reasonable to assume that the surface impedance varies linearly along a finite transition region between the land–sea boundaries. However, attempts to investigate the sloping-beach problem have indicated that the tangential electric- to magnetic-field ratio over a wedge-shaped subsurface region (overburden) does not vary linearly with distance along the path of propagation (see Fig. 1). Subsequently, it was suggested[5] that the variable surface impedance in the transition region may be assumed to be equal to the surface impedance of a parallel-stratified earth of depth equal to the local depth $h(x)$. In a recent study of the influence of the earth's subsurface on electromagnetic ground-wave propagation,[6] a geometrical–optical technique is first used to determine the surface impedance over a wedge-shaped overburden. This geometrical–optical approach, which has been used even when the overburden is a dissipative medium, was suggested intuitively since it converges to the exact solution as the wedge angle approaches zero (parallel stratification). However, it is found from these computations that the value of the surface impedance depends upon the direction of propagation of the incident wave. These values being used for the surface impedance, the compensation theorem was employed to derive an integral equation for the electromagnetic fields.[6] Schlak and Wait[7] later carried out a reciprocity test using these results. These tests indicate that the computed mutual impedances between two dipoles (at the surface of the earth) are dependent upon the direction of propagation. However, this discrepancy (attributed to the approximations inherent in using the geometrical optical approach) decreases with increasing distance between the two dipoles. It was conjectured at the time by Schlak and Wait[6] that the surface-impedance calculations are more dependable when the propagation is in the direction of increasing overburden depth. However, the results of preliminary laboratory experiments at 4.75 GHz, conducted recently by King and Hustig,[5] indicate that the contrary would be true.

In a recent full-wave solution of the problem of propagation over a layered medium of arbitrarily varying depth,[8] the electromagnetic fields are expressed in terms of a complete local set of basis functions consisting of surface-wave and radiation terms. Since no characteristic-function expansion is available, this approach reduces the problem to the solution of coupled first-order differential equations (for the forward- and backward-wave amplitudes) similar to those encountered in nonuniform waveguide structures. A major advantage of this method is that it requires neither the assumption nor calculation of the surface impedance, nor is the solution restricted by the surface-impedance concept. Indeed, it is possible to use these results to compute the tangential electric- to magnetic-field ratio at the earth's surface. Thus, this analysis

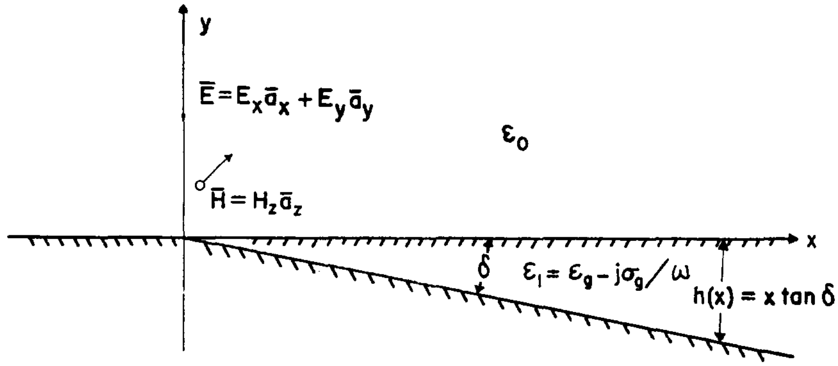


FIG. 1. Vertically polarized waves over a wedge-shaped overburden.

unables one to examine critically the use of the surface-impedance concept when the tangential electric to magnetic field ratio is rapidly varying. This analysis also assists in resolving some of the questions arising from the geometrical–optical analysis[6] and recent laboratory results of King and Hustig.[5]

In this paper, we derive the direct relationship between the geometrical–optical solution to the idealized problem with the wedge-shaped overburden and the full-wave solution which is valid for overburdens of arbitrarily varying depth. On deriving this relationship, we shall be able to justify the use of the geometrical–optical approach (even for the case in which the overburden is a dissipative medium) and to examine carefully the restrictions that must be made upon the use of this approach. The difficulties arising in the use of the geometrical–optical approach in the region near the apex of the wedge-shaped overburden are also examined. Finally, we determine the reason why the geometrical–optical solution is more in accord with the reciprocity theorem as the distance between transmitting and receiving dipoles increases.

2. RELATIONSHIP BETWEEN THE FULL-WAVE SOLUTION AND THE GEOMETRICAL–OPTICAL SOLUTION

A. Coupling Effects Neglected

Using the full-wave analysis of the problem of propagation of vertically polarized radio waves over a layered surface of nonuniform thickness (see Fig. 1), we have shown that the z-directed magnetic field can be expressed in terms of a complete set of basis functions consisting of the surface wave and the radiation terms,[8] as

$$H_z(x, y) = \sum_{n=1}^{N} I(x, \beta_n)\phi(y, \beta_n) + \int_\Gamma I(x, \beta)\phi(y, \beta)\, d\beta, \tag{2.1}$$

where the finite discrete sum is over the surface waves satisfying the modal equation

$$\tan u_1 h + \frac{i\epsilon_1}{\epsilon_0}\frac{u_0}{u_1} = 0, \tag{2.2a}$$

in which ϵ_0 and ϵ_1 are the dielectric coefficients for free space and for the overburden, respectively. The region below the overburden is, for simplicity, assumed to be highly conducting. The wave parameters are

$$u_{0,1} = (k_{0,1}^2 - \beta^2)^{\frac{1}{2}}, \quad k_{0,1} = \omega(\mu_0\epsilon_{0,1})^{\frac{1}{2}}. \tag{2.2b}$$

The square roots are chosen such that, for the surface waves, u_0 and u_1 are in the first quadrant and β is in the fourth quadrant.

The permeability for free space and the overburden is μ_0. The path of integration Γ, for the radiation term, is

$$k \geq \beta \geq 0 \quad \text{and} \quad 0 < i\beta < \infty. \tag{2.3a}$$

Thus,

$$0 \leq u_0 < \infty. \tag{2.3b}$$

In this paper, we are assuming an $e^{i\omega t}$ time dependence. The generalized currents $I(x, \beta)$ are only functions of the coordinate x. However, the local basis functions $\phi(y, \beta)$, which satisfy all the boundary conditions for a parallel-stratified medium of thickness equal to the local depth of the overburden $h(x)$ (see Fig. 2), are explicitly functions of y and implicitly

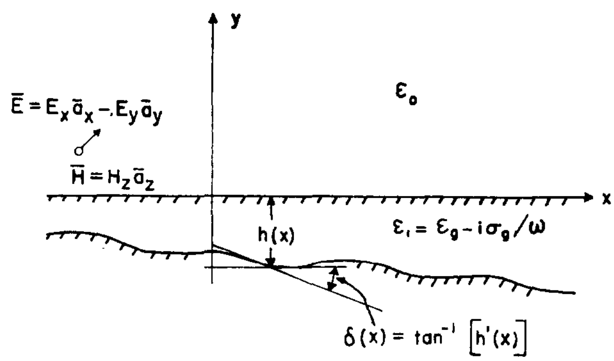


FIG. 2. Nonuniform overburden of depth $h(x)$.

functions of x except for regions where the overburden is of uniform depth:

$$\phi(y,\beta) = [N(\beta)]^{-1} \times \cos u_1(y+h), \quad -h < y < 0,$$
$$= [N(\beta)]^{-1} \times \left[\cos u_1 h \cos u_0 y - \frac{\epsilon_0 u_1}{\epsilon_1 u_0} \sin u_1 h \sin u_0 y\right], \quad y > 0. \quad (2.4)$$

The normalization coefficient $N(\beta)$ is chosen such that

$$\int_{-h}^{\infty} Z(\beta)\phi(y,\beta)\phi(y,\beta')\,dy = \delta_{\beta,\beta'}, \quad (2.5a)$$

for the surface-wave term, and

$$\int_{-h}^{\infty} Z(\beta)\phi(y,\beta)\phi(y,\beta')\,dy = \delta(\beta,\beta'), \quad (2.5b)$$

for the radiation term. The wave impedance is

$$Z(\beta) = Z_1(\beta) = \beta/\omega\epsilon_1, \quad -h < y < 0,$$
$$= Z_0(\beta) = \beta/\omega\epsilon_0, \quad y > 0, \quad (2.6)$$

where $\delta_{\beta,\beta'}$ and $\delta(\beta'-\beta')$ are the Kronecker and Dirac δ functions, respectively. Thus, for the radiation term,[8]

$$N^2(\beta) = Z_0(\beta)\left[\cos^2 u_1 h + \left(\frac{\epsilon_0}{\epsilon_1}\frac{u_1}{u_0}\right)^2 \sin^2 u_1 h\right]\frac{\pi}{2}\left(\frac{-u_0}{\beta}\right) \quad (2.7a)$$

and, for the surface waves,

$$N^2(\beta) = \frac{iZ_0(\beta)}{2u_0}\left[\cos^2 u_1 h + \left(\frac{\epsilon_0}{\epsilon_1}\right)^2(\sin^2 u_1 h + u_1 h \tan u_1 h)\right]. \quad (2.7b)$$

Note that for the surface waves the expression for $\phi(y,\beta)$ reduces to

$$\phi(y,\beta) = \cos u_1 h \exp(iu_0 y)/N(\beta), \quad y > 0, \quad (2.4')$$

and the radiation term consists of a continuous spectrum of standing waves along the y direction.[9] The generalized current $I(x,\beta)$ can be expressed in terms of coupled forward- and backward-wave amplitudes $a(x,\beta)$ and $b(x,\beta)$. Thus, in general,

$$I(x,\beta) = a(x,\beta) + b(x,\beta). \quad (2.8a)$$

However, when the coupling between the waves that constitute the general solution (2.1) is negligible, they can be shown to reduce to

$$a(x,\beta) \to a(0,\beta)e^{-i\beta x}$$
and
$$b(x,\beta) \to b(0,\beta)e^{i\beta x}, \quad (2.8b)$$

where $a(0,\beta)$ and $b(0,\beta)$ are the wave amplitudes at $x = 0$.

The wave parameters $u_{0,1}$ and β are related to the complex cosine and sine of the angles of propagation in the air and in the overburden θ_0 and θ_1, respectively. Thus,

$$u_{0,1} = k_{0,1}\cos\theta_{0,1} \quad \text{and} \quad \beta = k_0 \sin\theta_0 = k_1 \sin\theta_1.$$

The second relationship constitutes Snell's law, and, for the surface waves, these complex angles of propagation have been shown to be related to the Brewster angle.[10]

In this section, we assume that coupling is negligible. Thus, for a wave incident at an angle θ_0^i in the region above the overburden, the magnetic field may be expressed in terms of the basis function expansion (2.1) as

$$H_z^i(x,y) \approx \exp[-i(\beta^i x - u_0^i y)] + \left(\frac{\cos u_1^i h - i\alpha^i \sin u_1^i h}{\cos u_1^i h + i\alpha^i \sin u_1^i h}\right)\exp[-i(\beta^i x + u_0^i y)]$$
$$= \exp(-i\beta^i x)\phi(y,\beta^i)2N(\beta^i)/(\cos u_1^i h + i\alpha^i \sin u_1^i h), \quad (2.9a)$$

in which α^i is the normalized wave impedance,

$$\alpha^i = \frac{\epsilon_0 u_1^i}{\epsilon_1 u_0^i} = \frac{\eta_1 \cos\theta_1^i}{\eta_0 \cos\theta_0^i}, \quad (2.9b)$$

where the $\eta_{0,1}$ are the intrinsic impedances for free space and for the overburden, respectively, and

$$\beta^i = k_0 \sin\theta_0^i = k_1 \sin\theta_1^i \quad \text{and} \quad u_{0,1}^i = k_{0,1}\cos\theta_{0,1}^i. \quad (2.9c)$$

Note that, for the case $h(x) = \text{const}$, $\phi(y,\beta)e^{-i\beta x}$ consists of the incident and the specularly reflected uniform plane waves

$$\exp[-ik_0(\sin\theta_0^i x - \cos\theta_0^i y)]$$

and

$$R_T \exp[-ik_0(\sin\theta_0^i x + \cos\theta_0^i y)],$$

respectively. The composite reflection coefficient R_T, at the surface $y = 0$, can be expressed in terms of the reflection and transmission coefficients at the plane interface between two semi-infinite media with permittivities ϵ_1 and ϵ_0. Thus,

$$R_T \equiv (\cos u_1^i h - i\alpha^i \sin u_1^i h)/(\cos u_1^i h + i\alpha^i \sin u_1^i h)$$
$$= (R_{00} + e^{-i2u_1 h})/(1 - R_{11}e^{-2iu_1 h}), \quad (2.10a)$$

where R_{00} and R_{11} are the reflection coefficients for the waves incident from media ϵ_0 and ϵ_1, respectively; thus

$$R_{00} = -R_{11} = (1-\alpha^i)/(1+\alpha^i). \quad (2.10b)$$

On expanding the denominator in (2.10a) in a binomial series, we get

$$R_T = R_{00} + T_{01}T_{10}e^{-i2u_1'h} + T_{01}R_{11}T_{10}e^{-i4u_1'h} + T_{01}(R_{11})^2T_{10}e^{-i6u_1'h} + \cdots + T_{01}(R_{11})^{n-1}T_{10}e^{-i2nu_1'h} + \cdots, \quad (2.10c)$$

in which the transmission coefficients T_{10} and T_{01} are

$$T_{10} = 1 + R_{00} \quad \text{and} \quad T_{01} = 1 + R_{11}. \quad (2.10d)$$

The above expression for R_T constitutes the direct and the multiply reflected waves in the overburden. This expression for R_T is valid for both dissipative and nondissipative overburdens.

For a wedge-shaped overburden, we let the z axis lie along the apex of the wedge (see Fig. 1). Thus,

$$h(x) = x \tan \delta, \quad (2.11)$$

where δ is the wedge angle. In this case, R_T is also a function of x. Thus, the reflected component of (2.9a) is not a specularly reflected uniform plane wave. We can now show readily that, under certain conditions, this inhomogeneous reflected wave can be approximated by a finite series which can be identified with the geometrical-optical series used by Schlak and Wait[6] to determine the surface impedance of the structure. Substituting the expressions for R_T [(2.10c)] and $h(x)$[(2.11)] in the expression for the reflected component, we get

$$\begin{aligned} R_T \exp [-ik_0(\sin \theta_0^i x + \cos \theta_0^i y)] \approx{} & R_{00} \exp [-ik_0(\sin \theta_0^i x + \cos \theta_0^i y)] \\ & + T_{01}T_{10} \exp [-ik_0(\sin \theta_0^1 x + \cos \theta_0^1 y)] \\ & + T_{01}R_{11}T_{10} \exp [-ik_0(\sin \theta_0^2 x + \cos \theta_0^2 y)] + \cdots \\ & + T_{01}(R_{11})^{N-1}T_{01} \exp [-ik_0(\sin \theta_0^N x + \cos \theta_0^N y)], \end{aligned}$$

where

$$2N\delta < \tfrac{1}{2}\pi - \theta_1^i < 2(N + 1)\delta. \quad (2.12)$$

In the above expression, we have assumed that the following approximations are valid:

$$\begin{aligned} \sin \theta_1^1 &= \sin \theta_1^i + \cos \theta_1^i \, 2 \tan \delta \\ &= [\sin \theta_1^i(\cos 2\delta - \sin 2\delta) + \cos \theta^i \sin 2\delta]/\cos^2 \delta \\ &\approx \sin \theta_1^i \cos 2\delta + \cos \theta_1^i \sin 2\delta \\ &= \sin (\theta_1^i + 2\delta). \end{aligned} \quad (2.13a)$$

In (2.13a), we assume that $|\delta| \ll \frac{1}{4}\pi$ and neglect terms of order δ^2. Similarly,

$$\sin \theta_1^n \approx \sin (\theta_1^i + 2n\delta). \quad (2.13b)$$

Obviously, as n increases, this approximation becomes less valid. Thus, we must also require

$$T_{01}(R_{11})^n T_{01} \ll 1 \quad \text{as} \quad n\delta \to \tfrac{1}{4}\pi.$$

This condition simply assures that the amplitude of the nth multiply reflected wave is negligibly small as the approximation implied in (2.13b) becomes invalid. In addition, we assume that

$$\cos (\theta_1^i + 2\delta) = \cos \theta_1^i \cos 2\delta - \sin \theta_1^i \sin 2\delta \approx \cos \theta_1^i. \quad (2.13c)$$

For the above condition to be satisfied to within the same approximation as (2.13a), we must require, in addition, that

$$|\theta_1^i| \ll \tfrac{1}{4}\pi. \quad (2.13d)$$

The angles θ_1^n (for $y > 0$) are related to θ_0^n for the overburden through Snell's law:

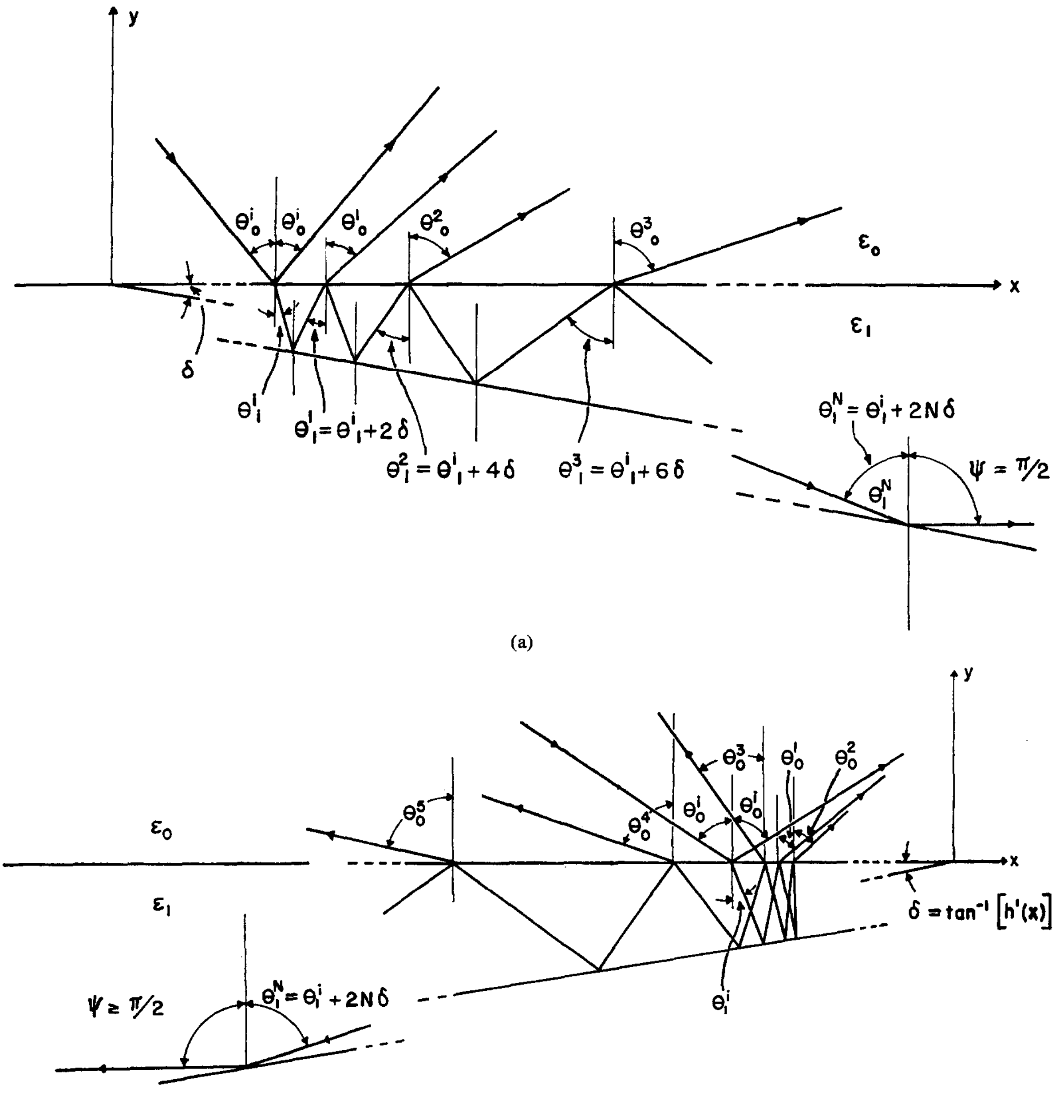
$$\beta^n = k_0 \sin \theta_0^n = k_1 \sin \theta_1^n. \quad (2.13e)$$

Summing up, we note that, for all the nonnegligible terms of the finite-series expansion (2.12), we require that

$$|\theta_1^i + 2n\delta| \ll \tfrac{1}{2}\pi. \quad (2.14)$$

The geometrical-optical-series expansion is illustrated in Fig. 3. In Figs. 3(a) and 3(b) the "incident" waves are assumed to be traveling in the direction of increasing and decreasing overburden depth $h(x)$, respectively. Thus, for case (a) $\delta > 0$, and for case (b) $\delta < 0$. For simplicity, ϵ_1 is assumed real in these figures. Examining the finite geometrical-optical series, we observe that, for $|\mathrm{Re}\,(\theta_1^i) + 2\delta| \to \frac{1}{2}\pi$, the expression consists of only the incident and the specularly reflected wave, a situation in which it is obvious that the geometrical-optical solution is invalid unless, of course, $|T_{01}T_{10}| \to 0$. Note that, if the overburden is nondissipative, β^n is real, and the multiply reflected waves both in and above the overburden are homogeneous plane waves. However, for a dissipative overburden, the multiply reflected terms of the geometrical-optical series, both in and above the overburden, are inhomogeneous waves since β^n is a complex number in the fourth quadrant. Thus, for these terms, the planes of constant phase do not coincide with the planes of constant amplitude. Note that condition (2.13d) on the complex angle θ_1^i in the overburden is a restriction, not only upon the real angle of incidence θ_0^i but also upon the complex value of the dielectric coefficient for the overburden ϵ_1.

It is interesting to point out that, for the limiting case in which $\epsilon_1 \to \epsilon_0$, the reflection coefficient R_{11} approaches 0 and the local modal expansion reduces to the incident wave and a single reflected wave $\exp [-ik_0(\sin \theta_0^1 x + \cos \theta_0^1 y)]$[(2.12)]. The latter component, of course, is a wave specularly reflected by the tilted conducting plane $y = x \tan \delta$ (and not the plane



(a)

(b)

FIG. 3. Geometrical–optical representation of the wave travelling in (a) the direction of increasing overburden depth, $\delta > 0$, (b) the direction of decreasing overburden depth, $\delta < 0$.

$y = 0$). This point emphasizes the advantages of using a local modal expansion [one that locally satisfies the boundary conditions of a parallel stratified overburden of local depth $h(x)$]. Use of a fixed modal expansion for the electromagnetic field [one in which $\phi(y, \beta)$ is strictly a function of the y coordinate only] would result in strong coupling between the component waves, since, in this case, $\phi(y, \beta)$ describes an incident wave and a wave specularly reflected by the surface $y = 0$ rather than the plane $y = x \tan \delta$. This aspect of the problem has also been noted in an earlier paper.[8] In the following section, we consider the problem of coupling between the terms of the local expansion and examine how the distances between the input and output reference planes and the finite conductivity of the overburden affect the validity of the geometrical–optical solutions.

B. Consideration of Coupling Effects

In the preceding section, we considered the case in which the forward- and backward-wave amplitudes $a(x, \beta)$ and $b(x, \beta)$ could be regarded as uncoupled

[(2.8b)]. However, for the general case (following the rigorous full-wave analysis[8]), the wave amplitudes are shown to satisfy the coupled first-order differential equations

$$-\frac{d}{dx}a(x,\beta') = i\beta' a(x,\beta') + \sum S^{BA}(\beta',\beta)a(x,\beta) + \sum S^{AA}(\beta',\beta)b(x,\beta), \quad (2.15a)$$

$$-\frac{d}{dx}b(x,\beta') = -i\beta' b(x,\beta') + \sum S^{AB}(\beta',\beta)b(x,\beta) + \sum S^{BB}(\beta',\beta)a(x,\beta), \quad (2.15b)$$

where $\sum$ indicates a summation of the finite number of surface waves and an integral over the contour as described by (2.1). The scattering coefficients S^{BA} and S^{AA} are defined as

$$S^{BA}(\beta',\beta) = \frac{1}{2}\frac{\epsilon_0 Z_0(\beta')\, dh/dx}{\epsilon_1 N(\beta')N(\beta)}\frac{u_1^2}{u_1'^2 - u_1^2}\left[1 + \frac{\beta}{\beta'}\left(\frac{u_1'}{u_1}\right)^2\right], \quad \beta' \neq \beta,$$
$$= 0, \quad \beta' = \beta,$$

and

$$S^{AA}(\beta',\beta) = \frac{1}{2}\frac{\epsilon_0 Z_0(\beta')\, dh/dx}{\epsilon_1 N(\beta')N(\beta)}\frac{u_1^2}{(u_1'^2 - u_1^2)}\left[1 - \frac{\beta}{\beta'}\left(\frac{u_1'}{u_1}\right)^2\right], \quad \beta' \neq \beta,$$
$$= -\frac{1}{2}\left(\frac{Z_0(\beta')\, dh/dx}{N^2(\beta')\epsilon_1/\epsilon_0} + \frac{1}{Z_0(\beta')}\frac{d}{dx}Z_0(\beta')\delta_{\beta',\beta_n'}\right), \quad \beta' = \beta, \quad (2.16a)$$

where $\delta_{\beta',\beta_n'}$ is the Kronecker δ function which vanishes unless the incident wave is a surface wave (characterized by the discrete parameter β_n'), and

$$S^{BA}(\beta',\beta) = S^{AB}(\beta',\beta), \quad S^{AA}(\beta',\beta) = S^{BB}(\beta',\beta). \quad (2.16b)$$

Obviously, (2.8b) is the solution to (2.15) if the scattering coefficients vanish. In order to solve (2.15) for the general case, we first determine the forward mode amplitude $a^i(x,\beta)$ for the incident magnetic field $H_z^i(x,y)$[(2.9a)]. Using (2.1) and the orthogonal properties of the basis functions $\phi(y,\beta)$[(2.5)] and noting that $I^i(x,\beta) \to a^i(x,\beta)$, we find that

$$a^i(x,\beta) = \frac{2N(\beta^i)e^{-i(\beta^i x)}}{\cos u_1^i h + i\alpha^i \sin u_1^i h} \times \begin{cases}\delta_{\beta_n,\beta^i}, & \text{surface wave,}\\ \delta(\beta - \beta^i), & \text{radiation term.}\end{cases} \quad (2.17)$$

We now seek an iterative solution of (2.15). Thus (neglecting reconversion) we substitute (2.17) for $a(x,\beta)$ in the right-hand side of (2.15a) and neglect $b(x,\beta)$ for the first-order iterative solution:

$$-\frac{d}{dx}[a(x,\beta')] - i\beta' a(x,\beta') = \frac{S^{BA}(\beta',\beta^i)2N(\beta^i)T_{10}^i \exp[-i(\beta^i x + u_1^i h)]}{1 - R_{11}^i \exp(-i2u_1^i h)}. \quad (2.18)$$

The solution of the above inhomogeneous equation for the scattered waves ($\beta' \neq \beta^i$) is[11]

$$a(x,\beta') = -2T_{10}^i \exp(-i\beta' x)\int_0^x \frac{S^{BA}(\beta',\beta^i)N(\beta')}{1 - R_{11}^i \exp(-i2u_1^i h)} \times \exp[i(\beta' - \beta^i - u_1^i \tan\delta)\zeta]\, d\zeta. \quad (2.19)$$

In the above, we have replaced x by the variable of integration ζ in the integrand. The coefficient $N^2(\beta)$ can be expressed as follows:

$$N^2(\beta) = \frac{Z_0(\beta)}{(T_{10})^2}(1 - 2R_{11}\cos 2u_1 h + R_{11}^2)\frac{\pi}{2}\left(\frac{-u_0}{\beta}\right), \quad \text{radiation term,} \quad (2.20a)$$

$$= \frac{iZ_0(\beta)}{2u_0}\left[\left(\frac{u_1^2 - u_0^2}{u_1^2}\right)\cos^2 u_1 h + \left(\frac{\epsilon_0}{\epsilon_1}\right)^2 u_1 h \tan u_1 h\right], \quad \text{surface waves.} \quad (2.20b)$$

For a nondissipative overburden, k_1^2 is real and u_1, for the radiation term, is also real. Thus, for $|R_{11}^i| \ll 1$, the coefficient $N^2(\beta)$[(2.20a)] is a very slowly varying function of x for the entire region $0 < x < \infty$. The denominator in the integrand of (2.19) can be expanded as a rapidly converging binomial series, and (2.19) may be integrated term by term. Thus, replacing $N^2(\beta)$ by its value for $h = 0$,

$$N^2(\beta) \approx -Z_0(\beta)\pi u_0/2\beta.$$

It now follows that

$$a(x,\beta') \approx -2\exp(-i\beta' x)S^{BA}(\beta',\beta^i)N(\beta^i)\sum_{n=0}^{\infty}(R_{11}^i)^n \times \int \exp(i\{\beta' - [\beta^i + (2n+1)u_1^i \tan\delta]\}\zeta)\, d\zeta. \quad (2.21)$$

Using the unilateral Fourier-transform pair, we note that

$$g(\beta') = \frac{1}{2\pi}\int_0^\infty\int_{-\infty}^\infty g(\beta)e^{i(\beta'-\beta)x}\, d\beta\, dx. \quad (2.22)$$

Thus,

$$\int_0^x \exp\,[i(\beta' - \beta_s^n)\zeta]\,d\zeta$$
$$\to \{1 - \exp\,[i(\beta' - \beta_s^n)x]\}/i(\beta' - \beta_s^n), \quad \text{for } x \text{ finite},$$
$$\to 2\pi\delta(\beta' - \beta_s^n), \quad \text{for } x \to \infty, \tag{2.23}$$

where, under the conditions established for the geometrical–optical approximations,

$$\beta_s^n = \beta^i + (2n+1)u_1^i \tan\delta \approx k_1 \sin\,[\theta_1^i + (2n+1)\delta]. \tag{2.24}$$

Thus, for the scattered field, the wave amplitude $a(x, \beta')$ becomes a very directive function of β' as $x \to \infty$. Even for finite values of x, it is obvious that $a(x, \beta')$ has maxima at $\beta' = \beta_s^n$ [i.e., when the phase in the integrand (2.21) is zero]. The forward scattered radiation field $H_z^s(x, y)$ is

$$H_z^s(x, y) = \int_\Gamma a(x, \beta')\phi(y, \beta')\,d\beta'. \tag{2.25}$$

Thus, taking into account the x dependence of the local basis function $\phi(y, \beta')$, it can be shown that the scattered field also radiates primarily in the directions θ_0^n[(2.13e)], as predicted by the geometrical–optical approach, provided, of course, all the conditions stipulated above are satisfied. The above discussion, therefore, also explains why the discrepancies in the geometrical–optical approach decrease as the distance between the transmitting and receiving dipoles increases.[7] In a similar manner, it is possible to obtain a first-order iterative solution for the backward-wave amplitude $b(x, \beta')$ starting with (2.15b). It should be pointed out that, using the geometrical–optical approach, we neglect these backward waves for the case in which the incident wave is traveling in the $+x$ direction (see Fig. 1). Higher-order iterative solutions for both $a(x, \beta')$ and $b(x, \beta')$ may be generated from the first-order solutions in a straightforward manner, and it can be shown that they exhibit the same directive properties as implied by (2.24).

We now consider the coupling into the various surface waves. Note that, even for the case in which the overburden is nondissipative, the surface-wave parameters u_0 and u_1 are complex. Thus, the coefficients $N^2(\beta)$[(2.20b)] are strongly dependent on the variable x. Moreover, for large values of x, $N(\beta)$ increases exponentially as $\exp\,(-iu_1 \tan\delta\, x)$ and, in view of (2.16), the coupling decreases exponentially as is expected. Thus, coupling into the surface waves is restricted to the region in the vicinity of the wedge apex, and the forward-wave amplitude $a(x, \beta)$ does not exhibit sharp maxima in the directions of the complex Brewster angles (surface-wave terms). It is for this reason that the geometrical–optical approach is restricted to regions distant from the wedge apex.[6]

Turning now to the case in which the overburden is dissipative, k_1 (and therefore u_1) is also complex for both the radiation and surface-wave terms. Thus, for the radiation term in this case, $N^2(\beta)$ is a slowly varying function of x only for the region in which

$$|2R_{11} \cos 2u_1 h/(1 - R_{11}^2)| \ll 1.$$

It is in this region that most of the coupling takes place. For large values of x, $N(\beta)$ increases exponentially with distance x, as in the case of the surface waves. As should be expected, the coupling decreases (exponentially) as x increases (see Fig. 2). Hence, when the overburden is dissipative, the scattered-wave amplitude $a(x, \beta')$ is not as directive a function as in the case of the nondissipative overburden. However, for the case of the dissipative overburden, the scattered fields H_z^s are less significant,and the zero-order solution (2.17) may be sufficient for most practical cases. Hence, in this case too, the geometrical–optical approach may yet be used to obtain an appropriate value for the surface impedance.

In the above analysis, we have restricted our attention to the case in which the incident wave is propagating in the $+x$ direction. For the case in which the direction of propagation is reversed, we need only invoke the reciprocity relationships, since the full-wave analysis is consistent with reciprocity.[8] Experimental evidence implies that the reflected-wave amplitude $b(x, \beta)$ may be neglected in cases of practical interest[8] when the incident field is propagating in the $+x$ direction. Thus, the indirect approach, involving the use of the reciprocity relationships, is preferable to the direct approach when the incident wave is propagating in the negative x direction.[12]

3. CONCLUDING REMARKS

In this paper, we have derived the relationship between the full-wave solution and a geometrical–optical approach to the problem of propagation over a wedge-shaped overburden. We have examined the applicability of the latter approach for both dissipative and nondissipative overburdens and determined the limitations of this approach relative to the magnitude of the wedge angle δ and the angle of incidence θ_1^i in the overburden. We have also resolved certain questions arising from the directive property of the calculated surface impedance and its relationship to the reciprocity theorem.[7] The question of the excitation of the surface waves in the vicinity of the wedge apex (neglected by the geometrical–optical approach) has also been investigated.

Finally, we wish to recall that, when applicable, the geometrical-optical approach is a convenient method to determine the appropriate value of the surface impedance. However, the actual field variations still need to be computed. Through a judicious use of the compensation theorem, Wait[1] derives an integral equation for the field which is solved by an iterative method. Thus, the latter solution depends upon the applicability of the surface impedance concept. For the surface impedance concept to be meaningful, it must be insensitive to variations of the angle of incidence θ_0^i. Furthermore, the fluctuations in the relative value of the surface impedance must be limited.[6] On the other hand, the full-wave solution (which is not restricted to wedge-shaped overburdens) does not depend upon the applicability of the surface-impedance concept.[8]

ACKNOWLEDGMENTS

The author wishes to thank J. R. Wait for stimulating his interest in this investigation and for the many valuable discussions on this topic. Thanks are also due to F. G. Ullman for his comments and to Mrs. M. Alles for preparing the manuscript.

[1] J. R. Wait, *Electromagnetic Surface Waves*, Advances in Radio Research, J. A. Saxton, Ed. (Academic, London, 1964), Vol. 1, pp. 157–217.
[2] R. J. King, S. W. Maley, and J. R. Wait, *Electromagnetic Wave Theory*, J. Brown, Ed. (Pergamon, New York, 1965), pp. 217–24.
[3] J. R. Wait, J. Res. Natl. Bur. Std. (U.S.) **V65D**, 401 (1961).
[4] J. R. Wait and L. C. Walters, IEEE Trans. Antennas Propagation **11**, 38 (1963).
[5] R. J. King and C. H. Hustig, "Microwave (4.75 *GHz*) Surface Impedance Measurements of a 1° Dielectric Wedge on a Perfect Conductor" (presented at the conference Environmental Effects on Antenna Performance Proceedings of EEAP Conference, Vol. II, Clearinghouse for Federal Scientific and Technical Information, AD-692 879, Springfield, Virginia, 1969).
[6] G. A. Schlak and J. R. Wait, Can. J. Phys. **45**, 3697 (1967).
[7] G. A. Schlak and J. R. Wait, Can. J. Phys. **46**, 1135 (1968).
[8] E. Bahar, "Propagation of Radio Waves Over a Nonuniform Layered Medium" (to be published in Radio Sci.).
[9] R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960).
[10] H. M. Barlow and J. Brown, *Radio Surface Waves*, International Monograms on Radio (Clarendon, Oxford, 1962).
[11] E. Bahar, Trans. Microwave Theory Techniques **16**, 503 (1968).
[12] E. Bahar and J. R. Wait, Radio Sci., J. Res., **96D**, 1445 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Kruskal's Perturbation Method

DAVID P. STERN
Laboratory for Theoretical Studies, Goddard Space Flight Center, Greenbelt, Maryland 20771

(Received 29 December 1969)

A general method for eliminating the angle variable from the equations of a perturbed periodic motion and for deriving an "adiabatic invariant" J has been given by Kruskal and, for a special class of Hamiltonian systems, McNamara and Whiteman have shown (to order ϵ^2) that J is related to a set of invariants I obtained from the expansion of Poisson-bracket relations. In this work, an order-by-order algorithm for Kruskal's method is introduced, and a new set of invariants Z_1 is obtained. It is shown that these invariants bear a close relation to those obtained from the Poisson-bracket expansion and, in the special case investigated by McNamara and Whiteman, the relation between I and Z_1 may be brought to the same form as the relation between I and J derived by those authors. Finally, the relationship between Z_1 and J is examined, and arguments are presented that in certain cases the two are equal to all orders.

INTRODUCTION

Let a perturbed periodic mechanical system be given, described by n canonical variables collectively represented by the vector $\mathbf{y}$ and by a Hamiltonian H dependent on a small parameter ϵ, of the form

$$H = y_1 + \sum_{k=1}^{\infty} \epsilon^k H^{(k)}(\mathbf{y}) \tag{1}$$

(here and in what follows, superscripts in parentheses denote order in ϵ). The vector $\mathbf{y}$ may be viewed as the solution of the Hamilton–Jacobi equation for the unperturbed motion, yielding an action variable y_1, its conjugate angle variable y_n, and a set of other variables y_i which are constants of the unperturbed motion.

A perturbation may now be employed to eliminate y_1 and y_n from the equations of motion to any desired order in ϵ. One such technique has been devised by Kruskal[1] and will be described in more detail further on (this method is also applicable to noncanonical systems). By Kruskal's approach, a near-identity transformation from the variables $\mathbf{y}$ to new "nice" variables $\mathbf{z}$ is performed, so that, of the n first-order

Finally, we wish to recall that, when applicable, the geometrical-optical approach is a convenient method to determine the appropriate value of the surface impedance. However, the actual field variations still need to be computed. Through a judicious use of the compensation theorem, Wait[1] derives an integral equation for the field which is solved by an iterative method. Thus, the latter solution depends upon the applicability of the surface impedance concept. For the surface impedance concept to be meaningful, it must be insensitive to variations of the angle of incidence θ_0^i. Furthermore, the fluctuations in the relative value of the surface impedance must be limited.[6] On the other hand, the full-wave solution (which is not restricted to wedge-shaped overburdens) does not depend upon the applicability of the surface-impedance concept.[8]

ACKNOWLEDGMENTS

The author wishes to thank J. R. Wait for stimulating his interest in this investigation and for the many valuable discussions on this topic. Thanks are also due to F. G. Ullman for his comments and to Mrs. M. Alles for preparing the manuscript.

[1] J. R. Wait, *Electromagnetic Surface Waves*, Advances in Radio Research, J. A. Saxton, Ed. (Academic, London, 1964), Vol. 1, pp. 157–217.

[2] R. J. King, S. W. Maley, and J. R. Wait, *Electromagnetic Wave Theory*, J. Brown, Ed. (Pergamon, New York, 1965), pp. 217–24.

[3] J. R. Wait, J. Res. Natl. Bur. Std. (U.S.) **V65D**, 401 (1961).

[4] J. R. Wait and L. C. Walters, IEEE Trans. Antennas Propagation **11**, 38 (1963).

[5] R. J. King and C. H. Hustig, "Microwave (4.75 *GHz*) Surface Impedance Measurements of a 1° Dielectric Wedge on a Perfect Conductor" (presented at the conference Environmental Effects on Antenna Performance Proceedings of EEAP Conference, Vol. II, Clearinghouse for Federal Scientific and Technical Information, AD-692 879, Springfield, Virginia, 1969).

[6] G. A. Schlak and J. R. Wait, Can. J. Phys. **45**, 3697 (1967).

[7] G. A. Schlak and J. R. Wait, Can. J. Phys. **46**, 1135 (1968).

[8] E. Bahar, "Propagation of Radio Waves Over a Nonuniform Layered Medium" (to be published in Radio Sci.).

[9] R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960).

[10] H. M. Barlow and J. Brown, *Radio Surface Waves*, International Monograms on Radio (Clarendon, Oxford, 1962).

[11] E. Bahar, Trans. Microwave Theory Techniques **16**, 503 (1968).

[12] E. Bahar and J. R. Wait, Radio Sci., J. Res., **96D**, 1445 (1965).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Kruskal's Perturbation Method

DAVID P. STERN

Laboratory for Theoretical Studies, Goddard Space Flight Center, Greenbelt, Maryland 20771

(Received 29 December 1969)

A general method for eliminating the angle variable from the equations of a perturbed periodic motion and for deriving an "adiabatic invariant" J has been given by Kruskal and, for a special class of Hamiltonian systems, McNamara and Whiteman have shown (to order ϵ^2) that J is related to a set of invariants I obtained from the expansion of Poisson-bracket relations. In this work, an order-by-order algorithm for Kruskal's method is introduced, and a new set of invariants Z_1 is obtained. It is shown that these invariants bear a close relation to those obtained from the Poisson-bracket expansion and, in the special case investigated by McNamara and Whiteman, the relation between I and Z_1 may be brought to the same form as the relation between I and J derived by those authors. Finally, the relationship between Z_1 and J is examined, and arguments are presented that in certain cases the two are equal to all orders.

INTRODUCTION

Let a perturbed periodic mechanical system be given, described by n canonical variables collectively represented by the vector $\mathbf{y}$ and by a Hamiltonian H dependent on a small parameter ϵ, of the form

$$H = y_1 + \sum_{k=1}^{\infty} \epsilon^k H^{(k)}(\mathbf{y}) \tag{1}$$

(here and in what follows, superscripts in parentheses denote order in ϵ). The vector $\mathbf{y}$ may be viewed as the solution of the Hamilton–Jacobi equation for the unperturbed motion, yielding an action variable y_1, its conjugate angle variable y_n, and a set of other variables y_i which are constants of the unperturbed motion.

A perturbation may now be employed to eliminate y_1 and y_n from the equations of motion to any desired order in ϵ. One such technique has been devised by Kruskal[1] and will be described in more detail further on (this method is also applicable to noncanonical systems). By Kruskal's approach, a near-identity transformation from the variables $\mathbf{y}$ to new "nice" variables $\mathbf{z}$ is performed, so that, of the n first-order differential equations describing the evolution of $\mathbf{z}$, $n-1$ may be separated and solved independently as an autonomous set. Furthermore, using these "nice variables," we may express an "adiabatic invariant" J, which is a constant of the perturbed motion, to any order of ϵ. If J is used to eliminate z_1, one winds up as required with a mechanical system containing only $n-2$ independent variables.

An alternate method of deriving an invariant I for the system described by Eq. (1) is based on an expansion of the Poisson-bracket relation

$$[I, H] = 0. \tag{2}$$

This expansion has been described by Whittaker[2] and was further explored by McNamara and Whiteman[3] (henceforth referred to as McNW). The latter authors were able to show—by a lengthy calculation and only to order ϵ^2—that the invariant thus obtained is, in a special case, related to Kruskal's J.

In what follows we shall show that, for a system given by Eq. (1), Kruskal's method may be modified to yield a family of invariants in a completely different way from that used in deriving J. It will then be shown that these invariants are solutions of Eq. (2), are related to the invariant of McNW, and also are connected with the adiabatic invariant J.

Notation: In order to obtain concise expressions, the notation used here departs somewhat from that of Kruskal. In Kruskal's work, $\mathbf{y}$ and $\mathbf{z}$ have $n-1$ components and special symbols v and ϕ stand for what we here denote by y_n and z_n. In what follows, such vectors with $n-1$ components, excluding the angle variable, will be denoted by a tilde, e.g., $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$. Furthermore, if the canonical variables in $\mathbf{y}$ are arranged in order

$$\mathbf{y} = (\mathbf{p}, \mathbf{q}), \tag{3}$$

with the action-angle variables in the extreme positions, we shall define a conjugate vector

$$\bar{\mathbf{y}} = (\mathbf{q}, -\mathbf{p}) \tag{4}$$

so that $\bar{\mathbf{y}}_1 = y_n$ and $\bar{\mathbf{y}}_n = -y_1$.

The use of $\bar{\mathbf{y}}$ enables one to write Poisson brackets concisely [the summation convention is henceforth used in all summations over n or $(n-1)$ components] as

$$[a, b] = \frac{\partial a}{\partial \bar{y}_i}\frac{\partial b}{\partial y_i}, \tag{5}$$

and Hamilton's equations become

$$\frac{dy_i}{dt} = -\frac{\partial H}{\partial \bar{y}_i}. \tag{6}$$

Finally, we shall assume (as in Kruskal's work) that the basic period in the dependence on y_n is unity and denote quantities averaged over y_n (which clearly depend on $\tilde{\mathbf{y}}$ only) by angular brackets

$$\langle a \rangle = \int_0^1 a\, dy_n.$$

1. KRUSKAL'S EXPANSION

Kruskal's method does not require the system to be canonical but assumes the evolution of $\mathbf{y}$ to be given by equations of the form

$$\frac{d\mathbf{y}}{dt} = \sum_{k=0}^{\infty} \epsilon^k \mathbf{g}^{(k)}(\mathbf{y}), \tag{7}$$

where the components of $\mathbf{g}^{(k)}$ are periodic in y_n with period unity and where $\mathbf{g}^{(0)}$ has only one nonzero component, namely the nth. Since we are interested in systems for which (7) reduces to (6) with H given by (1), we shall assume that this component is unity

$$\mathbf{g}^{(0)} = (0, 0, \cdots, 0, 1), \tag{8}$$

although what follows can be extended to more general cases. We now seek a near-identity transformation to new variables

$$\mathbf{z} = \mathbf{y} + \sum_{k=1} \epsilon^k \boldsymbol{\zeta}^{(k)}(\mathbf{y}) \tag{9}$$

such that the evolution of $\mathbf{z}$ satisfies

$$\frac{d\mathbf{z}}{dt} = \sum_{k=0} \epsilon^k \mathbf{h}^{(k)}(\tilde{\mathbf{z}}) \tag{10}$$

with $h^{(0)} = g^{(0)}$. Since z_n does not appear on the right-hand side, the first $n-1$ equations of (10) form an autonomous system which may be solved independently, the solution then being substituted in the remaining equation to provide the evolution of z_n.

Substitution of (9) in the lhs of (10) gives

$$\begin{aligned}\frac{d\mathbf{z}}{dt} &= \frac{d\mathbf{y}}{dt} + \sum_{k=1} \epsilon^k \frac{\partial \boldsymbol{\zeta}^{(k)}}{\partial y_i}\frac{dy_i}{dt} \\ &= \sum_{k=0} \epsilon^k \left(\mathbf{g}^{(k)} + \sum_{j=0}^{k-1} g_i^{(j)} \frac{\partial \boldsymbol{\zeta}^{(k-j)}}{\partial y_i}\right) \\ &= \sum_{k=0} \epsilon^k \left(\mathbf{g}^{(k)} + \frac{\partial \boldsymbol{\zeta}^{(k)}}{\partial y_n} + \sum_{j=1}^{k-1} \mathbf{g}^{(j)} \cdot \nabla \boldsymbol{\zeta}^{(k-j)}\right)\end{aligned} \tag{11}$$

with the ∇ operator defined in $\mathbf{y}$ space. We now convert the rhs of (10) to depend on $\mathbf{y}$ as well, using the Taylor expansion operators derived as follows. Defining the exponential of a differential operator by means of the power series for e^x, one can formally express the Taylor expansion of $h^{(k)}(\tilde{\mathbf{z}})$ as follows

(* stands for "operates on"):

$$\begin{aligned}\mathbf{h}^{(k)}(\tilde{\mathbf{z}}) &= \mathbf{h}^{(k)}(\tilde{\mathbf{y}} + \sum \epsilon^j \boldsymbol{\zeta}^{(j)}) \\ &= \exp(\sum \epsilon^j \tilde{\boldsymbol{\zeta}}^{(j)} \cdot \nabla) * \mathbf{h}^{(k)}(\tilde{\mathbf{y}}) \\ &= \sum_{j=0} \epsilon^j T^{(j)} * \mathbf{h}^{(k)}(\tilde{\mathbf{y}}),\end{aligned} \tag{12}$$

where the $T^{(j)}$ are differential operators obtained by expanding the exponential. The first few of them are[4]

$$\begin{aligned}T^{(0)} &= 1, \\ T^{(1)} &= \boldsymbol{\zeta}^{(1)} \cdot \nabla, \\ T^{(2)} &= \boldsymbol{\zeta}^{(2)} \cdot \nabla + \tfrac{1}{2}\boldsymbol{\zeta}^{(1)}\boldsymbol{\zeta}^{(1)} : \nabla\nabla, \\ T^{(3)} &= \boldsymbol{\zeta}^{(3)} \cdot \nabla + \boldsymbol{\zeta}^{(1)}\boldsymbol{\zeta}^{(2)} : \nabla\nabla + \tfrac{1}{6}\boldsymbol{\zeta}^{(1)}\boldsymbol{\zeta}^{(1)}\boldsymbol{\zeta}^{(1)} \therefore \nabla\nabla\nabla.\end{aligned} \tag{13}$$

Because $T^{(0)}$ equals unity, one can separate the $\mathbf{h}^{(k)}$ term from the rest. The rhs of (10) then becomes

$$\sum \epsilon^k \left(\mathbf{h}^{(k)}(\tilde{\mathbf{y}}) + \sum_{j=1}^{k-1} T^{(j)} * \mathbf{h}^{(k-j)}(\tilde{\mathbf{y}}) \right). \tag{14}$$

Both sides of (10) are now functions of y, and therefore the equality holds independently for every order. One thus obtains a set of equations

$$\frac{\partial \boldsymbol{\zeta}^{(k)}}{\partial y_n} - \mathbf{h}^{(k)}(\tilde{\mathbf{y}}) = \boldsymbol{\lambda}^{(k)}(\mathbf{y}), \tag{15}$$

where

$$\boldsymbol{\lambda}^{(k)}(\mathbf{y}) = \sum_{j=1}^{k-1} T^{(j)} * \mathbf{h}^{(k-j)} - \mathbf{g}^{(k)} - \sum_{j=1}^{k-1} \mathbf{g}^{(j)} \cdot \nabla \boldsymbol{\zeta}^{(k-j)} \tag{16}$$

and

$$k = 1, 2, \cdots.$$

This may be used as a recursion relation. Suppose all quantities entering here are (like the $\mathbf{g}^{(k)}$) either periodic in y_n or independent of it. Then $\boldsymbol{\lambda}^{(k)}$ will possess a "secular" part $\langle \boldsymbol{\lambda}^{(k)} \rangle$ independent of y_n and a purely periodic part averaging zero,

$$(\boldsymbol{\lambda}^{(k)})_{\text{per}} = \boldsymbol{\lambda}^{(k)} - \langle \boldsymbol{\lambda}^{(k)} \rangle.$$

On the lhs of (15), $\mathbf{h}^{(k)}$ is wholly secular by definition, while the other term there is purely periodic, since any secular part of $\boldsymbol{\zeta}^{(k)}$ is removed by the differentiation. Thus, if $\boldsymbol{\lambda}^{(k)}$ is given, we can derive the kth-order quantities through

$$\mathbf{h}^{(k)}(\tilde{\mathbf{y}}) = -\langle \boldsymbol{\lambda}^{(k)} \rangle, \tag{17}$$

$$\frac{\partial \boldsymbol{\zeta}^{(k)}}{\partial y_n} = \boldsymbol{\lambda}^{(k)} - \langle \boldsymbol{\lambda}^{(k)} \rangle, \tag{18}$$

from which

$$\boldsymbol{\zeta}^{(k)} = \int_0^{y_n} (\boldsymbol{\lambda}^{(k)} - \langle \boldsymbol{\lambda}^{(k)} \rangle)\, dy_n' + \boldsymbol{\mu}^{(k)}(\tilde{\mathbf{y}}), \tag{19}$$

with $\boldsymbol{\mu}^{(k)}$ an arbitrary secular vector. In Kruskal's work, $\boldsymbol{\mu}^{(k)}$ is chosen to vanish, so that at $y_n = 0$ the vector $\mathbf{z}$ is identical with $\mathbf{y}$; we shall denote the functions thus obtained by $\hat{\boldsymbol{\zeta}}^{(k)}$. If, however, we only demand that $\mathbf{z}$ should satisfy equations of the form (10), $\boldsymbol{\mu}^{(k)}$ may be arbitrarily chosen. In what follows we shall make use of this free choice in order to endow the "nice variables" $\mathbf{z}$ with additional desired properties.

2. THE NEW INVARIANT

So far we have treated the general case of Kruskal's expansion (apart from our choice of $g_n^{(0)}$) with no reference to the canonical character of y. Taking this now into account, one finds from (1) and (6)

$$g_i^{(k)} = -\frac{\partial H^{(k)}}{\partial \tilde{y}_i}. \tag{20}$$

Defining for convenience

$$\boldsymbol{\zeta}^{(0)} \equiv \mathbf{y}$$

and using (16) and (5) give

$$\boldsymbol{\lambda}^{(k)} = \sum_{j=1}^{k-1} (T^{(j)} * \mathbf{h}^{(k-j)}(\tilde{\mathbf{y}})) + \sum_{j=1}^{k} [H^{(j)}, \boldsymbol{\zeta}^{(k-j)}]. \tag{21}$$

We now pose the following question: Is it possible, by proper choice of the $\boldsymbol{\mu}^{(k)}$ in (19), to make some component z_i of $\mathbf{z}$ a constant Z_i of the motion?

If z_i is conserved, this means that $h_i^{(k)}$ vanishes for all k, which in turn implies the vanishing of $\langle \lambda_i^{(k)} \rangle$. By the last equation, this reduces to

$$\sum_{j=1}^{k} \langle [H^{(j)}, \zeta_i^{(k-j)}] \rangle = 0. \tag{22}$$

Suppose that at the stage when Eq. (22) is reached the $\boldsymbol{\mu}^{(i)}$ have been derived up to and including the $(k-2)$ order. We can then fulfil (22) by choosing $\mu_i^{(k-1)}$ to satisfy

$$[\mu_i^{(k-1)}, \langle H^{(1)} \rangle] = \langle [H^{(1)}, \hat{\zeta}_i^{(k-1)}] \rangle + \sum_{j=2}^{k} \langle [H^{(j)}, \zeta_i^{(k-j)}] \rangle, \tag{23}$$

where the rhs is assumed to be known at that stage. The above equation is a linear first-order partial differential equation, and solutions in general do exist. Deriving them explicitly is another problem, however. For the special case when all $H^{(k)}$ with $k > 1$ vanish, McNamara and Whiteman (who arrive at a similar equation) obtained from first principles formulas which allow $\mu_i^{(k)}$ to be derived (for $i = 1$, which is the relevant case, as will be seen) up to $k = 3$. A different approach to the problem will be outlined in Sec. 5.

The iteration for Z_i can thus be carried on—provided that it can be started. The first additive function encountered is $\mu_i^{(1)}$, used in ensuring the vanishing of $\langle \lambda_i^{(2)} \rangle$. There exists no adjustable variable to ensure the vanishing of $\langle\lambda_i^{(1)}\rangle$, so the iteration can be started if and only if this term vanishes of its own accord, which in turn implies that

$$\left\langle\frac{\partial H^{(1)}}{\partial \tilde{y}_i}\right\rangle = 0. \tag{24}$$

If (as assumed) y_n enters H only through angular terms, this will certainly hold for $i = 1$, since the y_n derivative which is applied in that case removes the secular part of $H^{(1)}$, leaving a purely periodic function. Thus an invariant Z_1 of the type discussed here may, in general, be derived. If $H^{(1)}$ itself is purely periodic, other invariants may be generated for $i \neq 1$.

3. THE POISSON-BRACKET METHOD

McNamara and Whiteman,[3] following Whittaker,[2] derive an invariant I (in their notation: J) in the following way. Let I have an expansion in ϵ

$$I = \sum_{k=0} \epsilon^k I^{(k)}(y), \tag{25}$$

and let H be expanded as in (1) (this is a slight generalization; in the work of McNW, all $H^{(i)}$ with $i > 1$ vanish[5]). Then

$$[I, H] = 0 = \sum_{k=0} \epsilon^k \sum_{j=0}^{k} [I^{(j)}, H^{(k-j)}]. \tag{26}$$

By (5)

$$[I^{(k)}, H^{(0)}] = [I^{(k)}, y_1] = \frac{\partial I^{(k)}}{\partial y_n}. \tag{27}$$

So the equation for the kth order is

$$\frac{\partial I^{(k)}}{\partial y_n} = -\sum_{j=0}^{k-1} [I^{(j)}, H^{(k-j)}], \tag{28}$$

which defines $I^{(k)}$ recursively within an arbitrary function of $\tilde{\mathbf{y}}$:

$$\begin{aligned} I^{(k)} &= -\int_0^{y_n} \sum_{j=0}^{k-1} [I^{(j)}, H^{(k-j)}]\, dy_n' + G^{(k)}(\tilde{\mathbf{y}}) \\ &= \hat{I}^{(k)} + G^{(k)}. \end{aligned} \tag{29}$$

By (19) and (21), this is exactly the same as the equation for $\zeta_i^{(k)}$, provided that all $h_i^{(k)}$ vanish.

Consider again Eq. (28): due to the y_n derivative, its lhs will be purely periodic, but unless special steps are taken the rhs may well contain a secular part. We therefore must assume that, at the stage at which $I^{(k)}$ is being derived, $G^{(k-1)}$ has not yet been determined. The rhs can then be made purely periodic by requiring that

$$[G^{(k-1)}, \langle H^{(1)}\rangle] = -\langle[\hat{I}^{(k-1)}, H^{(1)}]\rangle - \sum_{j=0}^{k-2} \langle[I^{(j)}, H^{(k-j)}]\rangle, \tag{30}$$

which has the same form as (23).

The lowest order of (26) gives

$$[I^{(0)}, H^{(0)}] = 0 = \frac{\partial I^{(0)}}{\partial y_n}$$

so that

$$I^{(0)} = I^{(0)}(\tilde{\mathbf{y}}).$$

The iteration can be started only if $I^{(0)}$ makes the rhs of (28) purely periodic for $k = 1$:

$$[I^{(0)}, \langle H^{(1)}\rangle] = 0. \tag{31}$$

Obviously, any function of $H^{(1)}$ can be chosen as $I^{(0)}$; McNW chose

$$I^{(0)} = \langle H^{(1)}\rangle. \tag{32}$$

This, however, is not entirely satisfactory, since we expect $I^{(0)}$ to tend to some natural invariant of the unperturbed system in the limit of vanishing ϵ, independent of the choice of $H^{(1)}$. A more suitable choice is

$$I^{(0)} = y_1, \tag{33}$$

which also satisfies (31), since (24) holds for $i = 1$. With this choice, I equals the action variable y_1 in the unperturbed limit, a property shared by the invariant Z_1 previously derived and also (it may be shown) by Kruskal's J. The alternative choice, made by McNW, will be explored in the next section.

With $I^{(0)}$ chosen as in (33), it also equals $\zeta_1^{(0)}$,and it is easy to show that the expansion equations of I match those of the invariant Z_1 stage by stage. One may then match

$$I^{(k)} = \zeta_1^{(k)},$$
$$G^{(k)} = \mu_1^{(k)}$$

by making identical choices of the arbitrary functions of $\langle H^{(1)}\rangle$ [and by virtue of (33) satisfying (31), functions of y_1 as well] which can be added to $G^{(k)}$ and to $\mu_1^{(k)}$ at every stage.

4. THE CHOICE $I^{(0)} = \langle H^{(1)}\rangle$

McNamara and Whiteman chose $I^{(0)}$ as in (32), in the special case where $H^{(k)}$ vanishes for $k > 1$. In that case, their recursion continues with

$$\begin{aligned} I^{(1)} &= -\int_0^{y_n} [\langle H^{(1)}\rangle, H^{(1)}]\, dy_n' + G^{(1)} \\ &= \hat{I}^{(1)} + G^{(1)} \end{aligned} \tag{34}$$

and

$$[G^{(1)}, \langle H^{(1)}\rangle] = \langle[H^{(1)}, \hat{I}^{(1)}]\rangle. \tag{35}$$

Expanding the invariant Z_1 by Kruskal's method, for the same H, we obtain

$$\lambda_1^{(1)} = \frac{\partial H^{(1)}}{\partial y_n}$$

and, by (19),

$$\zeta_1^{(1)} = H^{(1)} - H^{(1)}(y_n = 0) + \mu_1^{(1)}.$$

From (23) and the preceding equation, noting that secular functions may be taken out of the averaging brackets, we obtain

$$[\mu_1^{(1)}, \langle H^{(1)}\rangle] = -[\langle H^{(1)}\rangle, H^{(1)}(y_n = 0)].$$

Let us select

$$\mu_1^{(1)} = -\langle H^{(1)}\rangle + H^{(1)}(y_n = 0)$$

so that

$$\zeta_1^{(1)} = H^{(1)} - \langle H^{(1)}\rangle.$$

The second-order equations then are, by (21),

$$\zeta_1^{(2)} = -\int_0^{y_n} [H^{(1)}, \langle H^{(1)}\rangle]\, dy_n' + \mu_1^{(2)}$$

and

$$[\mu_1^{(2)}, \langle H^{(1)}\rangle] = \langle [H^{(1)}, \hat{\zeta}_1^{(2)}]\rangle.$$

This demonstrates that, with matching choices of additive functions,

$$I^{(1)} = -\zeta_1^{(2)}, \quad G^{(1)} = -\mu_1^{(2)}.$$

If additive functions of higher orders are matched as well, the subsequent parts of $-\zeta_1$ and of I are identical, except for an extra order of ϵ in the former. That means that

$$-\epsilon(I - I^{(0)}) = Z_1 - \zeta_1^{(0)} - \epsilon\zeta_1^{(1)}$$

or

$$Z_1 = H - \epsilon I, \tag{36}$$

a relation resembling that obtained (to order ϵ^2) by McNW, except that in their result J appears in place of Z_1.

5. RELATION TO J

Kruskal defined the adiabatic invariant J by

$$J = \oint \sum_k p_k\, dq_k = \int_0^1 \sum_k p_k \frac{\partial q_k}{\partial z_n}\, dz_n, \tag{37}$$

where the integration is carried over a set of points ("ring") sharing the same $\mathbf{z}$ and differing only in z_n. For details about J, the reader is referred to Kruskal's article[1]; its value is independent of the canonical set used in its derivation, though the components of $\mathbf{y}$ are the best choice for this role if they form a canonical set [the inverse transformation $\mathbf{z} \to \mathbf{y}$ must then be derived by means of the expansion operators (13)]. Here we shall merely sketch out the connection between Z_1 and J without deriving the details.

Suppose that among the many sets of "nice" variables possible, differing in their choices of $\mu_i^{(k)}$ but all obeying (10), there exists a set (or a family of sets) that is canonical, with z_1 conjugate to z_n. Obviously, this set, too, can be used in deriving (37), leading immediately to

$$J = z_1.$$

The existence of nice canonical variables has been proved by Kruskal[1] in Appendix 2 of his article. It is furthermore possible to express the $\mu_i^{(k)}$ which generate such variables. The details of this derivation are somewhat involved and will therefore be described in a separate article; here we shall just assume that these $\mu_i^{(k)}$ are known. Then the transformation which they describe belongs to the (much larger) family of transformations which make z_1 a constant of the motion, each of which in its turn provides a solution to the corresponding Poisson-bracket expansion. Thus, among the many possible solutions of Z_1 and I, there exist such ones for which

$$I = Z_1 = J.$$

Practically, given the $\mu_i^{(k)}$ which make $\mathbf{z}$ canonical, these functions probably offer the best way of deriving Z_1 or I, since they are known to solve Eq. (30). The only other possibility for solving this equation is to use the formulas of McNW, which are valid for the lowest few orders only and are specifically related to the choice of $I^{(0)}$ given in (32).

[1] M. Kruskal, J. Math. Phys. **3**, 806 (1962).

[2] E. T. Whittaker, *Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944), 4th ed.

[3] B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).

[4] These operators originated in the calculation of derivatives of a function of a function by F. De Bruno, Quart. J. Math. **1**, 359 (1855). They are listed up to $j = 8$ by P. Musen, J. Astronautical Sci. **12**, 129 (1965).

[5] The definition of Poisson brackets by McNamara and Whiteman, in Eq. (3.2) of their article, has a reversed sign compared to the usual convention. As a consequence, some equations of theirs differ in sign from those obtained here.

Direct Canonical Transformations

DAVID P. STERN
Laboratory for Theoretical Studies, Goddard Space Flight Center
Greenbelt, Maryland 20771

(Received 29 December 1969)

Some of the perturbation methods in classical Hamiltonian mechanics lead to near-identity transformations of the variables, with the new variables explicitly given as functions of the old ones. Two methods are used for identifying and characterizing the subclass of all such transformations which are also canonical: one approach is related to the conventional method of generating canonical transformations, while the other one uses the properties of Poisson brackets and is related to an operator method of Lie. Either of the methods may be used to derive certain steps in a perturbation method devised by Lacina, inadvertently omitted by that author.

INTRODUCTION

Let a classical canonical system be given, described by $2n$ canonical variables collectively denoted by the vector $\mathbf{y}$. The components y_i may be divided into canonical momenta p_i and canonical coordinates q_i, and we shall assume that their order is

$$\mathbf{y} = (p_1 \cdots p_n, q_1 \cdots q_n) = (\mathbf{p}, \mathbf{q}). \quad (1)$$

A transformation $\mathbf{y} \to \mathbf{z}$ is termed canonical if the new variables also form a canonical set

$$\mathbf{z} = (\mathbf{P}, \mathbf{Q}). \quad (2)$$

Canonical transformations are customarily defined by means of a generating function σ[1,2] depending on n new and n old variables (and possibly on time, though this will not be assumed in what follows). Of the several possible types of such functions, we choose to use those of the form

$$\sigma = \sigma(\mathbf{P}, \mathbf{q}), \quad (3)$$

for which the transformation equations are given by

$$\frac{\partial\sigma}{\partial P_i} = Q_i, \quad \frac{\partial\sigma}{\partial q_i} = p_i. \quad (4)$$

These equations give the transformations indirectly: If the old variables are given and the new ones are sought, additional "untangling" is generally in order. By contrast, a "direct" transformation

$$\mathbf{z} = \mathbf{z}(\mathbf{y})$$

is immediately usable. In particular, we shall be concerned with near-identity transformations expanded in terms of a small parameter $\epsilon \ll 1$ (the superscript in parentheses denotes order in ϵ):

$$\mathbf{z} = \mathbf{y} + \sum_{k=1} \epsilon^k \boldsymbol{\zeta}^{(k)}(\mathbf{y}). \quad (5)$$

In what follows, we shall investigate the conditions under which the transformation (5) is canonical.

The motivation for this investigation may be of some interest. In the classical perturbation theory developed for celestial mechanics, the basic problem is the solution of a canonical system depending on a small parameter ϵ, given that in the limit of vanishing ϵ ("unperturbed motion") the solution is known and periodic. One procedure used there (usually associated with the names of Von Zeipel and Poincaré) is as follows.[3] First, one prepares the ground by transforming the original problem to new variables—which will be labeled $\mathbf{y}$ as in Eq. (1)—so that one of the new variables is anglelike and represents the periodicity of the unperturbed motion, while the others, in the limit of vanishing ϵ, are constant in time. If the angle variable is conjugate to the variable y_1, the new variables may be chosen so as to make the Hamiltonian $H = y_1$ as well.

The same variables are then introduced into the finite-ϵ problem, in which case the Hamiltonian assumes the form

$$H = y_1 + \sum_{k=1} \epsilon^k H^{(k)}(\mathbf{y}). \quad (6)$$

We now seek a near-identity canonical transformation, generated by

$$\sigma(\mathbf{P}, \mathbf{q}) = \sum_{i=1}^{n} P_i q_i + \sum_{k=1}^{\infty} \epsilon^k \sigma^{(k)}(\mathbf{P}, \mathbf{q}) \quad (7)$$

such that, in the new variables $\mathbf{z}$, the angle variable is absent from the transformed Hamiltonian, making its conjugate z_1 a constant of the motion. Further solution of the motion involves neither of these variables as an unknown, so that the problem has been reduced by two variables or, equivalently, by one dimension.

A different approach, independently developed by Kruskal,[4] is related to the method of Bogoliubov and Krylov and describes the evolution of a system through $2n$ equations of the form

$$\frac{d\mathbf{x}}{dt} = \sum_{k=0} \epsilon^k \mathbf{f}^{(k)}(\mathbf{x}). \tag{8}$$

The method does not require the system to be canonical, but we shall restrict ourselves here to problems where this is the case. The preparatory derivation of "intermediate variables" $\mathbf{y}$ is then very much the same as before. To solve the perturbed problem, however, the method seeks a direct transformation of the form (5), with the angle variable transformed into one of the components of $\mathbf{z}$ which will be denoted (for reasons which will be clarified later) as $\bar{z}_1$. The transformation is such that, when Eqs. (8) are transformed to give the evolution of $\mathbf{z}$,

$$\frac{d\mathbf{z}}{dt} = \sum_{k=0} \epsilon^k \mathbf{h}^{(k)}(\mathbf{z}), \tag{9}$$

then $\bar{z}_1$ is absent on the right-hand side. One can now separately handle the $2n-1$ equations in $2n-1$ variables obtained by omitting the equation for $d\bar{z}_1/dt$ from the above set. One can also obtain a constant of the motion, the so-called adiabatic invariant J, thus removing one more unknown variable from the problem. The details of J will not be described here beyond noting that it resembles in many ways the constant action variable z_1 obtained by the preceding method—for instance, the two may be shown to be equal at least to order ϵ^0.

The two methods are evidently similar, and one may ask whether they can be made to coincide by a suitable choice of the arbitrariness existing in Kruskal's method.[5] For this to happen, the transformation (5) obtained by Kruskal's approach must be canonical, which leads one to the basic problem stated before. As will be shown, the matter is also intimately connected to the work of Lacina,[6,7] where it leads to the correction of an error.

Two approaches to the problem will be described here: The first one is concise and is derived from the conventional form of canonical transformations, while the second one appears to be more elegant and has interesting geometrical implications. It is related to an operator method for characterizing direct canonical transformations, originally devised by Lie[8] and recently applied to celestial mechanics by Hori[9] and Deprit.[10]

1. DIRECT SOLUTION

If the transformation (5) is canonical, then there must exist a generating function $\sigma(\mathbf{P}, \mathbf{q})$ of the form given in (7) that leads to it. Applying (3), one obtains

$$Q_i = q_i + \sum_{k=1} \epsilon^k \left(\frac{\partial \sigma^{(k)}(\mathbf{P}, \mathbf{q})}{\partial P_i}\right), \tag{10}$$

$$P_i = p_i - \sum_{k=1} \epsilon^k \left(\frac{\partial \sigma^{(k)}(\mathbf{P}, \mathbf{q})}{\partial q_i}\right). \tag{11}$$

If the ordering of the components of $\boldsymbol{\zeta}^{(k)}$ corresponds to the ordering of the components of $\mathbf{y}$ in (1), one may define "partial vectors," the sum of which equals $\boldsymbol{\zeta}^{(k)}$,

$$\boldsymbol{\pi}^{(k)} = (\zeta_1^{(k)} \cdots \zeta_n^{(k)}, 0 \cdots 0),$$
$$\boldsymbol{\theta}^{(k)} = (0 \cdots 0, \zeta_{n+1}^{(k)} \cdots \zeta_{2n}^{(k)}), \tag{12}$$

which allows (4) to be split up into

$$Q_i = q_i + \sum_{k=1} \epsilon^k \theta_{i+n}^{(k)}(\mathbf{y}), \tag{13}$$

$$P_i = p_i + \sum_{k=1} \epsilon^k \pi_i^{(k)}(\mathbf{y}), \quad i = 1, \cdots, n. \tag{14}$$

Substituting in (11), we obtain

$$P_i = p_i - \sum_{k=1} \epsilon^k \frac{\partial \sigma^{(k)}}{\partial q_i}\left(\mathbf{p} + \sum_{m=1} \epsilon^m \boldsymbol{\pi}^{(m)}(\mathbf{y}), \mathbf{q}\right). \tag{15}$$

We now introduce expansion operators[11,5] $S^{(k)}$ such that, for any function f of the canonical variables ($*$ means "operates on"),

$$f(\mathbf{P}, \mathbf{q}) = \sum_{m=0} \epsilon^m S^{(m)} * f(\mathbf{p}, \mathbf{q}). \tag{16}$$

The explicit expressions for $S^{(m)}$ may be obtained by replacing $\boldsymbol{\zeta}^{(k)}$ with $\boldsymbol{\pi}^{(k)}$ in Eqs. (13) of Ref. 5; they involve $\boldsymbol{\nabla}$ operators in $\mathbf{y}$ space acting on the $\boldsymbol{\pi}^{(k)}$. We then get ($S^{(0)}$ being unity)

$$P_i = p_i - \sum_{k=1} \epsilon^k \left(\frac{\partial \sigma^{(k)}}{\partial q_i} + \sum_{m=1}^{k-1} S^{(m)} * \frac{\partial \sigma^{(k-m)}}{\partial q_i}\right), \tag{17}$$

where the various orders of σ on the rhs are functions of $\mathbf{y}$ only and may be obtained from those appearing in Eq. (7) by replacing $\mathbf{P}$ everywhere with $\mathbf{p}$. This is formally equivalent to (14), and therefore

$$\pi_i^{(k)} = -\frac{\partial \sigma^{(k)}(\mathbf{y})}{\partial q_i} - \sum_{m=1}^{k-1} S^{(m)} * \frac{\partial \sigma^{(k-m)}}{\partial q_i}. \tag{18}$$

Since the $S^{(m)}$ appearing here all involve lower orders of $\boldsymbol{\pi}^{(m)}$, this is a usable recursion formula, allowing $\boldsymbol{\pi}^{(k)}$ for the direct transformation to be solved—provided that the $\sigma^{(k)}$ are known and provided that the lower orders have already been solved. Similarly, one gets

$$\theta_{i+n}^{(k)} = \frac{\partial \sigma^{(k)}(\mathbf{y})}{\partial p_i} + \sum_{m=1}^{k-1} S^{(m)} * \frac{\partial \sigma^{(k-m)}}{\partial p_i}. \tag{19}$$

The last two equations can be joined together by introducing the concept of the conjugate vector $\bar{\mathbf{y}}$, formed by permuting the order of components of $\mathbf{y}$ given in Eq. (1) to

$$\bar{\mathbf{y}} = (\mathbf{q}, -\mathbf{p}). \tag{20}$$

Several properties of $\bar{\mathbf{y}}$ are described in Ref. 5; using them, one may express Poisson brackets as ($\bar{\boldsymbol{\nabla}}$ operator defined in $\bar{\mathbf{y}}$ space)

$$[a, b] = \bar{\boldsymbol{\nabla}} a \cdot \boldsymbol{\nabla} b \tag{21}$$

and, in particular,

$$[a, y_i] = \frac{\partial a}{\partial \bar{y}_i}. \tag{22}$$

With this notation, (18) and (19) may be combined to

$$\boldsymbol{\zeta}^{(k)} = -\bar{\boldsymbol{\nabla}} \sigma^{(k)} + \boldsymbol{\rho}^{(k)}, \tag{23}$$

with

$$\rho_i^{(k)} = -\sum_{m=1}^{k-1} S^{(m)} * \frac{\partial \sigma^{(k-m)}}{\partial \bar{y}_i} \tag{24}$$

being determined only by lower orders. In spite of its external appearance, this relation is not free from partial vectors, since the $\boldsymbol{\pi}^{(m)}$ appear in the S operators. A criterion for canonicity is now easily established. If the transformation (5) is canonical up to and including the order $k - 1$, one may form the vector

$$\mathbf{u}^{(k)} = \boldsymbol{\zeta}^{(k)} - \boldsymbol{\rho}^{(k)},$$

and the transformation will be canonical if and only if $\mathbf{u}^{(k)}$ is a gradient in $\bar{\mathbf{y}}$ space, that is, if the components of the "conjugate curl" tensor of $\mathbf{u}^{(k)}$,

$$(\bar{\boldsymbol{\nabla}} \times \mathbf{u}^{(k)})_{ij} = \frac{\partial u_i^{(k)}}{\partial \bar{y}_j} - \frac{\partial u_j^{(k)}}{\partial \bar{y}_i}, \tag{25}$$

all vanish.

2. SOLUTION BY POISSON BRACKETS

A different approach involves no partial vectors and is based on the properties of Poisson brackets. If both $\mathbf{y}$ and $\mathbf{z}$ are canonical and equal in the zeroth order, then

$$[y_i, y_j] = [z_i, z_j], \tag{26}$$

the value of the brackets being 1 or 0 depending on whether canonical conjugacy exists or not. Substituting the expansion (5) on the rhs and equating each order in ϵ separately, one gets as the condition for canonicity a set of relations, which through (22) and (25) can be expressed as

$$(\bar{\boldsymbol{\nabla}} \times \boldsymbol{\zeta}^{(k)})_{ij} = -\sum_{m=1}^{k-1} [\zeta_i^{(m)}, \zeta_j^{(k-m)}]. \tag{27}$$

Of more interest than a criterion for canonicity would be the derivation of a method for actually generating direct canonical transformations. Equations (27) are not very convenient for this, since they involve the components of $\boldsymbol{\zeta}^{(k)}$ only through combinations of their derivatives. A method for integrating these equations would obviously be useful here. For instance, the first equation of the set,

$$\bar{\boldsymbol{\nabla}} \times \boldsymbol{\zeta}^{(1)} = 0,$$

may be integrated to

$$\boldsymbol{\zeta}^{(1)} = \bar{\boldsymbol{\nabla}} \chi^{(1)}, \tag{28}$$

with $\chi^{(1)}$ an arbitrary scalar. Similarly, an arbitrary conjugate gradient may be added to any of the $\boldsymbol{\zeta}^{(k)}$, since such a gradient is ignored by the curl operation. We may thus formally write the most general solution of (27) as

$$\boldsymbol{\zeta}^{(k)} = \bar{\boldsymbol{\nabla}} \chi^{(k)} + \mathbf{f}^{(k)}, \tag{29}$$

where $\mathbf{f}^{(k)}$ is any vector depending only on lower orders of $\boldsymbol{\zeta}^{(m)}$, which gives one particular solution of that equation. Comparing (29) with (25) shows that one possible choice of $\mathbf{f}^{(k)}$ is

$$\mathbf{f}^{(k)} = \boldsymbol{\rho}^{(k)}, \tag{30}$$

which carries with it the identification

$$\chi^{(k)} = -\sigma^{(k)}.$$

The drawback of this choice (more esthetical than practical) is that $\mathbf{f}^{(k)}$ contains partial vectors. Solutions which do not split up phase space into coordinates and momenta also do exist to any order. The first of these is almost trivial: by (28), $\mathbf{f}^{(1)}$ vanishes. The next three vectors are

$$\begin{aligned}
\mathbf{f}^{(2)} &= \tfrac{1}{2}\boldsymbol{\zeta}^{(1)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)}, \\
\mathbf{f}^{(3)} &= \boldsymbol{\zeta}^{(2)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)}, \\
\mathbf{f}^{(4)} &= \boldsymbol{\zeta}^{(3)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)} + \tfrac{1}{2}\boldsymbol{\zeta}^{(2)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(2)} \\
&\quad + \tfrac{1}{4}[(\boldsymbol{\zeta}^{(1)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)}) \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(2)} - \boldsymbol{\zeta}^{(2)} \cdot \boldsymbol{\nabla}(\boldsymbol{\zeta}^{(1)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)})] \\
&\quad - \tfrac{1}{4}[(\boldsymbol{\zeta}^{(1)} \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)}) \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)}] \cdot \boldsymbol{\nabla}\boldsymbol{\zeta}^{(1)}.
\end{aligned} \tag{31}$$

They are far from unique, since various "curl-free" expressions, involving only the $\boldsymbol{\zeta}^{(m)}$, may be added to any one of them. A general method for deriving such expressions, based on Lie's method for characterizing direct canonical transformations, will now be described.

3. LIE'S METHOD

Let $W(\mathbf{y})$ be an arbitrary function of a given set of canonical variables. One then defines[10] the *Lie derivative generated by W* of any function $f(\mathbf{y})$ as the function

$$L_W(f) = [f, W].$$

The operator $L_{\Pi'}$ is linear

$$L_{\Pi'}(\alpha f + \beta g) = \alpha L_{\Pi'}(f) + \beta L_{\Pi'}(g),$$

and its action on a product resembles that of the derivative

$$L_{\Pi'}(fg) = fL_{\Pi'}(g) + gL_{\Pi'}(f).$$

By means of Jacobi's identity, one can prove similar properties for $L_{\Pi'}$ acting on Poisson brackets, except that the order of terms must now be preserved:

$$L_{\Pi'}[f, g] = [L_{\Pi'}(f), g] + [f, L_{\Pi'}(g)].$$

By successive application of $L_{\Pi'}$, various powers $L^n_{\Pi'}$ of the operator may be defined, and for completeness one then includes $L^0_{\Pi'}$ as the identity operator. If α is a constant, $L^n_{\Pi'}(\alpha)$ thus vanishes for all values of n except zero, since all powers of $L_{\Pi'}$ except the zeroth involve differentiation.

Using the preceding definitions, one may define an exponential operator

$$\exp(\epsilon L_{\Pi'}) * f = (1 + \epsilon L_{\Pi'} + \tfrac{1}{2}\epsilon^2 L^2_{\Pi'} + \cdots) * f, \quad (32)$$

where ϵ is a constant much smaller than unity, which helps the expression converge (convergence will not be discussed here, however). The property of the exponential operator of importance here is[10]

$$\exp(\epsilon L_{\Pi'}) * [f, g] = [\exp(\epsilon L_{\Pi'}) * f, \exp(\epsilon L_{\Pi'}) * g].$$

The preceding transformation shows that, if $\mathbf{y}$ is canonical, then for any W

$$\mathbf{z} = \exp \epsilon L_{\Pi'} * \mathbf{y} \quad (33)$$

is also canonical, for we then have

$$[z_i, z_j] = \exp \epsilon L_{\Pi'} * [y_i, y_j] = [y_i, y_j],$$

the latter Poisson brackets always being a constant equal to 1 or 0. By (32), the transformation is a near-identity one, even if the "generating function" W does not depend on ϵ. It is, nevertheless, possible to include such a dependence, expressed in a power series in ϵ:

$$W(\mathbf{y}, \epsilon) = \sum_{k=0} \epsilon^k W^{(k)}(\mathbf{y}). \quad (34)$$

We now prove the following. Let a direct canonical transformation be given by Lie's method as in (33), with W expanded as in (34). Then the same transformation may also be expressed as in (4), with $\boldsymbol{\zeta}^{(k)}$ given as in (29). If we choose to identify

$$\chi^{(k)} = -W^{(k-1)}, \quad (35)$$

then the two approaches may be readily related and an explicit expression for $\mathbf{f}^{(k)}$ may be found.

We begin with the following result.

Lemma: Suppose that (35) holds, that $i < k$, that the "function" $g(\boldsymbol{\zeta})$ is an expression involving orders of $\boldsymbol{\zeta}^{(m)}$ lower than the kth, and that $f^{(m)}(\boldsymbol{\zeta})$ is known for values of m smaller than k; then a "function" $h(\boldsymbol{\zeta})$, similar in structure to g and $f^{(m)}$, may always be found so that

$$[g(\boldsymbol{\zeta}), W^{i-1}] = h(\boldsymbol{\zeta}).$$

Proof: Using (21), (29), and (35), we obtain

$$\begin{aligned}[g, W^{i-1}] &= [\chi^{(i)}, g] \\ &= \overline{\nabla}\chi^{(i)} \cdot \nabla g(\boldsymbol{\zeta}) \\ &= (\boldsymbol{\zeta}^{(i)} - \mathbf{f}^{(i)}(\boldsymbol{\zeta})) \cdot \nabla g(\boldsymbol{\zeta}),\end{aligned}$$

and the last expression has the required form.

Several corollaries are now easily derived:

Corollary 1: The preceding result is still valid if g is replaced by a vector $\mathbf{g}(\boldsymbol{\zeta})$ in $\mathbf{y}$ space, in which case h is also replaced by a vector $\mathbf{h}$.

Corollary 2: If multiple-nested Poisson brackets are given, under the same conditions as stated before and of the form

$$[\cdots[[\mathbf{g}(\boldsymbol{\zeta}), W^{i-1}], W^{j-1}], \cdots W^{s-1}], \quad (36)$$

they may still be reduced to the form $\mathbf{h}(\boldsymbol{\zeta})$.

To prove this, one only has to note that the innermost brackets may be thus reduced, then the innermost brackets of the remaining expression, and so on until all brackets have been eliminated.

Corollary 3: The preceding still holds if $\mathbf{g}(\boldsymbol{\zeta})$ in (36) is replaced by $\mathbf{y}$, for the innermost brackets then become

$$\begin{aligned}[\mathbf{y}, W^{i-1}] &= [\chi^{(i)}, \mathbf{y}] \\ &= \overline{\nabla}\chi^{(i)} \\ &= \boldsymbol{\zeta}^{(i)} - \mathbf{f}^{(i)}(\boldsymbol{\zeta}),\end{aligned}$$

which has the form of $\mathbf{h}(\boldsymbol{\zeta})$. The remaining brackets may then be removed as before.

We now prove the main assertion. Let

$$L^{(k)}_{\Pi'}(\mathbf{g}) = [\mathbf{g}, W^{(k-1)}]$$

(note displacement of order index) so that

$$\epsilon L_{\Pi'} = \sum_{k=1} \epsilon^k L^{(k)}_{\Pi'}.$$

With this notation, the expression (36) may be rewritten

$$L_W^{(s)} \cdots L_W^{(j)} L_W^{(i)}(\mathbf{g}) \tag{37}$$

and, as was noted, this can be reduced to the form $\mathbf{h}(\boldsymbol{\zeta})$, as can the analogous expression with y replacing g.

We now have, by (32),

$$\begin{aligned}\mathbf{z} &= [1 + \sum \epsilon^k L_W^{(k)} + \tfrac{1}{2}(\sum \epsilon^k L_W^{(k)})^2 + \cdots] * \mathbf{y} \\ &= \sum_{k=0} \epsilon^k M^{(k)} * \mathbf{y},\end{aligned} \tag{38}$$

where the $M^{(k)}$ are expansion operators resembling those of (16), with the difference that account must be taken of the fact that $L_W^{(k)}$ operators with different values of k do not commute. One has, for instance [compare the last of Eqs. (13) in Ref. 5],

$$M^{(3)} = L_W^{(3)} + \tfrac{1}{2}(L_W^{(1)} L_W^{(2)} + L_W^{(2)} L_W^{(1)}) + \tfrac{1}{6}(L_W^{(1)})^3.$$

Let us denote

$$N^{(k)} = M^{(k)} - L_W^{(k)}.$$

Then, since

$$L_W^{(k)}(\mathbf{y}) = [\mathbf{y}, W^{(k-1)}] = -\bar{\boldsymbol{\nabla}} W^{(k-1)},$$

one obtains, from (38),

$$\mathbf{z} = \mathbf{y} - \sum_{k=1} \epsilon^k \bar{\boldsymbol{\nabla}} W^{(k-1)} + \sum_{k=1} \epsilon^k N^{(k)} * \mathbf{y}. \tag{39}$$

Since $N^{(k)} * \mathbf{y}$ consists only of terms of the form (37), it may be reduced to a "function" $\mathbf{h}(\boldsymbol{\zeta})$. Comparison of the last equation with (5) and (29) then identifies this function with $\mathbf{f}^{(k)}(\boldsymbol{\zeta})$, provided that (35) holds. This completes the proof of our original assertion.

4. LACINA'S EXPANSION

Lacina[6,7] has published what he claims is a simple new canonical perturbation method, leading to results similar to those obtained from the Hamilton–Jacobi equation. Unfortunately, the simplicity is more apparent than real for two reasons: First, there exists an important omission in the calculation and, secondly, there is no assumption of near-periodicity, so that the elimination of secular terms may be dispensed with. Perturbation calculations for systems without periodic character are possible, but of little interest, since their range of validity in time is usually quite limited. It is the periodic character inherent in the problems of celestial mechanics and of guiding center motion which makes possible solutions valid over long intervals in time, provided that secular terms are eliminated.

Lacina's result is easily derived by the preceding formalism and, in fact, our notation allows more concise treatment than is found in the original articles, which use a separate notation for canonical coordinates and momenta. Let a near-identity canonical transformation be given by a direct relation as in (5); the transformation is then fully specified by the various orders of $\chi^{(k)}$ appearing in (29), assuming, of course, that a particular choice of $\mathbf{f}^{(k)}$ has been selected. This choice could be the one of (30), in which case $\chi^{(k)}$ is the kth order of the conventional generating function σ, with sign reversed, or it may be the one derived in (39): any such choice may be used in what follows.

By using (29), any single component of the direct expansion, e.g.,

$$\zeta_1^{(k)} = \frac{\partial \chi^{(k)}}{\partial \bar{y}_1} + f_1^{(k)} \tag{40}$$

may be used to define $\chi^{(k)}$ (with a certain arbitrariness) and consequently the transformation. This is essentially the idea behind Lacina's approach. To define the transformation by means of $\zeta_1^{(k)}$, he uses the $(i, 1)$ component of (27):

$$\frac{\partial \zeta_i^{(k)}}{\partial \bar{y}_1} = \frac{\partial \zeta_1^{(k)}}{\partial \bar{y}_i} - \sum_{m=1}^{k-1} [\zeta_i^{(m)}, \zeta_1^{(k-m)}], \tag{41}$$

from which we have

$$\zeta_i^{(k)} = \int_C^{\bar{y}_1} \left(\frac{\partial \zeta_1^{(k)}}{\partial \bar{y}_i} - \sum_{m=1}^{k-1} [\zeta_i^{(m)}, \zeta_1^{(k-m)}] \right) d\bar{y}_1' + \mu_i^{(k)}(\tilde{\mathbf{y}}), \tag{42}$$

where $\tilde{\mathbf{y}}$ is the vector formed of the $2n - 1$ components of y excluding $\pm \bar{y}_1$ (the sign being adjusted so that this *is* a component of y).

The lower limit is arbitrary, but its choice affects $\mu_j^{(k)}$, which has to be chosen in a way assuring that (27) also holds when neither i nor j equals 1. To handle such cases, we express the Poisson brackets by means of $\mathbf{f}^{(k)}$, which is known to be a solution for $\boldsymbol{\zeta}^{(k)}$ in (27):

$$\zeta_i^{(k)} = \int_C^{\bar{y}_1} \left(\frac{\partial \zeta_1^{(k)}}{\partial \bar{y}_i} + \frac{\partial f_i^{(k)}}{\partial \bar{y}_1} - \frac{\partial f_1^{(k)}}{\partial \bar{y}_i} \right) d\bar{y}_1' + \mu_i^{(k)}(\tilde{\mathbf{y}}). \tag{43}$$

As one forms the (i,j) element of the curl of the vector of which the above expression is one component, the contributions of $\zeta_1^{(k)}$ and of $f_1^{(k)}$ in the integrand cancel out, and one is left with

$$\begin{aligned}(\bar{\boldsymbol{\nabla}} \times \boldsymbol{\zeta}^{(k)})_{ij} &= \int_C^{\bar{y}_1} \frac{\partial}{\partial \bar{y}_1'} (\bar{\boldsymbol{\nabla}} \times \mathbf{f}^{(k)})_{ij}\, d\bar{y}_1' + (\bar{\boldsymbol{\nabla}} \times \boldsymbol{\mu}^{(k)})_{ij} \\ &= (\bar{\boldsymbol{\nabla}} \times \mathbf{f}^{(k)})_{ij} - (\bar{\boldsymbol{\nabla}} \times \mathbf{f}^{(k)}(\bar{y}_1 = C))_{ij} \\ &\quad + (\bar{\boldsymbol{\nabla}} \times \boldsymbol{\mu}^{(k)})_{ij}.\end{aligned} \tag{44}$$

Here the lhs cancels with the first term on the right, and one is left with the condition

$$\mu^{(k)}(\tilde{y}) = f^{(k)}(\tilde{y}_1 = C) + \tilde{\nabla}\psi^{(k)}(\tilde{y}), \qquad (45)$$

where $\psi^{(k)}$ is any function of $\tilde{y}$ (or zero). Because only the curl of $f^{(k)}$ is involved, any choice satisfying (29) may be used. None of this appears in Lacina's work,[7] because the contributions of the lower limit of integration are inadvertently omitted in the last steps of the equations following his Eqs. (15) and (16).

Let us assume that the unperturbed problem has been prepared so that the Hamiltonian has the form given in Eq. (6). Following Lacina, we now stipulate that in the new variables the Hamiltonian equals z_1. Since the transformation is time independent, this new Hamiltonian equals the old one, given in Eq. (6), leading to the identification

$$\zeta_1^{(k)}(y) = H^{(k)}(y). \qquad (46)$$

Substituting this choice of $\zeta_1^{(k)}$ into (42) and evaluating $\zeta_i^{(k)}$ by means of (45) allows the other z_i to be derived to any desired order. Since the new Hamiltonian equals z_1, the variable conjugate to z_1 will be linear in time and all other variables (z_1 included) constants of the motion. The problem is thus essentially solved.

If one defines

$$\zeta_i^{(0)} = y_i$$

and uses (46) and (22) in (41), then the basic equation converts to

$$\sum_{m=0}^{k} [\zeta_i^{(m)}, H^{(k-m)}] = 0.$$

This is the starting point of the method of McNamara and Whiteman,[12,5] which in turn is related to Whittaker's adelphic integral,[13] Contopoulos' third integral,[14] and to other approaches quoted by Contopoulos. The difference is that McNamara and Whiteman assume a periodic character of the motion and include an extra step to ensure elimination of secular terms. While their aim is to generate one invariant only, cases may exist[15] in which a complete set of invariants can be generated, yielding a solution similar to Lacina's but free from secularity.

[1] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1953).
[2] H. Corben and P. Stehle, *Classical Mechanics* (Wiley, New York, 1960), 2nd ed.
[3] Ref. 2, Sec. 74.
[4] M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[5] D. Stern, J. Math. Phys. **11**, 2771 (1970) (previous article). The notation in that work differs from the one used here in that n there denotes the *total* number of variables.
[6] J. Lacina, Czech. J. Phys. **B19**, 130 (1969).
[7] J. Lacina, Ann. Phys. (N.Y.) **51**, 381 (1969).
[8] S. Lie, *Theorie der Transformationgruppen, Band I* (Teubner, Leipzig, 1888).
[9] G. Hori, Publ. Astron. Soc. Japan **18**, 287 (1966).
[10] A. Deprit, Celestial Mech. **1**, 12 (1969).
[11] P. Musen, J. Astronautical Sci. **12**, 129 (1965).
[12] B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).
[13] E. T. Whittaker, *Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944), 4th ed.
[14] G. Contopoulos, Astron. J. **68**, 1 (1963).
[15] See Ref. 5, Eq. (24) and text following.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Second Quantized Atomic Wavefunctions with Definite Unitary and Rotational Symmetry*

M. J. Cunningham
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 14 November 1969; Revised Manuscript Received 6 April 1970)

Some atomic wavefunctions for equivalent electrons in the group scheme $SU_2 \times (U_{2l+1} \supset R_{2l+1} \supset R_3)$ are constructed in terms of electron fermion creation and annihilation operators. The concept of semiconjugacy is defined and shown to reduce the number of states that must be explicitly calculated. The states for the d shell are calculated and tabulated.

I. INTRODUCTION

It has been known for some time[1–4] that the bases for the irreducible representations of the classical groups can be constructed in terms of sums of products of boson creation and annihilation operators acting on a suitably defined vacuum. The group U_n has been extensively studied using the canonical chain $U_n \supset U_{n-1} \supset \cdots \supset U_1$ as the solution to the state-labeling problem.[3,4] SU_3 has also been investigated by this technique using the scheme $SU_3 \supset R_3$.[3,5]

Moshinsky[6] has shown that the same analysis can be repeated using fermion operators. Using this method, Flores *et al.*[7] analyzed the group U_6 in the

Here the lhs cancels with the first term on the right, and one is left with the condition

$$\boldsymbol{\mu}^{(k)}(\tilde{\mathbf{y}}) = \mathbf{f}^{(k)}(\tilde{y}_1 = C) + \tilde{\nabla}\psi^{(k)}(\tilde{\mathbf{y}}), \tag{45}$$

where $\psi^{(k)}$ is any function of $\tilde{\mathbf{y}}$ (or zero). Because only the curl of $\mathbf{f}^{(k)}$ is involved, any choice satisfying (29) may be used. None of this appears in Lacina's work,[7] because the contributions of the lower limit of integration are inadvertently omitted in the last steps of the equations following his Eqs. (15) and (16).

Let us assume that the unperturbed problem has been prepared so that the Hamiltonian has the form given in Eq. (6). Following Lacina, we now stipulate that in the new variables the Hamiltonian equals z_1. Since the transformation is time independent, this new Hamiltonian equals the old one, given in Eq. (6), leading to the identification

$$\zeta_1^{(k)}(\mathbf{y}) = H^{(k)}(\mathbf{y}). \tag{46}$$

Substituting this choice of $\zeta_1^{(k)}$ into (42) and evaluating $\zeta_i^{(k)}$ by means of (45) allows the other z_i to be derived to any desired order. Since the new Hamiltonian equals z_1, the variable conjugate to z_1 will be linear in time and all other variables (z_1 included) constants of the motion. The problem is thus essentially solved.

If one defines

$$\zeta_i^{(0)} = y_i$$

and uses (46) and (22) in (41), then the basic equation converts to

$$\sum_{m=0}^{k} [\zeta_i^{(m)}, H^{(k-m)}] = 0.$$

This is the starting point of the method of McNamara and Whiteman,[12,5] which in turn is related to Whittaker's adelphic integral,[13] Contopoulos' third integral,[14] and to other approaches quoted by Contopoulos. The difference is that McNamara and Whiteman assume a periodic character of the motion and include an extra step to ensure elimination of secular terms. While their aim is to generate one invariant only, cases may exist[15] in which a complete set of invariants can be generated, yielding a solution similar to Lacina's but free from secularity.

[1] H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, Mass., 1953).
[2] H. Corben and P. Stehle, *Classical Mechanics* (Wiley, New York, 1960), 2nd ed.
[3] Ref. 2, Sec. 74.
[4] M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[5] D. Stern, J. Math. Phys. **11**, 2771 (1970) (previous article). The notation in that work differs from the one used here in that n there denotes the *total* number of variables.
[6] J. Lacina, Czech. J. Phys. **B19**, 130 (1969).
[7] J. Lacina, Ann. Phys. (N.Y.) **51**, 381 (1969).
[8] S. Lie, *Theorie der Transformationgruppen, Band I* (Teubner, Leipzig, 1888).
[9] G. Hori, Publ. Astron. Soc. Japan **18**, 287 (1966).
[10] A. Deprit, Celestial Mech. **1**, 12 (1969).
[11] P. Musen, J. Astronautical Sci. **12**, 129 (1965).
[12] B. McNamara and K. J. Whiteman, J. Math. Phys. **8**, 2029 (1967).
[13] E. T. Whittaker, *Analytical Dynamics of Particles and Rigid Bodies* (Dover, New York, 1944), 4th ed.
[14] G. Contopoulos, Astron. J. **68**, 1 (1963).
[15] See Ref. 5, Eq. (24) and text following.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Second Quantized Atomic Wavefunctions with Definite Unitary and Rotational Symmetry*

M. J. CUNNINGHAM
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 14 November 1969; Revised Manuscript Received 6 April 1970)

Some atomic wavefunctions for equivalent electrons in the group scheme $SU_2 \times (U_{2l+1} \supset R_{2l+1} \supset R_3)$ are constructed in terms of electron fermion creation and annihilation operators. The concept of semiconjugacy is defined and shown to reduce the number of states that must be explicitly calculated. The states for the d shell are calculated and tabulated.

I. INTRODUCTION

It has been known for some time[1–4] that the bases for the irreducible representations of the classical groups can be constructed in terms of sums of products of boson creation and annihilation operators acting on a suitably defined vacuum. The group U_n has been extensively studied using the canonical chain $U_n \supset U_{n-1} \supset \cdots \supset U_1$ as the solution to the state-labeling problem.[3,4] SU_3 has also been investigated by this technique using the scheme $SU_3 \supset R_3$.[3,5]

Moshinsky[6] has shown that the same analysis can be repeated using fermion operators. Using this method, Flores *et al.*[7] analyzed the group U_6 in the scheme $U_6 \supset R_6 \supset R_5 \supset R_3$ for all representations of U_6 of the form $\{4^a3^b2^c1^d\}$, i.e., for nucleons, while Jahn[8] has calculated the coefficients of fractional parentage for the scheme $U_5 \supset R_5 \supset R_3$, also for nucleons.

In this paper, the fermion creation operator is identified as a single-electron creation operator with orbital angular momentum l and spin of $\frac{1}{2}$, in the scheme widely used by atomic spectroscopists for a configuration of equivalent electrons[9,10]; viz., $SU_2 \times (U_{2l+1} \supset R_{2l+1} \supset R_3)$. Some of the basis states are constructed for the representations $\{2^a1^b\}$, $S = \frac{1}{2}b$ (S is the total spin quantum number) of the group $SU_2 \times U_{2l+1}$ in terms of sums of products of single-electron wavefunctions.

The method used to construct these states is essentially one of projection with the Casimir operator at the R_{2l+1} level, coupled with the demand at the R_3 level that L_{+1} and S_{+1} on the state be zero (Sec. III). Hence, unless otherwise stated, every state in this paper has $M_L = L$ and $M_S = S = \frac{1}{2}b$. Lower states can of course be derived by cranking with L_{-1} and S_{-1}. This differs from the technique of Flores *et al.* which demands that the raising operators of R_{2l+1} on the state be zero and then obtains the R_3 state by the use of a lowering operator. The method of this paper exploits the inherent greater simplicity of the atomic wavefunctions over nuclear wavefunctions, so that an explicit closed formula can be given for the states $|\{2^a1^b\}[2^c1^d]L_M\rangle$, where L_M is the maximum value of L in the branching rule for the $[2^c1^d]$ representation of R_{2l+1} upon being restricted to R_3. Closed formulas can also be given for some R_{2l+1} representations for all values of L, viz., [0], [1], [11], and [2].

In Sec. V, the concept of semiconjugacy is defined which, together with some group theory, allows any state to be written down once the state $|\{2^c1^d\}[2^c1^d]L\rangle$ is known. In Sec. VI, the preceding theory is illustrated by calculating and tabulating the wavefunctions for the multiplicity-free case of the d shell.

II. GROUP GENERATORS

Judd[10] has shown that the generators of the group U_{2l+1} are

$$X_{ab} = \sum_{k,q} (-1)^{l-a}[k]^{\frac{1}{2}} \begin{pmatrix} l & k & l \\ -a & q & b \end{pmatrix} V_q^{(k)}$$

while the generators of the group R_{2l+1} are

$$W_{ab} = \sum_{k,q} (-1)^{l-a}[1-(-1)^k][k]^{\frac{1}{2}} \begin{pmatrix} l & k & l \\ -a & q & b \end{pmatrix} V_q^{(k)}.$$

If we substitute for $V_q^{(k)}$, using the formula[11]

$$W_{\pi q}^{(\kappa k)} = -\tfrac{1}{2}(\mathbf{a}^\dagger \mathbf{a})_{\pi q}^{(\kappa k)},$$

these become

$$X_{ab} = \sum_{m_s} a^\dagger_{m_s,a} a_{m_s,b}$$

and

$$W_{ab} = \sum_{m_s} (a^\dagger_{m_s,a} a_{m_s,b} - (-1)^{a+b} a^\dagger_{m_s,-b} a_{m_s,-a}) = X_{ab} - (-1)^{a+b} X_{-b-a}, \quad (1)$$

where the a's are single-electron creation or annihilation operators, with subscripts respectively identifying spin and orbital angular-momentum projections. These operators are just proportional to Moshinsky's[6] $\mathcal{C}_a^b$ and Λ_a^b, respectively; in fact, $X_{ab} = \mathcal{C}_a^b$ and $\Lambda_a^b = \frac{1}{2}W_{ab}$.

If we invert the expression for the U_{2l+1} generators, we have

$$V_q^{(k)} = \sum_{a,b} (-1)^{l-a}[k]^{\frac{1}{2}} \begin{pmatrix} l & k & l \\ -a & q & b \end{pmatrix} X_{ab}$$

and, by noting that[10]

$$\mathbf{V}^{(1)} = \mathbf{L}[3/l(l+1)(2l+1)]^{\frac{1}{2}},$$

we have that the R_3 generators, i.e., the components of the vector L, are given by

$$\mathbf{L} = [l(l+1)(2l+1)]^{\frac{1}{2}} \times \sum_b (-1)^{l-b} \begin{pmatrix} l & 1 & l \\ -b & q & b-q \end{pmatrix} X_{b\,b-q} = \sum_b t_b^q X_{q\,b-q}, \quad (2)$$

where

$$t_b^q = (-1)^{l-b}[l(l+1)(2l+1)]^{\frac{1}{2}} \begin{pmatrix} l & 1 & l \\ -b & q & b-q \end{pmatrix}.$$

We note that $t_b^q = t_{q-b}^q$.

The generators for the group SU_2, namely $S_{\pm 1}$ and S_0, are given by $\mathbf{S} = (2l+1)/2^{\frac{1}{2}} W^{(10)}$ so that, for instance, S_{+1} in second-quantized form becomes

$$S_{+1} = \frac{1}{\sqrt{2}} \sum_{m_l} a^\dagger_{\frac{1}{2}m_l} a_{-\frac{1}{2}m_l}.$$

If we now define

$$\nabla^{\frac{1}{2}}_{a_1\cdots a_n} \nabla^{-\frac{1}{2}}_{b_1\cdots b_m} = a^\dagger_{\frac{1}{2}a_1} \cdots a^\dagger_{\frac{1}{2}a_n} a^\dagger_{-\frac{1}{2}b_1} \cdots a^\dagger_{-\frac{1}{2}b_m},$$

a somewhat specialized form of Moshinsky's[3,6] $\boldsymbol{\nabla}$, then it follows immediately from definitions that

$$X_{ab} \nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_n} \nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_m} = \delta(\mu_i, b) \nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_{i-1}a\mu_{i+1}\cdots\mu_n} \nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_m} + \delta(\nu_i, b) \nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_n} \nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_{i-1}a\nu_{i+1}\cdots\nu_m} \quad (3)$$

and

$$S_{+1} \nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_n} \nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_m} = \frac{1}{\sqrt{2}} \sum_i (-1)^i \nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_n\nu} \nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_{i-1}\nu_{i+1}\cdots\nu_m}. \quad (4)$$

III. $U_{2l+1} \supset R_{2l+1}$

The condition on the form of the state is found in this section in order that it transform according to definite U_{2l+1}, R_{2l+1}, and R_3 symmetry. Clearly,

$$\mathcal{U} = \nabla^{\frac{1}{2}}_{ll-1\cdots l-a-b+1}\nabla^{-\frac{1}{2}}_{ll-1\cdots l-a+1}$$

is the greatest-weight state of $\{2^a1^b\}$, $S = \frac{1}{2}b$ for $SU_2 \times U_{2l+1}$ since $X_{\mu\nu}\mathcal{U} = 0$, $\mu > \nu$ and $S_{+1}\mathcal{U} = 0$ while $X_{ii}\mathcal{U} = k\mathcal{U}$, where $k = 2$, 1, or 0 if $l - a + 1 < i < l$, $l - a - b + 1 < i < l - a$, or $i < l - a - b$, respectively, while $S_0\mathcal{U} = \frac{1}{2}b\mathcal{U}$. By application of the step-down operators of U_{2l+1}, namely $X_{\mu\nu}$, $\mu < \nu$, it can be seen that a state of arbitrary weight of the representation $\{2^a1^b\}$ of U_{2l+1} (but of maximum weight in SU_2) takes the form of a linear combination of states like $\nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_{a+b}}\nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_a}$. To apply the additional constraint that this state $|T\rangle$ transforms according to the representation $[2^c1^d]$ of R_{2l+1}, we demand that

$$G\,|T\rangle = g\,|T\rangle, \tag{5}$$

where G is the Casimir operator of R_{2l+1} having the eigenvalue g for the representation $[2^c1^d]$.

The Casimir operator for R_{2l+1} is[10]

$$G = \sum_{a,b} W_{ab}W_{ba}$$

which, following Moshinsky,[6] we write as [from Eq. (1)]

$$\sum_{a,b}[X_{ab} - (-1)^{a+b}X_{-b-a}][X_{ba} - (-1)^{a+b}X_{-a-b}] = 2\sum_{a,b}X_{ab}X_{ba} - 2\sum_{a,b}(-1)^{a+b}X_{ab}X_{-a-b},$$

i.e.,

$$G = 2\Gamma - 2\mathcal{T}, \tag{6}$$

where

$$\mathcal{T} = \sum_{a,b}(-1)^{a+b}X_{ab}X_{-a-b}$$

and Γ is the Casimir operator for U_{2l+1}.

Since the eigenvalues of Γ acting on a state transforming according to $\{2^a1^b\}$, which are found by operating Γ on the greatest-weight state of $\{2^a1^b\}$, are $(2l + 2 - a - b)(a + b) + (2l + 4 - a)a$, while the eigenvalues of G for the representation $\{2^c1^a\}$ are[10] $4c(2l + 2 - c) + 2d(2l + 1 - 2c - d)$, it follows that the R_{2l+1} condition (5) becomes

$$\begin{aligned}\mathcal{T}\,|T\rangle &= [(a + b)(2l + 2 - a - b) + a(2l + 4 - a)\\ &\quad -2c(2l + 2 - c) - d(2l + 1 - 2c - d)]\,|T\rangle\\ &= p\,|T\rangle.\end{aligned} \tag{7}$$

If $\mathcal{T}$ is expanded as

$$\begin{aligned}\mathcal{T} &= \sum_{a>0}(X_{-a-a}X_{aa} + X_{aa}X_{-a-a})\\ &\quad + \sum_{a>0}(X_{a-a}X_{-aa} + X_{-aa}X_{a-a})\\ &\quad + \sum_{\substack{b\\ b\neq\pm a}}\left(\sum_{a>0}(-1)^{a+b}[X_{-b-a}X_{ba} + X_{-ba}X_{b-a}]\right)\\ &\quad + \sum_{\substack{b\\ b\neq 0}}(-1)^{b}X_{-b0}X_{b0} + X_{00}X_{00}\end{aligned} \tag{8}$$

and $|T\rangle$ is expanded as $|T\rangle = \sum_s \alpha_s S$, where S is a mononomial, and if the action of each term of Eq. (8) on a general mononomial is examined, we get a set of simultaneous equations in the coefficients α_s of S, viz.

$$\alpha_s = (2a + b + 2P_s^{\uparrow\downarrow})\alpha_s + \delta(P_s^{\uparrow\downarrow} > 0)2\sum_{\substack{\text{each}\\ \text{pair}}}\sum_{\substack{b\\ b\neq a}}(-1)^{a+b}\nabla^{\frac{1}{2}}_{\cdots b\cdots}\nabla^{-\frac{1}{2}}_{\cdots -b\cdots}, \tag{9}$$

where $P_s^{\uparrow\downarrow}$ is the number of $s = \frac{1}{2}m_l$ and $s = -\frac{1}{2} - m_l$ pairs in S (e.g., $P_s^{\uparrow\downarrow}$ for $\nabla^{\frac{1}{2}}_{2-1}\nabla^{-\frac{1}{2}}_{1}$ is one, and for $\nabla^{\frac{1}{2}}_{320}\nabla^{-\frac{1}{2}}_{0-2}$ is two) which will be called external pairs. The second term of Eq. (9) is a sum, effective only if $P^{\uparrow\downarrow} > 0$ [hence, $\delta(P_s^{\uparrow\downarrow} > 0)$], through the coefficients of those mononomials which are related to S by having each of its a, $-a$ external pairs replaced by a b, $-b$ external pair, for all $b \neq a$. Equation (9), then, is the condition on a state $|T\rangle$ in order that it transform according to $[2^c1^d]$ of R_{2l+1}.

If the set of Eqs. (9) are written as a matrix array A in the coefficients α_s, then the R_{2l+1} condition becomes the following: If the eigenvalues of matrix A are equal to p, then the corresponding eigenstate transforms according to $\{2^a1^b\}[2^c1^d]$ of U_{2l+1} and R_{2l+1}. It can be seen that A breaks up into block diagonal form, each block associated with a different number of external pairs. The problem thus reduces to finding the eigenvalues and eigenstates for each block separately.

We note now that the $U_{2l+1} \supset R_{2l+1}$ branching rules can be divided into two classes,[10,12,13] viz.,

$$a + b \leq l, \quad \{2^a1^b\} \to \sum_{x=0}^{a}[2^{a-x}1^b], \tag{10a}$$

$$a + b > l, \quad \{2^a1^b\} = \sum_{x=0}^{\min(a,a+b-l-1)}[2^{a-x,2l+1-2a-b+2x}] + \sum_{x=0}^{a}[2^{a-x}1^b]. \tag{10b}$$

In both of the above cases it is assumed that $2a + b \leq 2l + 1$. States with $2a + b > 2l + 1$ can be found by utilizing the particle–hole equivalence.[9,10] In both cases above,

$$p = 4xl + 2x(x + 2) + 2(1 - 2x)a + (1 - 2x)b.$$

We now consider the function

$$F_m = \sum_{\mu_1>\mu_2>\cdots>\mu_m} (-1)^{\mu_1+\mu_2+\cdots+\mu_m} \nabla^{\frac{1}{2}}_{q_1q_2\cdots q_{c'+d'}\mu_1\mu_2\cdots\mu_m} \times \nabla^{-\frac{1}{2}}_{r_1r_2\cdots r_{c'}-\mu_1-\mu_2\cdots-\mu_m}, \quad (11)$$

which has

$$\alpha_{\nabla^{\frac{1}{2}}_{q_1\cdots q_{c'+d'}\mu_1\cdots\mu_m}\nabla^{-\frac{1}{2}}_{r_1\cdots r_{c'}-\mu_1\cdots-\mu_m}} = (-1)^{\mu_1+\mu_2+\cdots+\mu_m},$$

where $m = a - c'$ and c' and d' are the values of c and d before Murnagham's[13] modification rules have been applied, that is, $c' = a - x$ and $d' = b$. Substitution for these values of α_s in (9) readily verifies that F_m transforms according to $\{2^a1^b\}[2^{a-m}1^b]$ or

$$\{2^a1^b\}[2^{a-m}1^{2+1-2a-b+2m}]$$

according to the relative sizes of a, b, and m.

As a special case, the state $|\{2^a1^b\}[2^b1^d]L_m\rangle$, where L_M is the maximum L in the $R_{2l+1} \rightarrow R_3$ branching rule for $[2^c1^d]$, can now be written down as

$$\mathcal{F}_m = N \sum_{\mu_1>\mu_2>\cdots>\mu_m} (-1)^{\mu_1+\mu_2+\cdots+\mu_m} \times \nabla^{\frac{1}{2}}_{l\,l-1\cdots l-c'-d'+1\,\mu_1\mu_2\cdots\mu_m} \nabla^{-\frac{1}{2}}_{l\,l-1\cdots l-c'+1-\mu_1\cdots-\mu_m}, \quad (12)$$

where $m = a - c'$, as before, and N is a normalization constant. That this is so can be seen by applying the raising operators W_{ab}, $b < a$, of R_{2l+1} to $\mathcal{F}_m$ to give a null result. It follows from (12) that

$$\begin{aligned} L_M &= 2 \sum_{i=l-c'+1}^{l} i + \sum_{i=l-c'-d'+1}^{l-c'} i \\ &= (2l + 1 - c')c' + \tfrac{1}{2}(2l + 1 - 2c' - d')d'. \end{aligned}$$

Since each term in $\mathcal{F}_m$ has a coefficient of modulus one, the normalization can be derived simply by counting the number of mononomials in $\mathcal{F}_m$. Hence,

$$\begin{aligned} \frac{1}{N^2} &= \sum_{\mu_1=1}^{M} \sum_{\mu_2=1}^{\mu_1} \cdots \sum_{\mu_m=1}^{\mu_{m-1}} 1 = \frac{1}{m!} \prod_{r=0}^{m-1} (M + r) \\ &= \frac{(M + m - 1)!}{m!\,(M - 1)!}, \end{aligned}$$

where M is the maximum number of different values that μ_1 (or μ_i) can take, i.e., $M = 2l + 2 - a - b - c'$.

Finally, then,

$$\begin{aligned} &|\{2^a1^b\}[2^c1^d]L_M\rangle \\ &= \left(\frac{(a - c')!\,(2 + 1 - a - c' - b)!}{(2 + 1 - 2c' - b)!}\right)^{\frac{1}{2}} \\ &\times \sum_{\mu_1>\mu_2>\cdots>\mu_{a-c'}} (-1)^{\mu_1+\mu_2+\cdots+\mu_{a-c'}} \\ &\times \nabla^{\frac{1}{2}}_{l\,l-1\cdots l-c'-d'+1\,\mu_1\mu_2\cdots\mu_{a-c'}} \nabla^{-\frac{1}{2}}_{l\,l-1\cdots l-c'+1\,-\mu_1-\mu_2\cdots-\mu_{a-c'}}. \end{aligned} \quad (13)$$

A second eigenstate of the submatrix A associated with one pair has been obtained in the form of a single-pair identity viz.

$$\sum_{\text{all } a} (-1)^a \alpha_{\nabla^{\frac{1}{2}}_{\cdots a}\,\nabla^{\frac{1}{2}}_{\cdots -a}} = 0. \quad (14)$$

In summary, the submatrix associated with no external pairs gives, as its complete set of solutions, mononomials transforming according to the pair of U_{2l+1} and R_{2l+1} representations associated with $x = 0$ [see (10)], which we shall shorten to transforming as $x = 0$. The $n \times n$ matrix associated with one pair gives a 1-dimensional solution F_1[(11)] transforming as $x = 1$ and an $(n - 1)$-dimensional solution (14) transforming as $x = 0$. This clearly exhausts the one-pair solutions. For a general number of pairs, only a 1-dimensional solution F_m[(11)], which transforms as $x = m$, has been found. F_m has, as a special case, the state

$$\mathcal{F}_m = |\{2^a1^b\}[2^c1^d]L_M\rangle.$$

It will be shown in Sec. VI that this special case can be derived by the use of group-theoretical arguments.

IV. $R_{2l+1} \supset R_3$

We now consider the chain $R_{2l+1} \supset R_3$. No general formula for arbitrary L contained within a given R_{2l+1} representation has been found. The 1- and 2-particle case, however, can be solved generally since $[0]S$ and $[1]l$ are special cases of $\mathcal{F}_m$; writing

$$|\{11\}[11]L\rangle \quad \text{as} \quad N \sum_a \beta_a \nabla^{\frac{1}{2}}_{L-a\,a}$$

and applying the L_{+1} condition, we get a set of simultaneous equations, namely,

$$t^1_{L-l+1}\beta_{\nabla^{\frac{1}{2}}_{L-l\,l}} + t^1_l\beta_{\nabla^{\frac{1}{2}}_{L-l+1\,l-1}} = 0, \quad (15a)$$

$$t^1_{L-l+2}\beta_{\nabla^{\frac{1}{2}}_{L-l+1\,l-1}} + t^1_{l-1}\beta_{\nabla^{\frac{1}{2}}_{L-l+2\,l-2}} = 0, \quad (15b)$$

$$t^1_{l-1}\beta_{\nabla^{\frac{1}{2}}_{l-2\,L-l+2}} + t^1_{L-l+2}\beta_{\nabla^{\frac{1}{2}}_{l-1\,L-l+1}} = 0, \quad (15b')$$

$$t^1_l\beta_{\nabla^{\frac{1}{2}}_{l-1\,L-l+1}} + t^1_{L-l+1}\beta_{\nabla^{\frac{1}{2}}_{l\,L-l}} = 0, \quad (15a')$$

which has the solution

$$\beta_a = (-1)^a \prod_{y=a+1}^{l} \frac{t_{L-y+1}}{t_y}, \quad (16)$$

which is understood to be $(-1)^a = (-1)^l$ if $a = l$. Since the Eqs. (15a), (15a′), (15b), (15b′), etc., are equal, the summation is restricted to $\frac{1}{2}(L + 1) \leq a \leq l$, i.e.,

$$|\{11\}[11]L\rangle = N \sum_{a=\frac{1}{2}(L+1)}^{l} \beta_a \nabla^{\frac{1}{2}}_{L-a\,a}.$$

No solution exists if L is even as required by the branching rule. We get a similar solution for $|\{2\}\{2\}L\rangle$, namely,

$$|\{2\}[2]L\rangle = N \sum_{a=L-l}^{l} \beta_a \nabla^{\frac{1}{2}}_{L-a} \nabla^{-\frac{1}{2}}_{a}$$

with the same β_a as for [11], (16), the summation in this case being unrestricted. Note that

$$S_{+1}\,|\{2\}[2]L\rangle = N \sum_a (\beta_a \nabla^{\frac{1}{2}}_{L-a\,a} + \beta_{L-a} \nabla^{\frac{1}{2}}_{a\,L-a})$$

so that, since $\beta_a = (-1)^L \beta_{L-a}$, this expression vanishes for L even, as required by the branching rule.

To find N, we must evaluate the expression

$$\sum_a \beta_a^2 = \sum_a \prod_{y=a+1}^{l} \left(\frac{t^1_{L-y+1}}{t^1_y}\right)^2, \tag{17}$$

with the appropriate limits for each case. We have that

$$\frac{t^1_a}{t^1_b} = \frac{(-1)^a \begin{pmatrix} l & 1 & l \\ -a & 1 & a-1 \end{pmatrix}}{(-1)^b \begin{pmatrix} l & 1 & l \\ -b & 1 & b-1 \end{pmatrix}} = \left(\frac{(l+a)(l-a+1)}{(l+b)(l-b+1)}\right)^{\frac{1}{2}},$$

so that (17) becomes

$$\begin{aligned}\sum_a \prod_{y=a+1}^{l} &\frac{(l+L-y+1)(l-L+y)}{(l+y)(l-y+1)} \\ &= \frac{(2l-L)!}{(2l)!\,L!} \sum_a \frac{(l+a)!\,(l+L-a)!}{(l-a)!\,(l-L+a)!} \\ &= \frac{(2l-L)!\,L!}{(2l)!} \sum_a \binom{l+a}{l-L+a}\binom{l+L-a}{l-a} \\ &= (\tfrac{1}{2})\frac{(2l-L)!\,L!}{(2l)!}\binom{2l+L+1}{2l-L},\end{aligned} \tag{18}$$

as a consequence of one of the many variants of Vandermonde's convolution identity. The $\frac{1}{2}$ occurs only in the restricted summation case of [11]. Equation (18) implies that

$$N = \left(\frac{(2)(2l)!\,(2L+1)!}{L!\,(2l+1+L)!}\right)^{\frac{1}{2}}$$

so that finally

$$|\{11\}[11]L\rangle = \left(\frac{2(2L+1)!\,(2l-L)!}{(L!)^2(2l+1+L)!}\right)^{\frac{1}{2}} \sum_{a=\frac{1}{2}(L+1)}^{l} (-1)^a \times \left(\frac{(l+a)!\,(l+L-a)!}{(l-a)!\,(l-L+a)!}\right)^{\frac{1}{2}} \nabla^{\frac{1}{2}}_{L-a\,a} \tag{19a}$$

and

$$|\{2\}[2]L\rangle = \left(\frac{(2L+1)!\,(2l-L)!}{(L!)^2(2l+1+L)!}\right)^{\frac{1}{2}} \sum_{a=L-l}^{l} (-1)^a \times \left(\frac{(l+a)!\,(l+L-a)!}{(l-a)!\,(l-L+a)!}\right)^{\frac{1}{2}} \nabla^{\frac{1}{2}}_{L-a} \nabla^{-\frac{1}{2}}_{a}. \tag{19b}$$

To find the expansion of a 3-or-more-particle state of general L and greatest weight in SU_2, in lieu of a general formula, we must demand that

$$L_{+1} = \sum_b t^q_b X_{q\,b-q}(2) \quad \text{and} \quad S_{+1}$$

on the state be zero, giving a set of simultaneous equations, which in the general case will have a set of solutions corresponding to the multiplicity of L in the branching rule of $[2^c1^d]$ upon restriction to R_3. If at any stage in the $U_{2l+1} \supset R_{2l+1} \supset R_3$ chain, one attempts to construct a state forbidden by the corresponding branching rule, one will get a vanishing result.

V. SEMICONJUGACY

In this section, it is shown that once the states $|\{2^c1^d\}[2^c1^d]L\rangle$ have been calculated explicitly, all other general states may be written down. We start by noting that the branching rules given in Sec. III imply that a general state falls into two classes, namely,

$$|\{2^{c+x}1^d\}[2^c1^d]L\rangle \quad \text{and} \quad |\{2^{c+x}1^{2l+1-2c-d}\}[2^c1^d]L\rangle.$$

We consider now the product state

$$|\{2^c1^d\}[2^c1^d]L\rangle\,|\{2^x\}[0]S\rangle.$$

Since the only unitary representation of the form $\{2^a1^b\}$ contained in the direct product $\{2^c1^d\} \times \{2^x\}$ is $\{2^{c+x}1^d\}$, the product state is proportional to $|\{2^{c+x}1^d\}[2^c1^d]L\rangle$, the first of the two classes of a general state. Similarly,

$$|\{2^c1^{2l+1-2c-d}\}[2^c1^d]L\rangle\,|\{2^x\}[0]S\rangle$$

is proportional to

$$|\{2^{c+x}1^{2l+1-2c-d}\}[2^c1^d]L\rangle,$$

i.e., the second class. Since $|\{2^x\}[0]S\rangle$ is known (equals $\mathcal{F}_x$, Sec. VI) and $|\{2^c1^d\}[2^c1^d]L\rangle$ is assumed known, we now have the first class of a general state. To find the second class, we have only to find how to form

$$|\{2^c1^{2l+1-2c-d}\}[2^c1^d]L\rangle, \quad \text{given } |\{2^c1^d\}[2^c1^d]L\rangle.$$

For this, the operation of semiconjugacy is introduced and represented by $\mathcal{R}$. Associated with a given set $\mathcal{A}$ of m_l quantum is a complementary set $\bar{\mathcal{A}}$, defined such that $\mathcal{A} \cup \bar{\mathcal{A}} = \{l, l-1, \cdots, 1, 0, -1, \cdots, -l+1, -l\}$ and $\mathcal{A} \cap \bar{\mathcal{A}} = \phi$, the null set; for example, if $l = 2$, and $\mathcal{A} = \{2, 1\}$, then $\bar{\mathcal{A}} = \{0, -1, -2\}$. Semiconjugacy is now defined by its action on a mononomial, namely,

$$\mathcal{R}\nabla^{\frac{1}{2}}_{\mu_1\cdots\mu_m}\nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_n} = \bar{\nabla}^{\frac{1}{2}}_{\mu_1\cdots\mu_m}\nabla^{-\frac{1}{2}}_{\nu_1\cdots\nu_n}. \tag{20}$$

Here

$$\bar{\nabla}^{\frac{1}{2}}_{\mu_1\cdots\mu_m} = \nabla^{\frac{1}{2}}_{\mu_1'\cdots\mu'_{2l+1-m}},$$

where $\{\mu_1' \cdots \mu_{2l+1-m}'\}$ is the negative of the complementary set to $\{\mu_1 \cdots \mu_m\}$. For instance, if $\mathcal{A} = \{2, 1\}$ and $l = 2$, $\bar{\mathcal{A}} = \{0, -1, -2\}$ and the negative of this set is $\{210\}$. The phase given in the definition (20) of $\mathfrak{R}$ (namely $+1$) holds only if $\mu_1 \cdots \mu_m$ and $\mu_1' \cdots \mu_{2l+1-m}'$ are ordered after the fashion $\mu_1 > \mu_2 > \cdots > \mu_m$ and $\mu_1' > \mu_2' > \cdots > \mu_{2l+1-m}'$. The operation is equivalent to replacing spin-up electrons only, with their corresponding holes; hence, its name.

It remains to show that $\mathfrak{R}$ has the desired property that $\mathfrak{R}\,|\{2^c1^d\}[2^c1^d]L\rangle = |\{2^c1^{2l+1-2c-d}\}[2^c1^d]L\rangle$, allowing the right-hand state to be written down simply by placing a bar over the $\nabla^{\frac{1}{2}}$ of each mononomial in the expansion of the left-hand state, assumed known. We see first that each mononomial

$$\bar{\nabla}^{\frac{1}{2}}_{q_1\cdots q_{a+b}}\nabla^{\frac{1}{2}}_{r_1\cdots r_a}$$

has the same M_L value as

$$\nabla^{\frac{1}{2}}_{q_1\cdots q_{a+b}}\nabla^{-\frac{1}{2}}_{r_1\cdots r_a}$$

since, if

$$q_1 + q_2 + \cdots + q_{a+b} = Q,$$

then the complementary set to q sums to $-Q$. Moreover,

$$L_{+1}\nabla^{\frac{1}{2}}_{\cdots\mu_i\,\mu_i-k\cdots} = \cdots + (t^1_{\mu_i+1}\nabla^{\frac{1}{2}}_{\cdots\mu_i+1\,\mu_i-k\cdots}) + t^1_{\mu_i-k+1}\nabla^{\frac{1}{2}}_{\cdots\mu_i\,\mu_i-k+1} + \cdots,$$

where the first term does not exist if $\mu_{i-1} = \mu_i + 1$ while

$$\begin{aligned} L_{+1}\bar{\nabla}^{\frac{1}{2}}_{\cdots\mu_i\,\mu_i-k\cdots} &= L_{+1}\nabla^{\frac{1}{2}}_{(-\mu_i-1)-\mu_i+1\,-\mu_i+2\cdots-\mu_i+k-1\cdots} \\ &= \cdots + (t^1_{-\mu_i}\nabla^{\frac{1}{2}}_{(-\mu_i)-\mu_i+1\cdots-\mu_i+k-1\cdots}) \\ &\quad + t^1_{-\mu_i+k}\nabla^{\frac{1}{2}}_{(-\mu_i-1)-\mu_i+1\cdots\mu_i+k-2\,-\mu_i+k\cdots} + \cdots \\ &= \cdots + (t^1_{\mu_i+1}\bar{\nabla}^{\frac{1}{2}}_{\cdots\mu_i+1\,\mu_i-k\cdots}) \\ &\quad + t_{\mu_i-k+1}\bar{\nabla}^{\frac{1}{2}}_{\cdots\mu_i\,\mu_i-k+1} + \cdots. \end{aligned}$$

Hence the simultaneous equations, derived by $\mathfrak{R}$-transforming the L_{+1} simultaneous equations for the state $|\{2^c1^d\}[2^c1^d]\rangle$, are again a set of simultaneous equations satisfying the L_{+1} condition.

We consider now the action of $\mathfrak{R}$ on the $(n-1)$-dimensional solution to the one-pair submatrix of A [(14)]:

$$\sum_x (-1)^x\nabla^{\frac{1}{2}}_{q\,x}\nabla^{-\frac{1}{2}}_{r\,-x},$$

where

$$q \equiv q_1q_2\cdots q_{a+b-1} \quad\text{and}\quad r \equiv r_1r_2\cdots r_{a-1},$$

$$\sum_x (-1)^x\nabla^{\frac{1}{2}}_{q\,x}\nabla^{-\frac{1}{2}}_{r-x} = \sum_{x=-q_i} (-1)^{q_i}\nabla^{\frac{1}{2}}_{q-q_i}\nabla^{-\frac{1}{2}}_{r\,q_i} + (\nabla^{\frac{1}{2}}_{q\,0}\nabla^{-\frac{1}{2}}_{r\,0}) + \sum_{\substack{x\neq 0\\ \neq -q_i}} (-1)^x\nabla^{\frac{1}{2}}_{q\,x}\nabla^{-\frac{1}{2}}_{r-x}.$$

The zero term is bracketed, since, if r_i or q_i equals zero, it does not exist. Thus,

$$\begin{aligned} \mathfrak{R}\sum_x (-1)^x\nabla^{\frac{1}{2}}_{q\,x}\nabla^{-\frac{1}{2}}_{r-x} &= \sum_{q_i} (-1)^{q_i}M_{-q_i}\nabla^{\frac{1}{2}}_{\bar{q}\dot{-}q_i}\nabla^{-\frac{1}{2}}_{r\,q_i} + (M_0\nabla^{\frac{1}{2}}_{\bar{q}\dot{-}0}\nabla^{-\frac{1}{2}}_{r\,0}) \\ &\quad + \sum_{\substack{x\neq 0\\ \neq -q_i}} (-1)^x M_x\nabla^{\frac{1}{2}}_{\bar{q}\dot{-}x}\nabla^{-\frac{1}{2}}_{r\,x}, \end{aligned} \tag{21}$$

where the first summation is through q_i, such that there exists a $q_j = -q_i$ in q, while M_g is the phase required to order $\nabla^{\frac{1}{2}}_{\bar{q}g}$. So, if g belongs to position k in $\nabla^{\frac{1}{2}}$ counting from the left, $M_g = (-1)^{k-a-b}$. We take $\bar{q}$ as the negative of the complementary set to q, i.e.,

$$\mathfrak{R}\nabla^{\frac{1}{2}}_q = \bar{\nabla}^{\frac{1}{2}}_q = \nabla^{\frac{1}{2}}_{\bar{q}},$$

while $\dot{-}$ means "without," e.g., $\{2, 1, 0\} \dot{-} 0 = \{21\}$. The three summations of Eq. (21) taken together are thus an instruction to take each q_i from $\bar{q}$ and place it in the spin-down space. We wish now to find the position i of $-g$ in $\bar{q}$. To do so, we reason that, since the number of blanks in the sequence $l\; l-1 \cdots g$ for the set of quantum numbers $q \dotplus g$ is both equal to $l - g - k + 1$ and the position of $-g$ in $\bar{q}$ counting from the *right*, then the position of $-g$ equals the difference of the length of $\bar{q}$ and the *right* position of g plus one, i.e.,

$$\begin{aligned} i &= (2l + 1 - a - b) - (l - g - k + 1) + 1 \\ &= l + g + k + 1 - a - b. \end{aligned}$$

Finally, then,

$$\mathfrak{R}\sum_x (-1)^x\nabla^{\frac{1}{2}}_{q\,x}\nabla^{-\frac{1}{2}}_{r-x} = (-1)^{l+1}\sum_i (-1)^i\nabla_{\bar{q}\dot{-}g_i}\nabla^{-\frac{1}{2}}_{r\,g_i}. \tag{22}$$

This, however, is just proportional to the general type of equation we get if we demand that the states

$$\bar{\nabla}^{\frac{1}{2}}_{q_1\cdots q_{a+b}}\nabla^{-\frac{1}{2}}_{r_1\cdots r_a}$$

satisfy the S_{+1} condition [cf. (4)]. Conversely, we find that any S_{+1} equation for the

$$\nabla^{\frac{1}{2}}_{q_1\cdots q_{a+b}}\nabla^{-\frac{1}{2}}_{r_1\cdots r_a}$$

states transform under $\mathfrak{R}$ to a single-pair R_{2l+1} condition (18) for

$$\bar{\nabla}^{\frac{1}{2}}_{q_1\cdots q_{a+b}}\nabla^{-\frac{1}{2}}_{r_1\cdots r_a}.$$

Since $\mathfrak{R}$ can clearly be applied in the reverse sense (i.e., from $\bar{\nabla}$ to ∇) giving the same result, we have that the *total* set of equations required to solve for the state $|\{2^c1^d\}[2^c1^d]L\rangle$ is identical apart from a barred $\nabla^{\frac{1}{2}}$ to the *total* set of equations for $\mathfrak{R}\,|\{2^c1^d\}[2^c1^d]L\rangle$. Since

$$\bar{\nabla}^{\frac{1}{2}}_{q_1\cdots q_{a+b}}\nabla^{-\frac{1}{2}}_{r_1\cdots r_a}$$

has $2l + 1 - a - b$ terms in the spin-up space and

M_S and M_L are maximum, we have as required

$$\mathcal{R}\,|\{2^c1^d\}[2^c1^d]L\rangle = |\{2^c1^{2l+1-2c-d}\}[2^c1^d]L\rangle. \quad (23)$$

Now since a general state is proportional to the product of $|\{2^c1^d\}[2^c1^d]L\rangle$ or $\mathcal{R}\,|\{2^c1^d\}[2^c1^d]L\rangle$ with $|\{2^x\}[0]S\rangle$, it is clear that the proportionality constant is dependent only on the values of c and d and is thus equal to the normalization of $\mathcal{F}_x$ divided by the normalization of $|\{2^x\}[0]S\rangle$, i.e., is equal to

$$\left(\frac{(2l+1-2c-d-x)!\,(2l+1)!}{(2l+1-2c-d)!\,(2l+1-x)!}\right)^{\frac12}.$$

VI. THE d SHELL

We have shown how to reduce the problem of constructing atomic wavefunctions in the $U_{2l+1} \supset R_{2l+1} \supset R_3$ to finding only the states $|\{2^c1^d\}[2^c1^d]L\rangle$. Unfortunately, it has not been possible to find a closed algebraic expression for these terms in general (with the exception of a few special cases given in Sec. IV) so that they must be found by solving the set of simultaneous equations given by the L_{+1} and S_{+1} conditions and applying additionally the single-paired condition (14) given in Sec. III. Even this may not be enough; since in Sec. III only one solution, corresponding to one eigenvalue, could be given to the multipaired case (11), it may happen that some of the eigenvalues for two or more pairs are also equal to the p value (7) of the state in question. Any attempt to solve for this state, with only zero and one pair, will give a vanishing result. In practice, this serves as a guide to finding which states are of this type.

In the d shell the only term of this type is $|\{22\}[22]S\rangle$. The double-paired submatrix for this case has eigenvalues 10, 5 (four times), and 2 (five times). The eigenvalue of 10, being the same as the p value for $|\{22\}[0]S\rangle$, corresponds to the doubly paired solution to this state, namely $\mathcal{F}_2$ (12). The eigenvalue of 2 equals the p value for $|\{22\}[22]S\rangle$ and gives an eigenstate spanning a 5-dimensional space, described by the following

TABLE I. The states $|\{2^c1^d\}[2^c1^d]L\rangle$ of the d shell.

State quantum numbers	Mononomial expansion
$\{0\}[0]S$	—
$\{1\}[1]D$	$\nabla_2^{\frac12}$
$\{11\}[11]F$	$\nabla_{2\;1}^{\frac12}$
$\{11\}[11]P$	$5^{-\frac12}(2^{\frac12}\nabla_{-1\;2}^{\frac12} - 3^{\frac12}\nabla_{0\;1}^{\frac12})$
$\{2\}[2]G$	$\nabla_2^{\frac12}\nabla_2^{-\frac12}$
$\{2\}[2]D$	$7^{-\frac12}(2^{\frac12}\nabla_0^{\frac12}\nabla_2^{-\frac12} - 3^{\frac12}\nabla_1^{\frac12}\nabla_1^{-\frac12} + 2^{\frac12}\nabla_2^{\frac12}\nabla^{-\frac12})$
$\{21\}[21]H$	$\nabla_{2\;1}^{\frac12}\nabla_2^{-\frac12}$
$\{21\}[21]G$	$5^{-\frac12}(3^{\frac12}\nabla_{2\;1}^{\frac12}\nabla_1^{-\frac12} - 2^{\frac12}\nabla_{2\;0}^{\frac12}\nabla_1^{-\frac12})$
$\{21\}[21]F$	$12^{-\frac12}(\nabla_{2\;1}^{\frac12}\nabla_0^{-\frac12} - \nabla_{2\;0}^{\frac12}\nabla_1^{-\frac12} + 6^{\frac12}\nabla_{2-1}^{\frac12}\nabla_2^{-\frac12} - 2\nabla_{1\;0}^{\frac12}\nabla_2^{-\frac12})$
$\{21\}[21]D$	$(\frac{3}{28})^{\frac12}(\nabla_{2\;1}^{\frac12}\nabla_{-1}^{-\frac12} - \nabla_{2\;0}^{\frac12}\nabla_0^{-\frac12} - \frac13\nabla_{2-1}^{\frac12}\nabla_1^{-\frac12} + \frac53\nabla_{2-2}^{\frac12}\nabla_2^{-\frac12} - \frac43\nabla_{1-1}^{\frac12}\nabla_2^{-\frac12} + \frac83^{\frac12}\nabla_{1\;0}^{\frac12}\nabla_1^{-\frac12})$
$\{21\}[21]P$	$(\frac{8}{35})^{\frac12}(\frac34\nabla_{1\;0}^{\frac12}\nabla_0^{-\frac12} - \nabla_{1\;2}^{\frac12}\nabla_{-2}^{-\frac12} - \frac34\nabla_{1-1}^{\frac12}\nabla_1^{-\frac12} - \frac12\nabla_{1-2}^{\frac12}\nabla_2^{-\frac12} + \frac12(6)^{-\frac12}\nabla_{2-1}^{\frac12}\nabla_0^{-\frac12} - (\frac23)^{\frac12}\nabla_{2\;0}^{\frac12}\nabla_{-1}^{-\frac12}$ $+ \frac52(6)^{-\frac12}\nabla_{0-1}^{\frac12}\nabla_2^{-\frac12} + \frac12\nabla_{2-2}^{\frac12}\nabla_1^{-\frac12})$
$\{22\}[22]I$	$\nabla_{2\;1}^{\frac12}\nabla_{2\;1}^{-\frac12}$
$\{22\}[22]G$	$(\frac{1}{33})^{\frac12}(\nabla_{2\;1}^{\frac12}\nabla_{2-1}^{-\frac12} + 2\nabla_{2\;0}^{\frac12}\nabla_{2\;0}^{-\frac12} + \nabla_{2-1}^{\frac12}\nabla_{2\;1}^{-\frac12} - 3(\frac32)^{\frac12}\nabla_{2\;1}^{\frac12}\nabla_{1\;0}^{-\frac12} - 3(\frac32)^{\frac12}\nabla_{1\;0}^{\frac12}\nabla_{2\;1}^{-\frac12})$
$\{22\}[22]F$	$(\frac{3}{40})^{\frac12}(\nabla_{2\;1}^{\frac12}\nabla_{1-1}^{-\frac12} - \nabla_{2\;0}^{\frac12}\nabla_{1\;0}^{-\frac12} + 2\nabla_{2-2}^{\frac12}\nabla_{1\;2}^{-\frac12} + \nabla_{1-1}^{\frac12}\nabla_{2-1}^{-\frac12} - \nabla_{1\;0}^{\frac12}\nabla_{2\;0}^{-\frac12} + 2\nabla_{2\;1}^{\frac12}\nabla_{2-2}^{-\frac12}$ $- (\frac23)^{\frac12}\nabla_{2\;0}^{\frac12}\nabla_{2-1}^{-\frac12} - (\frac23)^{\frac12}\nabla_{2-1}^{\frac12}\nabla_{2\;0}^{-\frac12})$
$\{22\}[22]D$	$(\frac{1}{21})^{\frac12}(\nabla_{2-1}^{\frac12}\nabla_{1\;0}^{-\frac12} + \nabla_{2\;0}^{\frac12}\nabla_{2-2}^{-\frac12} - (\frac32)^{\frac12}\nabla_{2\;1}^{\frac12}\nabla_{1-2}^{-\frac12} - 6^{\frac12}\nabla_{1\;0}^{\frac12}\nabla_{1\;0}^{-\frac12} + \nabla_{1\;0}^{\frac12}\nabla_{2-1}^{-\frac12} + \nabla_{2-2}^{\frac12}\nabla_{2\;0}^{-\frac12}$ $- (\frac32)^{\frac12}\nabla_{1-2}^{\frac12}\nabla_{2\;1}^{-\frac12} + \nabla_{2\;0}^{\frac12}\nabla_{1-1}^{-\frac12} - 6^{\frac12}\nabla_{2-1}^{\frac12}\nabla_{2-1}^{-\frac12} + \nabla_{1-1}^{\frac12}\nabla_{2\;0}^{-\frac12}$
$\{22\}[22]S$	$(\frac{3}{70})^{\frac12}(\nabla_{2\;1}^{\frac12}\nabla_{-1-2}^{-\frac12} - \nabla_{0-1}^{\frac12}\nabla_{1\;0}^{-\frac12} - \nabla_{0-2}^{\frac12}\nabla_{2\;0}^{-\frac12} - \frac13\nabla_{1-2}^{\frac12}\nabla_{2-1}^{-\frac12} - \frac43\nabla_{1-1}^{\frac12}\nabla_{2-2}^{-\frac12} + (\frac83)^{\frac12}\nabla_{1\;0}^{\frac12}\nabla_{1-2}^{-\frac12}$ $+ 2(\frac23)^{\frac12}\nabla_{2-1}^{\frac12}\nabla_{0-1}^{-\frac12}$ + their opposites[a] + $(\frac53\nabla_{2-2}^{\frac12}\nabla_{2-2}^{-\frac12} - \frac13\nabla_{1-1}^{\frac12}\nabla_{1-1}^{-\frac12}))$

[a] E.g., the opposite of $-\frac13\nabla_{1-2}^{\frac12}\nabla_{2-1}^{-\frac12}$ is $-\frac13\nabla_{2-1}^{\frac12}\nabla_{1-2}^{-\frac12}$.

equations:

$$\alpha_{\nabla^{\frac12}_{2\,1}\nabla^{-\frac12}_{-1-2}} - \alpha_{\nabla^{\frac12}_{0-1}\nabla^{-\frac12}_{1\,0}} + \alpha_{\nabla^{\frac12}_{0-2}\nabla^{-\frac12}_{2\,0}} - \alpha_{\nabla^{\frac12}_{-1-2}\nabla^{-\frac12}_{21}} = 0,$$

$$\alpha_{\nabla^{\frac12}_{2\,0}\nabla^{-\frac12}_{0-2}} - \alpha_{\Delta^{\frac12}_{1-1}\nabla^{-\frac12}_{1-1}} + \alpha_{\nabla^{\frac12}_{1-2}\nabla^{-\frac12}_{2-1}} + \alpha_{\nabla^{\frac12}_{1-2}\nabla^{-\frac12}_{2\,1}} = 0,$$

$$\alpha_{\nabla^{\frac12}_{2-1}\nabla^{-\frac12}_{1-2}} - \alpha_{\nabla^{\frac12}_{1-1}\nabla^{-\frac12}_{1-1}} + \alpha_{\nabla^{\frac12}_{0-2}\nabla^{-\frac12}_{2\,0}} - \alpha_{\nabla^{\frac12}_{-1-2}\nabla^{-\frac12}_{1\,0}} = 0,$$

$$\alpha_{\nabla^{\frac12}_{1\,0}\nabla^{-\frac12}_{0-1}} - \alpha_{\nabla^{\frac12}_{1-1}\nabla^{-\frac12}_{1-1}} + \alpha_{\nabla^{\frac12}_{1-2}\nabla^{-\frac12}_{2-1}} + \alpha_{\nabla^{\frac12}_{0-1}\nabla^{-\frac12}_{1\,0}} - \alpha_{\nabla^{\frac12}_{0-2}\nabla^{-\frac12}_{2\,0}} + \alpha_{\nabla^{\frac12}_{-1-2}\nabla^{-\frac12}_{2\,1}} = 0.$$

This, taken with the L_{+1} and S_{+1} equations, gives the expansion of $|\{22\}[22]S\rangle$ as tabulated in Table I. There is no single-paired condition for this state, for,if it has one pair,it has two.

As an example, the equations for the state $|\{21\}[21]D\rangle$ are displayed: the R_{2l+1} equation

$$\alpha_{\nabla^{\frac12}_{2\,1}\nabla^{-\frac12}_{-1}} - \alpha_{\nabla^{\frac12}_{2\,0}\nabla^{-\frac12}_{0}} + \alpha_{\nabla^{\frac12}_{2-1}\Delta^{-\frac12}_{1}} - \alpha_{\nabla^{\frac12}_{2-2}\nabla^{-\frac12}_{2}} = 0,$$

the L_{+1} equations

$$t^1_1\alpha_{\nabla^{\frac12}_{2\,0}\nabla^{-\frac12}_{0}} + t^1_0\alpha_{\nabla^{\frac12}_{2\,1}\nabla^{-\frac12}_{-1}} = 0,$$

$$t^1_2\alpha_{\nabla^{\frac12}_{1\,0}\nabla^{-\frac12}_{1}} + t^1_0\alpha_{\nabla^{\frac12}_{2-1}\nabla^{-\frac12}_{1}} + t^1_1\alpha_{\nabla^{\frac12}_{\ 0}\nabla^{-\frac12}_{0}} = 0,$$

$$t^1_2\alpha_{\nabla^{\frac12}_{1-1}\nabla^{-\frac12}_{2}} + t^1_{-1}\alpha_{\nabla^{\frac12}_{2-2}\nabla^{-\frac12}_{2}} + t^1_2\alpha_{\nabla^{\frac12}_{2-1}\nabla^{-\frac12}_{1}} = 0,$$

and the S_{+1} equation

$$\alpha_{\nabla^{\frac12}_{2\,1}\nabla^{-\frac12}_{-1}} - \alpha_{\nabla^{\frac12}_{2-1}\nabla^{-\frac12}_{1}} + \alpha_{\nabla^{\frac12}_{1-1}\nabla^{-\frac12}_{1}} = 0.$$

Since $t^1_1/t^1_2 = \sqrt{\frac32}$, the normalized solution to these equations is

$$(\tfrac{3}{28})^{\frac12}(\nabla^{\frac12}_{2\,1}\nabla^{-\frac12}_{-1} - \nabla^{\frac12}_{2\,0}\nabla^{-\frac12}_{0} - \tfrac13\nabla^{\frac12}_{2-1}\nabla^{-\frac12}_{1} + \tfrac53\nabla^{\frac12}_{2-2}\nabla^{-\frac12}_{2} - \tfrac43\nabla^{\frac12}_{1-1}\nabla^{\frac12}_{2} + \tfrac83\nabla^{\frac12}_{1\,0}\nabla^{-\frac12}_{1}).$$

One point to note: for the $|\{22\}[22]L\rangle$ states, the S_{+1} condition becomes

$$\alpha_{\nabla^{\frac12}_{a\,b}\nabla^{-\frac12}_{c\,a}} - \alpha_{\nabla^{\frac12}_{a\,c}\nabla^{-\frac12}_{b\,d}} = 0,$$

i.e.,

$$\alpha_{\nabla^{\frac12}_{a\,b}\nabla^{-\frac12}_{c\,a}} = \alpha_{\nabla^{\frac12}_{c\,a}\nabla^{-\frac12}_{ab}},$$

which gives us, with the exception of $L = 0$ (it has other S_{+1} equations as well), nearly half the number of unknowns.

The states of the form $|\{2^c1^d\}[2^c1^d]L\rangle$ are found tabulated in Table I.

VII. CONCLUSIONS

In this paper, the atomic states $|\{2^a1^b\}[2^c1^d]L\rangle$ for maximum M_L and M_S and for $2a + b \leq 2l + 1$ have been constructed in the scheme $U_{2l+1} \supset R_{2l+1} \supset R_3$. Clearly, states with $2a + b > 2l + 1$ can be derived by using particle hole equivalence[9,10]; in terms of the notation of this paper, one would simply place a bar over both $\nabla^{\frac12}$ and $\nabla^{-\frac12}$ to derive the hole state equivalent to a given particle state. The form of the states in this paper imply a phase convention, for the state $\nabla^{\frac12}_{321}\,\nabla^{-\frac12}_{-2-1}$ could equally well be written as $\nabla^{\frac12}_{321}\,\nabla^{-\frac12}_{-1-2}$ (any other permutation of m_l indices is clearly equivalent to one of these two states). This matter of phase has not been pursued in this paper, but more care may have to be taken if Clebsch–Gordan coefficients and isoscalar factors are to be derived using this basis as a starting point. Defining semiconjugacy makes it necessary to calculate only the states $|\{2^c1^d\}[2^c1^d]L\rangle$; for these states, however, no general closed algebraic expression could be found. In fact, in view of the immense combinatorial complexity of this problem, it seems doubtful that such an expression exists.

ACKNOWLEDGMENT

The author would like to thank Professor B. G. Wybourne for his continued interest and helpful comments.

* Research sponsored in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under AFOSR Grant No. 1275-67.

[1] J. Schwinger, Technical Information Service, Oak Ridge, Tennessee, Report W-23091, No. NYO-3071, 1952, published in *Quantum Theory of Angular Momentum*, L. C. Biedenharn and H. Van Dam, Eds. (Academic, New York, 1965).

[2] V. Bargmann, Commun. Pure Appl. Math. **14**, 198 (1961).

[3] M. Moshinsky, Nucl. Phys. **31**, 384 (1962); Rev. Mod. Phys. **34**, 813 (1962); J. Math. Phys. **4**, 1128 (1963).

[4] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 436, 1449 (1963); **5**, 1723, 1730 (1964).

[5] M. Moshinsky and V. S. Devi, J. Math. Phys. **10**, 455 (1969).

[6] M. Moshinsky, "Group Theory and the Many-Body Problem" in *Many-Body Problems and Other Selected Topics in Theoretical Physics*, M. Moshinsky, T. A. Brody, and G. Jacob, Eds. (Gordon and Breach, New York, 1966).

[7] J. Flores, E. Chacon, P. A. Mello, and M. De Llano, "Studies on Nuclear Structure in the 2*s* — 1*d* Shell, I and II", Ref. 6.

[8] H. A. Jahn, Proc. Roy. Soc. (London) **A205**, 192 (1951).

[9] G. Racah, Phys. Rev. **76**, 1352 (1949).

[10] B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).

[11] B. R. Judd, *Second Quantization and Atomic Spectroscopy* (John Hopkins, Baltimore, 1967).

[12] D. E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford U.P., London, 1940).

[13] F. D. Murnaghan, *The Unitary and Rotation Groups* (Spartan, Washington, D.C., 1962).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Some Topological Properties Connected to the Parametrized Feynman Amplitudes

C. J. LIU AND YUTZE CHOW
Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201

(Received 2 August 1968; Revised Manuscript Received 3 September 1969)

An approach first developed by one of the present authors for computation involving Poincaré incidence matrix is now carried over to the so-called "loop-matrix" formalism. A matrix $\overleftrightarrow{\tau}^{\{v\}}_{\{l\}}$ is introduced to prove some important properties of transformations among different sets of basic loops. Several properties in the form of lemmas and theorems are presented here. In particular, a formula of very concise form is found for the total number of possible tree graphs of a given Feynman graph in terms of its corresponding Poincaré incidence matrix. Also derived here is the formula for the total number of tree graphs in terms of any loop matrix. Their proofs show that the previously developed approach for Poincaré incidence matrix can be nicely generalized to a surprising extent to the loop matrix formalism and thus demonstrate the *duality* between the two formalisms.

I. INTRODUCTION

In dealing with a Feynman amplitude in its parametrized form, some important properties can often be extracted from its corresponding Feynman graph by means of the topological properties imbedded in the graph. In this paper, we present some results of our investigations of topological properties by means of the *loop matrix*[1,2] and some applications of these results to the U function and the V function.[2,3]

Our notation here essentially follows that of Ref. 2 unless otherwise specified. Given a Feynman graph containing N vertices and J internal lines, a set of *basic momenta* $\{k_l\}$ can be introduced; each basic momentum flows along a loop (or closed path) due to conservation. Whenever $\{k_l\}$ is given, a set of independent *basic loops* $\{l\}$ is uniquely defined; each basic momentum is required to belong to one and only one basic loop. Hereafter, we refer to the set $\{k_l\}$ as a *basic momentum set*, and the set $\{l\}$ as a *basic loop set*. Now, if we denote the jth internal momentum (i.e., the momentum of the jth internal line) by q_j and the set of the external momenta[1,2,4] by $\{p_i\}$, then we can write q_j into a linear combination of $\{k_l\}$ and $\{p_i\}$:

$$\vec{q} = \overleftrightarrow{\varepsilon} \cdot \vec{k} + \overleftrightarrow{\lambda} \cdot \vec{p}$$

and

$$q_j = K_j + P_j,$$

where $\overleftrightarrow{\varepsilon}$ is the so-called *loop matrix* with its entries defined by

$\varepsilon_{jl} = 0$, if line j does not belong to loop l,
$\varepsilon_{jl} = 1$, if line j belongs to loop l and if j is parallel to l,
$\varepsilon_{jl} = -1$, if j line belongs to loop l and if j is antiparellel to l

and where λ_{ji} can take any arbitrary real value provided momentum conservation at every vertex is satisfied. We note that the loop matrix $\overleftrightarrow{\varepsilon}$ is not uniquely defined for a given Feynman graph since the choice of the basic momentum set $\{k_l\}$ is not unique. However, once $\{k_l\}$ is chosen, then the *corresponding* $\overleftrightarrow{\varepsilon}$ is uniquely determined. The different choices of a basic momentum set $\{k_l\}$ correspond to the different possible ways of assigning, if topologically admissible,

$$\{k_1, \cdots, k_L\} = \{q_{\nu_1}, \cdots, q_{\nu_L}\},$$

where $\{\nu_1, \cdots, \nu_L\}$ is a proper subset of $\{1, \cdots, J\}$ and L is the total number of independent loops. It is well known that N, J, and L satisfy the condition

$$L = J - N + 1,$$

where N is the total number of vertices involved in the given graph.

Remark: If $\{\nu_1, \cdots, \nu_L\}$ is a *basic momentum set* and $\{1, \cdots, L\}$ is the *corresponding basic loop set* determined by $\{\nu_1, \cdots, \nu_L\}$, then it is necessary that

$$\varepsilon_{\nu_j l} = \pm\delta_{jl}, \quad l = 1, \cdots, L, j = 1, \cdots, J.$$

For any given Feynman amplitude, there are different parametrized forms of Feynman amplitude due to different ways of parametrization. However, here we concern ourselves with the following parametrized form of Feynman amplitude:

$$F = \text{const} \int_0^1 \cdots \int_0^1 \prod_{j=1}^{J} d\alpha_j \frac{\delta\left(1 - \sum_{j=1}^{L} \alpha_j\right)}{U^2(V - i0)^{J-2L}},$$

with

$$U \equiv \det \overleftrightarrow{A}$$

and

$$\overleftrightarrow{A} \equiv \overleftrightarrow{\varepsilon}^{\mathrm{T}}\overleftrightarrow{\alpha}\overleftrightarrow{\varepsilon},$$

where α is a diagonal matrix of order J, defined by

$$\alpha_{ij} \equiv \alpha_j\delta_{ij}, \forall j.$$

The V function takes the form

$$V \equiv \sum_{j=1}^{J}\alpha_j(m_j^2 + P_j^2) - \vec{b}^{\mathrm{T}}\overleftrightarrow{A}^{-1}\vec{b},$$

where

$$\vec{b} \equiv \overleftrightarrow{\varepsilon}^{\mathrm{T}}\overleftrightarrow{\alpha}\vec{P}.$$

II. SOME PROPERTIES OF MATRIX $\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}}$

In the study of a Feynman amplitude by means of the Poincaré incidence matrix $\overleftrightarrow{\epsilon}$, one of us has introduced the matrix[2]

$$\overleftrightarrow{\sigma}_{\{\mu\}} \equiv \begin{pmatrix} \epsilon_{\mu_1 1}\epsilon_{\mu_1 1} & \epsilon_{\mu_1 1}\epsilon_{\mu_1 2} & \cdots & \epsilon_{\mu_1 1}\epsilon_{\mu_1 N-1} \\ \epsilon_{\mu_2 2}\epsilon_{\mu_2 1} & \epsilon_{\mu_2 2}\epsilon_{\mu_2 2} & \cdots & \epsilon_{\mu_2 2}\epsilon_{\mu_2 N-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \epsilon_{\mu_{N-1} N-1}\epsilon_{\mu_{N-1} 1} & \epsilon_{\mu_{N-1} N-1}\epsilon_{\mu_{N-1} 2} & \cdots & \epsilon_{\mu_{N-1} N-1}\epsilon_{\mu_{N-1} N-1} \end{pmatrix}$$

whose determinant has been shown to be of considerable interest in connection with the so-called "tree set" (i.e., the set of all tree graphs) as discussed in Ref. 2.

From a *kind of "duality"* (since it is not really *duality* in the strict sense), we are led intuitively to define a similar matrix for a set of basic loops, namely,

$$\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}} \equiv \begin{pmatrix} \varepsilon_{\nu_1 1}\varepsilon_{\nu_1 1} & \varepsilon_{\nu_1 1}\varepsilon_{\nu_1 2} & \cdots & \varepsilon_{\nu_1 1}\varepsilon_{\nu_1 L} \\ \varepsilon_{\nu_2 2}\varepsilon_{\nu_2 1} & \varepsilon_{\nu_2 2}\varepsilon_{\nu_2 2} & \cdots & \varepsilon_{\nu_2 2}\varepsilon_{\nu_2 L} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \varepsilon_{\nu_L L}\varepsilon_{\nu_L 1} & \varepsilon_{\nu_L L}\varepsilon_{\nu_L 2} & \cdots & \varepsilon_{\nu_L L}\varepsilon_{\nu_L L} \end{pmatrix}, \quad (1)$$

where $\{\nu\} = \{\nu_1, \cdots, \nu_L\}$ is an *arbitrary* set of distinct L internal lines. The second subscript of $\overleftrightarrow{\varepsilon}$ refers to the different members of an *arbitrary* set of basic loops which are labeled by $\{l\} = \{1, \cdots, L\}$.

Indeed, the introduction of this matrix $\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}}$ as guided purely in the beginning by our intuition can be seen to be very useful. In this section, we discuss two of the interesting properties of $\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}}$ which are then used to prove two theorems in the next section.

For convenience in our subsequent discussions we introduce here a matrix for a given *arbitrary* set of distinct internal lines $\{\nu\}$ and an *arbitrary set of basic loops* $\{l\}$:

$$\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}} \equiv \begin{pmatrix} \varepsilon_{\nu_1 1} & \varepsilon_{\nu_1 2} & \cdots & \varepsilon_{\nu_1 L} \\ \varepsilon_{\nu_2 1} & \varepsilon_{\nu_2 2} & \cdots & \varepsilon_{\nu_2 L} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \varepsilon_{\nu_L 1} & \varepsilon_{\nu_L 2} & \cdots & \varepsilon_{\nu_L L} \end{pmatrix}. \quad (2)$$

Lemma 1: In defining the matrix $\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}$, if the set $\{\nu\}$ of L distinct internal lines happens to be a *basic momentum set*, then

$$\det \overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}} = \pm 1. \quad (3)$$

Proof: Let $\{\tilde{l}\}$ be the *corresponding* basic loop set determined by $\{\tilde{\nu}\}$. It is trivial that

$$\det \overleftrightarrow{\Theta}_{\{\tilde{l}\}}^{\{\tilde{\nu}\}} = \pm 1, \quad (4)$$

since each diagonal element in $\overleftrightarrow{\Theta}_{\{\tilde{l}\}}^{\{\tilde{\nu}\}}$ is ± 1 and each off-diagonal element in $\overleftrightarrow{\Theta}_{\{\tilde{l}\}}^{\{\tilde{\nu}\}}$ is zero.

Since the transformation between any two sets of basic loops is linear, we can carry out the transformation $\{\tilde{l}\} \to \{l\}$ in $n + 1$ steps:

$$\{\tilde{l}\} \to \{l_1\} \to \{l_2\} \to \cdots \to \{l_n\} \to \{l\},$$

such that in each step, e.g., $\{l_i\} \to \{l_{i+1}\}$, the elements in $\{l_{i+1}\}$ are obtained from $\{l_i\}$ by one of the three possible processes:

(i) Relabel some (or all) of the elements in $\{l_i\}$.

(ii) Choose two elements (i.e., two loops) a and b in $\{l_i\}$. Add (or subtract) them if loops a and b have different (or the same) senses in the overlapping path, to form a new element. Replace a or b by the new loop.

(iii) *Reverse* the sense of some (or all) of the loops in $\{l_i\}$.

These steps can only result in the following corresponding changes in $\overleftrightarrow{\Theta}$: $\overleftrightarrow{\Theta}_{\{l_i\}}^{\{\tilde{\nu}\}}$ differs from $\overleftrightarrow{\Theta}_{\{l_{i+1}\}}^{\{\tilde{\nu}\}}$ only in:

(1) exchange of columns;

(2) addition (or subtraction) of columns like

$$\begin{pmatrix} a & a' \\ b & b' \end{pmatrix} \to \begin{pmatrix} a + a' & a' \\ b + b' & b' \end{pmatrix};$$

(3) change of signs of all the elements in the same column; this can result in only a change of sign for the determinant.

Consequently, we have

$$\det \overleftrightarrow{\Theta}_{\{l_i\}}^{\{\tilde{\gamma}\}} = \pm\det \overleftrightarrow{\Theta}_{\{l_{i+1}\}}^{\{\tilde{\gamma}\}}. \tag{5}$$

Therefore,

$$\det \overleftrightarrow{\Theta}_{\{\bar{l}\}}^{\{\tilde{\gamma}\}} = \pm\det \overleftrightarrow{\Theta}_{\{l\}}^{\{\tilde{\gamma}\}}. \tag{6}$$

Finally, (3) follows from (4) and (6). QED

Lemma 2: Let $\overleftrightarrow{X}$ be the transformation on $\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}$ induced by a transformation connecting two sets of basic loops; then

$$\det \overleftrightarrow{X} = \pm 1. \tag{7}$$

Proof: When the basic loop set $\{l\}$ is transformed into another basic loop set $\{\bar{l}\}$, i.e.,

$$\{l\} \to \{\bar{l}\},$$

we have the induced transformation matrix $\overleftrightarrow{X}$,

$$\overleftrightarrow{X}: \overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}} \to \overleftrightarrow{\Theta}_{\{\bar{l}\}}^{\{\nu\}} \equiv \overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}\overleftrightarrow{X}, \tag{8}$$

where X is an $L \times L$ matrix and is uniquely defined whether $\{\nu\}$ is a *basic momentum set* or not.

But, from Lemma 1, we have

$$\det \overleftrightarrow{\Theta}_{\{\bar{l}\}}^{\{\nu\}} = \pm\det \overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}},$$

if $\{\nu\}$ is a basic momentum set. Therefore,

$$\det X = \pm 1. \qquad \text{QED}$$

Theorem 1: In defining the matrix $\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}}$, if the set $\{\nu\}$ of L distinct internal lines is a *basic momentum set*, then

$$\sum_{\forall\{\nu\}} \det \overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}} = +1, \tag{9}$$

where the summation with respect to $\{\nu\}$ is taken over all the possible permutations of the given basic momentum set. (It should be emphasized that in defining $\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}}$, the set $\{l\}$ is not in general the *corresponding* basic loop set determined by $\{\nu\}$.)

Proof: Let us consider $(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}})$. The ll'th matrix element of $(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}})$ is simply

$$(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}})_{ll'} = \sum_{j=1}^{L} \varepsilon_{jl}\varepsilon_{jl'}; \tag{10}$$

therefore,

$$\det(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}) = \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \sum_{j=1}^{J} \varepsilon_{j\gamma_l}\varepsilon_{jl}, \tag{11}$$

where $\{\gamma\} = \{\gamma_1, \cdots, \gamma_L\}$ are all the possible sets $\mathbb{P}\{1, \cdots, L\}$, P = permutation, $C_{\{\gamma\}} = (-)^{P(\gamma)}$, and P is the number of transpositions involved in $\mathbb{P}$. The summation sign and the product sign of (11) can be exchanged by properly adjusting some of the subscripts and range of summation:

$$\det(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}) = \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \sum_{\forall\{\rho\}} \prod_{l=1}^{L} \varepsilon_{\rho_l\gamma_l}\varepsilon_{\rho_l l}, \tag{12}$$

where $\{\rho\} = \{\rho_1, \cdots, \rho_L\}$ with $\rho_l = 1, \cdots, L$. We note that, *in principle*, any kind of repetition of elements is allowed for $\{\rho\}$. However, it will be shown that there is actually no repetition of elements in $\{\rho\}$.

First, let us rewrite (12) into the form

$$\begin{aligned}\det(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}) &= \sum_{\forall\{\rho\}} \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \varepsilon_{\rho_l\gamma_l}\varepsilon_{\rho_l l} \\ &= \sum_{\forall\{\rho\}} \left[\left(\prod_{l=1}^{L} \varepsilon_{\rho_l l}\right)\left(\sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \varepsilon_{\rho_l\gamma_l}\right)\right].\end{aligned} \tag{13}$$

If there is a repetition in $\{\rho\}$, say

$$\rho_{l'} = \rho_{l''},$$

then, in the summation

$$\sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \varepsilon_{\rho_l\gamma_l},$$

there are two terms

$$C_{\{\gamma'\}} \prod_{l=1}^{L} \varepsilon_{\rho_l\gamma_l'} \quad \text{and} \quad C_{\{\gamma''\}} \prod_{l=1}^{L} \varepsilon_{\rho_l\gamma_l''}$$

which cancel each other exactly because

$$\{\gamma'\} = \mathbb{P}_{l'l''}\{\gamma''\},$$

where $\mathbb{P}_{l'l''}$ is the transposition operator for l' and l''. Further, any repetition of elements in $\{\rho\}$ more than twice is clearly impossible, since the same argument applies to successive transpositions. Therefore, we conclude that there is actually no repetition of elements in $\{\rho\}$, i.e., $\{\rho\}$ is just $\{\nu\}$, differing only by some *permutation* of elements. Thus, (13) can be written in the form

$$\begin{aligned}\det(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}) &= \sum_{\forall\{\rho\}} \left(\sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \varepsilon_{\rho_l l}\varepsilon_{\rho_l\gamma_l}\right) \\ &= \sum_{\forall\{\rho\}} \det \overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}},\end{aligned} \tag{14}$$

where $\forall\{\rho\}$ under the summation sign means the sum over all the possible $\{\rho\}$ obtained by different permutations of $\{\rho\}$. The last step follows from the definition of $\det \overleftrightarrow{\tau}_{\{l\}}^{\{\rho\}}$.

However, since the $\{\rho\}$ are just the $\{\nu\}$ which differ from each other only by some permutation of elements and, using Lemma 1, we have

$$\det(\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}\mathrm{T}}\overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}) = +1.$$

We have, therefore, completed the proof. QED

Theorem 2: In dealing with the matrix $\overleftrightarrow{\Theta}^{\{\nu\}}_{\{l\}}$, if the set $\{\nu\}$ of L distinct internal lines is not a *basic momentum set*, then it is necessary that

$$\det \overleftrightarrow{\Theta}^{\{\nu\}}_{\{l\}} = 0. \tag{15}$$

Proof: We use mathematical induction on the number L. Let the theorem be true for the case of

$$\{l\} = \{1, \cdots, L\}$$

and

$$\{\nu\} = \{\nu_1, \cdots, \nu_L\}.$$

We want to show that the theorem is also true for the case of

$$\{l\} = \{1, \cdots, L, L+1\}$$

and

$$\{\nu\} = \{\nu_1, \cdots, \nu_L, \nu_{L+1}\};$$

that is the case of $L + 1$ basic loops for the set $\{l\}$ and $L + 1$ distinct internal lines for the set $\{\nu\}$. By Lemma 2, we can transform $\{l\}$ into a new set $\{\tilde{l}\}$ in such a way that there is at least one internal line, say ν_i, satisfying

$$\nu_i \in \tilde{i} \quad \text{and} \quad \nu_i \notin \tilde{j} \quad \text{for} \quad \tilde{j} \neq \tilde{i}, \quad \tilde{j}, \tilde{i} \in \{\tilde{l}\}.$$

This leads to

$$\det \overleftrightarrow{\Theta}^{\{\nu\}}_{\{l\}} = \pm \det \overleftrightarrow{\Theta}^{\{\nu\}}_{\{\tilde{l}\}} = \pm \begin{vmatrix} \varepsilon_{\nu_1\tilde{1}} & \cdots & \varepsilon_{\nu_1\widetilde{i-1}} & \varepsilon_{\nu_1\tilde{i}} & \varepsilon_{\nu_1\widetilde{i+1}} & \cdots & \varepsilon_{\nu_1\widetilde{L+1}} \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \varepsilon_{\nu_{i-1}\tilde{1}} & \cdots & \varepsilon_{\nu_{i-1}\widetilde{i-1}} & \varepsilon_{\nu_{i-1}\tilde{i}} & \varepsilon_{\nu_{i-1}\widetilde{i+1}} & \cdots & \varepsilon_{\nu_{i-1}\widetilde{L+1}} \\ 0 & \cdots & 0 & \varepsilon_{\nu_i\tilde{i}} & 0 & \cdots & 0 \\ \varepsilon_{\nu_{i+1}\tilde{1}} & \cdots & \varepsilon_{\nu_{i+1}\widetilde{i-1}} & \varepsilon_{\nu_{i+1}\tilde{i}} & \varepsilon_{\nu_{i+1}\widetilde{i+1}} & \cdots & \varepsilon_{\nu_{i+1}\widetilde{L+1}} \\ \varepsilon_{\nu_{L+1}\tilde{1}} & \cdots & \varepsilon_{\nu_{L+1}\widetilde{i-1}} & \varepsilon_{\nu_{L+1}\tilde{i}} & \varepsilon_{\nu_{L+1}\widetilde{i+1}} & \cdots & \varepsilon_{\nu_{L+1}\widetilde{L+1}} \end{vmatrix} \tag{16}$$

$$= \pm \begin{vmatrix} \varepsilon_{\nu_1\tilde{1}} & \cdots & \varepsilon_{\nu_1\widetilde{i-1}} & \varepsilon_{\nu_1\widetilde{i+1}} & \cdots & \varepsilon_{\nu_1\widetilde{L+1}} \\ \varepsilon_{\nu_{i-1}\tilde{1}} & \cdots & \varepsilon_{\nu_{i-1}\widetilde{i-1}} & \varepsilon_{\nu_{i-1}\widetilde{i+1}} & \cdots & \varepsilon_{\nu_{i-1}\widetilde{L+1}} \\ \varepsilon_{\nu_{i+1}\tilde{1}} & \cdots & \varepsilon_{\nu_{i+1}\widetilde{i-1}} & \varepsilon_{\nu_{i+1}\widetilde{i+1}} & \cdots & \varepsilon_{\nu_{i+1}\widetilde{L+1}} \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & & \cdot \\ \varepsilon_{\nu_{L+1}\tilde{1}} & \cdots & \varepsilon_{\nu_{L+1}\widetilde{i-1}} & \varepsilon_{\nu_{L+1}\widetilde{i+1}} & \cdots & \varepsilon_{\nu_{L+1}\widetilde{L+1}} \end{vmatrix} \tag{17}$$

$$= 0, \tag{18}$$

since the right-hand side of (17) only depends on L loops and L internal lines, if $\{\nu_1, \cdots, \nu_{i-1}, \nu_{i+1}, \cdots, \nu_{L+1}\}$ is a *nonbasic momentum set* in the subgraph. Further, one can show that $\{\nu_1, \cdots, \nu_{i-1}, \nu_{i+1}, \cdots, \nu_{L+1}\}$ should be a *nonbasic momentum set* in the subgraph. This is discussed in full in Appendix A. Then the mathematical induction is completed. QED

Remark: Given a graph G, the total number of independent loops L is determined uniquely. Therefore, if a mathematical induction on L is used, we have to consider all the subgraph of G. For a strongly connected graph (i.e., after any single internal line is deleted the subgraph remains connected), the deletion of an internal line results in the necessary removal of a loop. Consequently, the total number of independent loops for the subgraph has to be one less than before. However, for a graph which is not strongly connected, such as Fig. 1, then the deletion of line a yields the new graph shown in Fig. 2. In this case, the total number of the independent loops in the *subgraph* is equal to the total number of the independent loops in the *original* graph. It appears, at first, that the mathematical induction may not be applicable in this case. However, this apparent difficulty can be removed by the following considerations:

By definition of $\overleftrightarrow{\Theta}^{\{\nu\}}_{\{l\}}$, we can easily see that, if $a \in \{\nu\}$, in Fig. 1, it follows immediately that

$$\det \overleftrightarrow{\Theta}^{\{\nu\}}_{\{l\}} = 0,$$

since

$$\varepsilon_{al} = 0, \quad \forall\, l.$$

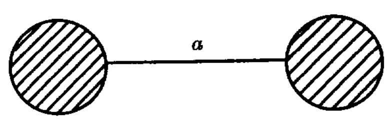


FIG. 1. A graph which is not strongly connected.

FIG. 2. A subgraph of Fig. 1.

On the other hand, if any internal line in $\{\nu_1, \cdots, \nu_{L+1}\}$ does not behave like the line a in Fig. 1, the mathematical induction is obviously applicable.

Corollary: For the $\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}}$ matrix, if the set $\{\nu\}$ of L distinct internal lines is not a *basic momentum set*, then it is necessary that

$$\det \overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}} = 0. \tag{19}$$

Proof:

$$\overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}} = \begin{pmatrix} \varepsilon_{\nu_1 1}\varepsilon_{\nu_1 1} & \varepsilon_{\nu_1 1}\varepsilon_{\nu_1 2} & \cdots & \varepsilon_{\nu_1 1}\varepsilon_{\nu_1 L} \\ \varepsilon_{\nu_2 2}\varepsilon_{\nu_2 1} & \varepsilon_{\nu_2 2}\varepsilon_{\nu_2 2} & \cdots & \varepsilon_{\nu_2 2}\varepsilon_{\nu_2 L} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \varepsilon_{\nu_L L}\varepsilon_{\nu_L 1} & \varepsilon_{\nu_L L}\varepsilon_{\nu_L 1} & \cdots & \varepsilon_{\nu_L L}\varepsilon_{\nu_L L} \end{pmatrix}$$

$$= \begin{pmatrix} \varepsilon_{\nu_1 1} & & & & \\ & \varepsilon_{\nu_2 2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & \varepsilon_{\nu_L L} \end{pmatrix} \times \begin{pmatrix} \varepsilon_{\nu_1 1} & \varepsilon_{\nu_1 2} & \cdots & \varepsilon_{\nu_1 L} \\ \varepsilon_{\nu_2 1} & \varepsilon_{\nu_2 2} & \cdots & \varepsilon_{\nu_2 L} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \varepsilon_{\nu_L 1} & \varepsilon_{\nu_L 2} & \cdots & \varepsilon_{\nu_L L} \end{pmatrix}$$

$$= \begin{pmatrix} \varepsilon_{\nu_1 1} & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \varepsilon_{\nu_L L} \end{pmatrix} \overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}}. \tag{20}$$

By Theorem 2, we have

$$\det \overleftrightarrow{\tau}_{\{l\}}^{\{\nu\}} = (\varepsilon_{\nu_1 1} \cdots \varepsilon_{\nu_L L}) \det \overleftrightarrow{\Theta}_{\{l\}}^{\{\nu\}} = 0.$$

III. SOME PROPERTIES OF THE MATRIX $\overleftrightarrow{\varepsilon}$

Theorem 3: For any loop matrix $\overleftrightarrow{\varepsilon}$, $\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon})$ is independent of the particular choice of the basic loop set $\{l\}$ used to define $\overleftrightarrow{\varepsilon}$.

Proof: For any two given $\overleftrightarrow{\varepsilon}$ and $\overleftrightarrow{\varepsilon}'$ corresponding to the basic loop sets $\{l\}$ and $\{l'\}$, there is a transformation connecting $\overleftrightarrow{\varepsilon}$ and $\overleftrightarrow{\varepsilon}'$, namely,

$$\overleftrightarrow{X}: \overleftrightarrow{\varepsilon} \to \overleftrightarrow{\varepsilon}' = \overleftrightarrow{\varepsilon} \cdot \overleftrightarrow{X}, \tag{21}$$

where $\overleftrightarrow{X}$ is an $L \times L$ matrix to be multiplied to the *right* of $\overleftrightarrow{\varepsilon}$.

Consider now

$$\begin{aligned} \det(\overleftrightarrow{\varepsilon}'^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}') &= \det(\overleftrightarrow{X}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon} \cdot \overleftrightarrow{X}) \\ &= \det(\overleftrightarrow{X}^{\mathrm{T}} \cdot \overleftrightarrow{X}) \det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}) \\ &= \det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}), \end{aligned} \tag{22}$$

where Lemma 2 is used in the last step. QED

Theorem 4: For any loop matrix $\overleftrightarrow{\varepsilon}$, we have

$$\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}) = \text{total number of all possible choices of the basic momentum set.} \tag{23}$$

Proof: Since the ll'th matrix element of $(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon})$ is simply

$$(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon})_{ll'} = \sum_{j=1}^{J} \varepsilon_{jl}\varepsilon_{jl'}, \tag{24}$$

we have, therefore,

$$\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}) = \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \sum_{j=1}^{J} \varepsilon_{j\gamma_l}\varepsilon_{jl}, \tag{25}$$

where $\{\gamma\}$ and $C_{\{\gamma\}}$ have the same meaning as that in Theorem 1. The summation sign and the product sign of (25) can be exchanged by properly adjusting some of the subscripts and range of summation:

$$\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}) = \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \sum_{\forall\{\mu\}} \prod_{l=1}^{L} \varepsilon_{\mu_l}\varepsilon_{\mu_l l}, \tag{26}$$

where $\{\mu\} = \{\mu_1, \cdots, \mu_L\}$ with $\mu_l = 1, \cdots, J$. We know that, *in principle*, any kind of repetition of elements is allowed for $\{\mu\}$. However, it can be shown that there is actually no repetition of elements in $\{\mu\}$, in a manner exactly like that given in Theorem 1.

Next, (26) can be written in the form

$$\begin{aligned} \det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}) &= \sum_{\forall\{\mu\}} \left(\sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \varepsilon_{\mu_l l}\varepsilon_{\mu_l \gamma_l} \right) \\ &= \sum_{\{\mu\}} \det \tau_{\{l\}}^{\{\mu\}}. \end{aligned} \tag{27}$$

The last step follows from the definition of $\det \overleftrightarrow{\tau}_{\{l\}}^{\{\mu\}}$. At this stage, we can split the summation over $\{\mu\}$ into two summations $\sum'_{\forall\{\mu\}} + \sum''_{\forall\{\mu\}}$, where the single-primed summation is carried over all the possible *basic momentum sets* and all their distinct permutations. The double-primed summation is carried over all the sets that are not basic momentum sets (and their permutations).

From Theorem 1 and the Corollary to Theorem 2, we know that

$$\sum_{\forall\{\mu\}}{}' \det \tau^{\{\mu\}}_{\{l\}} = +1, \tag{28a}$$

if $\{\mu\}$ is a basic momentum set and the sum over here is only carried over all the distinct permutations of a particular basic momentum set, and

$$\sum_{\forall\{\mu\}}{}'' \det \overleftrightarrow{\tau}^{\{\mu\}}_{\{l\}} = 0, \tag{28b}$$

if $\{\mu\}$ is not a basic momentum set.

Therefore, the nonvanishing contributions to (27) come from (28a), or more precisely, $+1$ for each choice of basic momentum set. That is,

$\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon})$ = total number of all possible choices of the basic momentum set.

QED

Remark: For a given graph, the *tree set* is defined as the *complement* of the *corresponding* basic momentum set. This leads to

$\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon})$ = total number of all possible choices of the tree set. (29)

IV. SOME THEOREMS ON U FUNCTIONS AND V FUNCTIONS

The following theorem, Theorem 5, is not new, but the existing proof, to our knowledge, is rather tedious.[1] The proof given below following our present approach is most straightforward.

Theorem 5: The U function is independent of the particular choice of the basic loop set $\{l\}$, used to define $\overleftrightarrow{\varepsilon}$.

Proof: Let $\overleftrightarrow{X}$ transform $\overleftrightarrow{\varepsilon}$ into $\overleftrightarrow{\varepsilon}'$:

$$\overleftrightarrow{X}: \overleftrightarrow{\varepsilon} \to \overleftrightarrow{\varepsilon}' = \overleftrightarrow{\varepsilon} \cdot \overleftrightarrow{X}.$$

Then $\overleftrightarrow{A}$ transforms according to

$$\overleftrightarrow{X}: \overleftrightarrow{A} \to \overleftrightarrow{A}', \tag{30}$$

where

$$\overleftrightarrow{A} \equiv \overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\alpha}\overleftrightarrow{\varepsilon} \tag{31}$$

and

$$\overleftrightarrow{A}' \equiv \overleftrightarrow{\varepsilon}'^{\mathrm{T}} \cdot \overleftrightarrow{\alpha} \cdot \overleftrightarrow{\varepsilon}' \tag{32}$$

as defined before.

Since

$$\begin{aligned}\overleftrightarrow{A}' &= \overleftrightarrow{X}^{\mathrm{T}} \cdot \overleftrightarrow{\varepsilon}^{\mathrm{T}} \cdot \overleftrightarrow{\alpha} \cdot \overleftrightarrow{\varepsilon}\overleftrightarrow{X} \\ &= \overleftrightarrow{X}^{\mathrm{T}} \cdot \overleftrightarrow{A} \cdot \overleftrightarrow{X},\end{aligned} \tag{33}$$

we have, therefore,

$$\det \overleftrightarrow{A}' = (\det \overleftrightarrow{X})^2 \det \overleftrightarrow{A}. \tag{34}$$

By Lemma 1, (34) becomes

$$\det \overleftrightarrow{A}' = \det A. \tag{35}$$

QED

We note that, in the proof of Theorem 5, we have not used the diagonal properties of the matrix $\overleftrightarrow{\alpha}$. Actually, for any $L \times L$ matrix $\overleftrightarrow{\beta}$, the value of $\det(\overleftrightarrow{\varepsilon}^{\mathrm{T}}\overleftrightarrow{\beta}\overleftrightarrow{\varepsilon})$ is independent of our particular choice of the basic loop set in defining $\overleftrightarrow{\varepsilon}$.

The following well-known property can also be proved by our present approach; the emphasis is on the uniformity of the nature of our proof.

Theorem 6: For any Feynman graph,[1,3]

$$\det \overleftrightarrow{A} = \sum_{\forall\{\nu\}} \prod_{l=1}^{L} \alpha_{\nu_l}, \tag{36}$$

where $\{\nu\}$ is a basic momentum set.

Proof: By definition, we have

$$\det \overleftrightarrow{A} = \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} a_{\gamma_l l} \tag{37}$$

and

$$a_{ll'} \equiv \sum_{j=1}^{J} \alpha_j \varepsilon_{jl} \varepsilon_{jl'}, \tag{38}$$

where both $\{\gamma\}$ and $C_{\{\gamma\}}$ have the same meaning as that in Theorem 6. Substitution of (38) into (37) gives

$$\begin{aligned}\det \overleftrightarrow{A} &= \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \sum_{j=1}^{J} \alpha_j \varepsilon_{j\gamma_l} \varepsilon_{jl} \\ &= \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \sum_{\forall\{t\}} \prod_{l=1}^{L} \alpha_{t_l} \varepsilon_{t_l\gamma_l} \varepsilon_{t_l l} \\ &= \sum_{\forall\{t\}} \left[\left(\prod_{l=1}^{L} \alpha_{t_l}\right) \sum_{\forall\{\gamma\}} C_{\{\gamma\}} \prod_{l=1}^{L} \varepsilon_{t_l\gamma_l} \varepsilon_{t_l l}\right],\end{aligned} \tag{39}$$

where $\{t\} = \{t_1, \cdots, t_L\}$ with $t_l = 1, \cdots, J$. Following the arguments and steps as those in the proof of Theorem 6, we conclude that

$$\det \overleftrightarrow{A} = \sum_{\forall\{t\}} \prod_{l=1}^{L} \alpha_{t_l} \det \overleftrightarrow{\tau}^{\{t\}}_{\{l\}}. \tag{40}$$

In virtue of (29a) and (29b), we have immediately

$$\det \overleftrightarrow{A} = \sum_{\forall\{\nu\}} \prod_{l=1}^{L} \alpha_{\nu_l},$$

where $\{\nu\}$ is now a set of *basic momenta*. QED

Theorem 7: The V function, defined by

$$V \equiv \sum_{j=1}^{J} \alpha_j (m_j^2 + P_j^2) + \vec{P}^{\mathrm{T}} \overleftrightarrow{\alpha}\overleftrightarrow{\varepsilon}\overleftrightarrow{A}^{-1}\overleftrightarrow{\varepsilon}^{\mathrm{T}}\overleftrightarrow{\alpha}\vec{P}, \tag{41}$$

is independent of the particular choice of the basic loop set $\{l\}$ used to define $\overleftrightarrow{\varepsilon}$.

Proof: We note that both $\vec{P}$ and $\overleftrightarrow{\alpha}$ in (41) are independent of $\{l\}$. We need only consider $\overleftrightarrow{\varepsilon}A^{-1}\overleftrightarrow{\varepsilon}$. If there is a transformation connecting $\overleftrightarrow{\varepsilon}$ and $\overleftrightarrow{\varepsilon}'$,

$$\overleftrightarrow{X}:\overleftrightarrow{\varepsilon} \to \overleftrightarrow{\varepsilon}' = \overleftrightarrow{\varepsilon}\overleftrightarrow{X},$$

then

$$\overleftrightarrow{X}:\overleftrightarrow{A}^{-1} \to \overleftrightarrow{A}'^{-1}. \tag{42}$$

We know from Lemma 2 that the matrix $\overleftrightarrow{X}$ has an inverse. Therefore, by making use of (33), we have

$$\overleftrightarrow{A}'^{-1} = \overleftrightarrow{X}^{-1}\overleftrightarrow{A}^{-1}\overleftrightarrow{X}^{\mathrm{T}^{-1}}; \tag{43}$$

then

$$\begin{aligned}\overleftrightarrow{\varepsilon}\overleftrightarrow{A}^{-1}\overleftrightarrow{\varepsilon}^{\mathrm{T}} &= \overleftrightarrow{\varepsilon}\overleftrightarrow{X}\overleftrightarrow{X}^{-1}\overleftrightarrow{A}^{-1}\overleftrightarrow{X}^{\mathrm{T}^{-1}}\overleftrightarrow{X}^{\mathrm{T}}\overleftrightarrow{\varepsilon}^{\mathrm{T}} \\ &= (\overleftrightarrow{\varepsilon}\overleftrightarrow{X})\overleftrightarrow{X}^{-1}\overleftrightarrow{A}^{-1}\overleftrightarrow{X}^{\mathrm{T}^{-1}}(\overleftrightarrow{\varepsilon}\overleftrightarrow{A})^{\mathrm{T}} \\ &= \overleftrightarrow{\varepsilon}'\overleftrightarrow{A}'^{-1}\overleftrightarrow{\varepsilon}'^{\mathrm{T}}.\end{aligned} \tag{44}$$

APPENDIX A

We want to show that the subset $\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}$ of (17) should be a *nonbasic momentum set* in the subgraph of the original graph G. This can be proved by contradiction.

If $\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}$ is a *basic momentum set* in the subgraph of G, then we can find the *corresponding basic loop set*, $\{\hat{1}, \cdots, \widehat{i-1}, \widehat{i+1}, \cdots, \widehat{L+1}\}$, where each element of $\{\hat{1}, \cdots, \widehat{i-1}\, \widehat{i+1}, \cdots, \widehat{L+1}\}$ is a linear combination of elements in $\{\tilde{1}, \cdots, \widetilde{i-1}, \widetilde{i+1}, \cdots, \widetilde{L+1}\}$ which is the basic loop set corresponding to $\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}$. It follows that

$$v_j \in \delta_{jk}\hat{k}, \quad j \neq i, \tag{A1}$$

since v_i was deleted. In the subgraph of G, the complement of $\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}$, denoted by

$$\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}_c, \tag{A2}$$

has the following property:

(number of vertices of the subgraph of G) $-$ 1
$=$ (number of internal lines of the subgraph of G), (A3)

i.e., $\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}_c$ in the subgraph of G is a *tree set*.

Now let us look at the original graph G. $\{v_1, \cdots, v_{i-1}, v_i, v_{i+1}, \cdots, v_{L+1}\}_c$ in G is the same as $\{v_1, \cdots, v_{i-1}, v_{i+1}, \cdots, v_{L+1}\}_c$ in the subgraph G since the subgraph of G is obtained from G by the deletion of v_i. Therefore, $\{v_1, \cdots, v_{i-1}, v_i, v_{i+1}, \cdots, v_{L+1}\}$ in G is a *tree set*. That means $\{v_1, \cdots, v_{i-1}, v_i, v_{i+1}, \cdots, v_{L+1}\}$ in G is *basic momentum set*. This is a contradiction.

Chow and Kleitman[3] have derived a systematic way of calculating the U function for a given Feynman graph. Their method consists of reducing the graph by shrinking all possible paths connecting any two arbitrary vertices. By repeating this, they obtained graphs, each of which has only a single vertex. This result can be easily verified by our expression (40).

By (14) and (40), we have

$$\sum_{\forall\{v\}} \prod_{l=1}^{L} \alpha_{v_l} \det \overleftrightarrow{\tau}_{\{l\}}^{\{v\}} = \det (\overleftrightarrow{\Theta}_{\{l\}}^{\{v\}\mathrm{T}}\overleftrightarrow{\alpha}^{\{v\}}\overleftrightarrow{\Theta}_{\{l\}}^{\{v\}}), \tag{A4}$$

where

$$\overleftrightarrow{\alpha}^{\{v\}} \equiv \begin{pmatrix} \alpha_{v_1} & & & \mathrm{O} \\ & \cdot & & \\ & & \cdot & \\ \mathrm{O} & & & \alpha_{v_L} \end{pmatrix}. \tag{A5}$$

However, $\det (\overleftrightarrow{\Theta}_{\{l\}}^{\{v\}\mathrm{T}}\overleftrightarrow{\alpha}^{\{v\}}\overleftrightarrow{\Theta}_{\{l\}}^{\{v\}})$ is the U function of L internal lines and L loops, i.e., there is only one vertex for the graph. For any graph, the U function can be written as

$$U = \sum_{\forall s} U_s, \tag{A6}$$

where

$$U_s \equiv \prod_{j=1}^{L} \alpha_{s_j}, \quad \forall\, s, \tag{A7}$$

is the U function for a single-vertex graph G_s containing L lines and the sum is over all such graphs obtainable from G by choosing different basic momentum sets.

In the Introduction, the V function is expressed in terms of internal momenta explicitly using loop matrices. However, we can write it in terms of external momenta explicitly using incidence matrices[1,2,4-7]

$$V = \sum_{j=1}^{J} \alpha_j m_j^2 + \sum_{n,n'=1}^{N-1} (\overleftrightarrow{h}^{-1})_{nn'} p_n p_{n'}, \tag{A8}$$

with

$$\overleftrightarrow{h} = \overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\tilde{\alpha}}\overleftrightarrow{\epsilon}^{\natural}, \tag{A9}$$

$$\tilde{\alpha}_{ij} \equiv \alpha_j^{-1}\delta_{ij}, \tag{A10}$$

$$\overleftrightarrow{\epsilon}^{\natural} \equiv \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1N-1} \\ \vdots & & \vdots \\ \epsilon_{J1} & \cdots & \epsilon_{JN-1} \end{pmatrix}, \tag{A11}$$

where

$\epsilon_{jn} = 0$, if the internal line j does not initiate or terminate at the vertex n,

$\epsilon_{jn} = +1$, if the internal line j initiates from the vertex n,

$\epsilon_{jn} = -1$, if the internal line j terminates at the vertex n.

We would like to show that V function, in terms of external momentum explicitly, is independent of the particular choice of the $N-1$ vertices out of the total N vertices in the graph. For different choice of $N-1$ vertices, $\overleftrightarrow{\epsilon}^{\natural}$ and h are transformed according to

$$\overleftrightarrow{Y}:\overleftrightarrow{\epsilon}^{\natural} \to \overleftrightarrow{\epsilon}^{\natural\prime} \equiv \overleftrightarrow{\epsilon}^{\natural}\overleftrightarrow{Y} \tag{A12}$$

and

$$\begin{aligned}\overleftrightarrow{Y}:\overleftrightarrow{n} \to \overleftrightarrow{h}' &\equiv \overleftrightarrow{\epsilon}^{\natural\prime\mathrm{T}}\overleftrightarrow{\alpha}\overleftrightarrow{\epsilon}^{\natural\prime}\\ &= \overleftrightarrow{Y}^{\mathrm{T}}\overleftrightarrow{h}\overleftrightarrow{Y},\end{aligned} \tag{A13}$$

where $\overleftrightarrow{Y}$ is a $(N-1)\times(N-1)$ matrix. For a given graph, a 1-to-1 correspondence exists between the incidence matrix; therefore, the matrix $\overleftrightarrow{Y}$ has an inverse. (Since $\det\overleftrightarrow{h}$ is independent of the particular choice of the $N-1$ vertices,[2,7] it follows immediately that $\det\overleftrightarrow{Y} = \pm 1$.) By making use of (A13), we have

$$\overleftrightarrow{Y}:\overleftrightarrow{h}^{-1} \to \overleftrightarrow{h}'^{-1} = \overleftrightarrow{Y}^{-1}\overleftrightarrow{h}^{-1}(\overleftrightarrow{Y}^{\mathrm{T}})^{-1}.$$

The relation between external momenta and internal momenta is given by

$$\vec{P} = \overleftrightarrow{\epsilon}^{\mathrm{T}}\vec{p}$$

and

$$\overleftrightarrow{\epsilon} \equiv \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1N} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \epsilon_{J1} & \cdots & \epsilon_{JN} \end{pmatrix}.$$

Since momenta are conserved, i.e.,

$$\sum_{n=1}^{N} p_n = 0,$$

we need only

$$\vec{P}^{\natural} = \overleftrightarrow{\epsilon}^{\natural}\vec{p}, \tag{A14}$$

where

$$\vec{P}^{\natural} \equiv (p_1, \cdots, p_{N-1}). \tag{A15}$$

For different choices of $N-1$ vertices, $\vec{P}^{\natural}$ is transformed according to

$$\begin{aligned} Y:\vec{P}^{\natural} \to \vec{P}^{\natural\prime} &\equiv \overleftrightarrow{\epsilon}^{\natural\prime\mathrm{T}}\vec{p} \\ &= \overleftrightarrow{Y}^{\mathrm{T}}\vec{P}^{\natural};\end{aligned} \tag{A16}$$

thus

$$\begin{aligned}\vec{P}^{\natural\mathrm{T}}\overleftrightarrow{h}^{-1}\vec{P}^{\natural} &= \vec{P}^{\natural\mathrm{T}}\overleftrightarrow{Y}(\overleftrightarrow{Y}^{-1}\overleftrightarrow{h}^{-1}\overleftrightarrow{Y}^{\mathrm{T}^{-1}})\overleftrightarrow{Y}^{\mathrm{T}}\vec{P}^{\natural}\\ &= \vec{P}^{\natural\prime\mathrm{T}}\overleftrightarrow{h}'^{-1}\vec{P}^{\natural\prime}.\end{aligned} \tag{A17}$$

Consequently, the V function is independent of the particular choice of the $N-1$ vertices when the V function is expressed in terms of external momenta using incidence matrix.

By using (A14), we have

$$V = \sum_{j=1}^{J}\alpha_j m_j^2 + \vec{p}^{\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural}\overleftrightarrow{h}^{-1}\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\vec{p}. \tag{A18}$$

However,

$$\begin{aligned}\overleftrightarrow{\epsilon}^{\natural}\overleftrightarrow{h}^{-1}\overleftrightarrow{\epsilon}^{\natural\mathrm{T}} &= \overleftrightarrow{\epsilon}^{\natural}\overleftrightarrow{Y}(\overleftrightarrow{Y}^{-1}\overleftrightarrow{h}^{-1}(Y^{\mathrm{T}})^{-1})\overleftrightarrow{Y}^{\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\\ &= \overleftrightarrow{\epsilon}^{\natural}\cdot\overleftrightarrow{h}'^{-1}\cdot\overleftrightarrow{\epsilon}^{\natural\prime\mathrm{T}};\end{aligned} \tag{A19}$$

i.e., the V function is also independent of the particular choice of the $N-1$ vertices when the U function is expressed in terms of internal momenta using incidence matrices.

We know that[2]

$$\begin{aligned}\det(\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural}) &= \sum_{\forall\{\mu\}}\det\overleftrightarrow{\sigma}_{\{\mu\}}\\ &= \text{total number of all possible choices of the tree sets.}\end{aligned}$$

That mean that the matrix $(\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural})$ has an inverse, so that,by means of (A14), we have

$$\vec{p} = \overleftrightarrow{\epsilon}^{\natural}(\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural})^{-1}\vec{P}^{\natural}. \tag{A20}$$

Therefore, (41) can be written in the form

$$\begin{aligned}V &= \sum_{j=1}^{J}\alpha_j m_j^2\\ &+ \vec{P}^{\natural\mathrm{T}}(\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural})^{-1}\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}(\overleftrightarrow{\alpha} - \overleftrightarrow{\alpha}\overleftrightarrow{\varepsilon}\overleftrightarrow{A}^{-1}\overleftrightarrow{\varepsilon}^{\mathrm{T}}\overleftrightarrow{\alpha})\overleftrightarrow{\epsilon}^{\natural}(\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural})^{-1}\vec{P}^{\natural}.\end{aligned} \tag{A21}$$

By virtue of (A8), Eq. (A21) gives

$$\overleftrightarrow{h}^{-1} = (\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural})^{-1}\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}(\overleftrightarrow{\alpha} - \overleftrightarrow{\alpha}\overleftrightarrow{\varepsilon}\overleftrightarrow{A}^{-1}\overleftrightarrow{\varepsilon}^{\mathrm{T}}\alpha)\overleftrightarrow{\epsilon}^{\natural}(\overleftrightarrow{\epsilon}^{\natural\mathrm{T}}\overleftrightarrow{\epsilon}^{\natural})^{-1}. \tag{A22}$$

APPENDIX B

In this appendix, we give some simple examples to illustrate the use of the formulas (29) and its counterpart given in Appendix A and also to show a certain "dual" nature between the *Poincaré incidence matrix* and the *loop matrix* associated with a Feynman graph. First, the usefulness of knowing the total number of possible tree graphs of a given Feynman graph lies in the fact that the U function can be evaluated by drawing all the possible tree graphs. [See Eq. (11) of Ref. 2.] It is easy to miss one in drawing the set of all tree graphs; thus it is useful to know the total number of trees we should have each time.

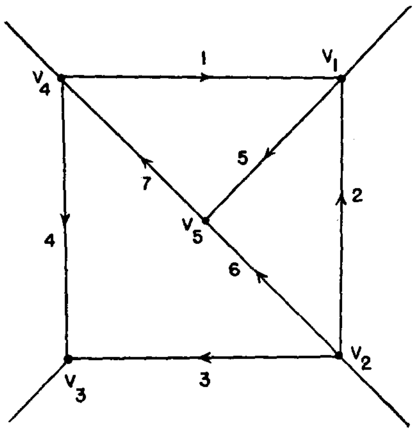


FIG. 3. An example.

Our first example is the one given in Fig. 3. The corresponding Poincaré incidence matrix is

$$\overleftrightarrow{\epsilon} \equiv \begin{pmatrix} -1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \quad \text{(B1)}$$

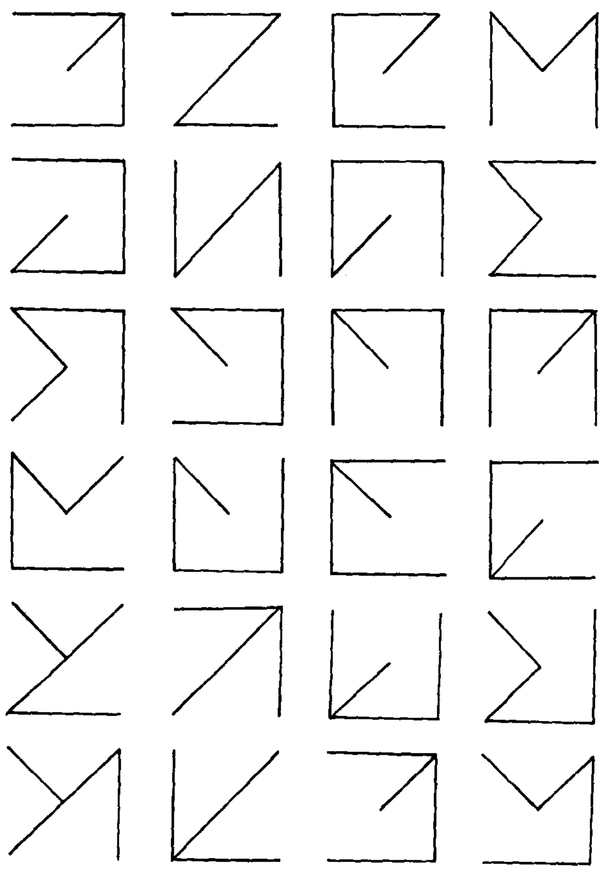


FIG. 4. Trees of Fig. 3.

To get $\overleftrightarrow{\epsilon}^{\natural}$, let us delete, say, the second column:

$$\overleftrightarrow{\epsilon}^{\natural} = \begin{pmatrix} -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \quad \text{(B2)}$$

Therefore,

$$\det(\overleftrightarrow{\epsilon}^{\natural T} \overleftrightarrow{\epsilon}^{\natural}) = \begin{pmatrix} 3 & 0 & -1 & -1 \\ 0 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ -1 & 0 & -1 & 3 \end{pmatrix} = 24, \quad \text{(B3)}$$

i.e., there are 24 trees from Fig. 3. They are given in Fig. 4.

On the other hand, one can use Eq. (29). If we draw the loops as in Fig. 5, then we have the loop matrix

$$\overleftrightarrow{\varepsilon} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \end{pmatrix}. \quad \text{(B4)}$$

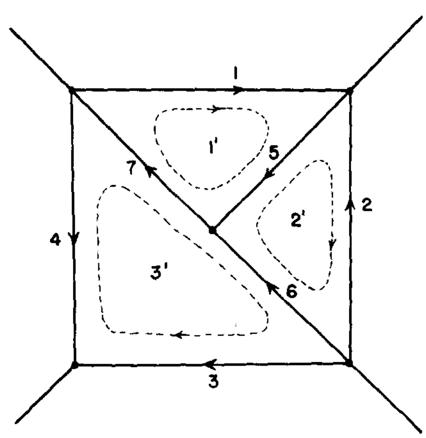


FIG. 5. Loop matrix analysis of Fig. 3.

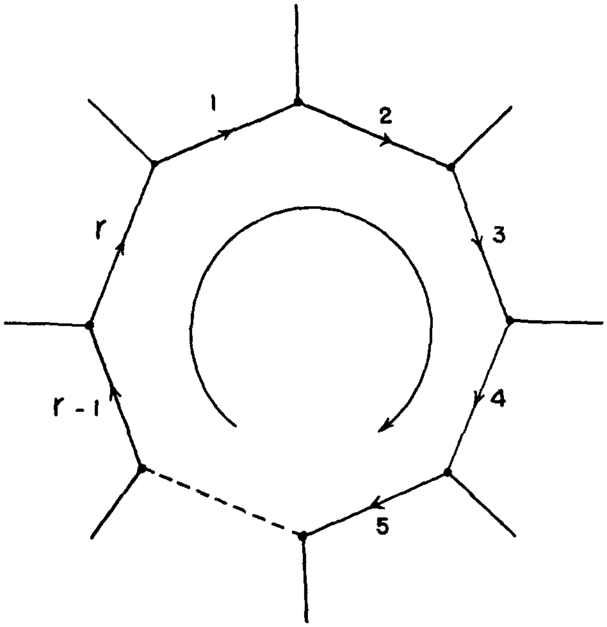


FIG. 6. An example.

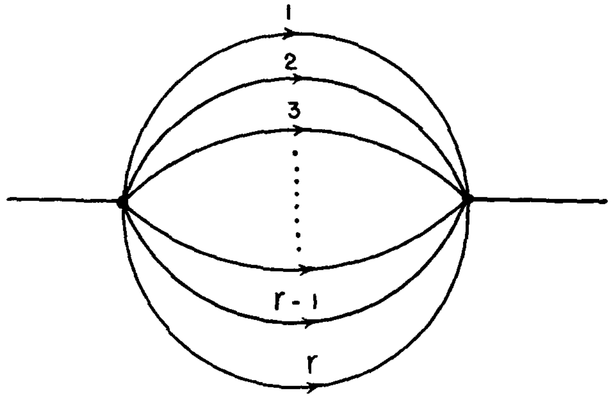


FIG. 7. An example.

Therefore,

$$\det(\overleftrightarrow{\varepsilon}^{T}\overleftrightarrow{\varepsilon}) = \begin{vmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 4 \end{vmatrix} = 24, \quad \text{(B5)}$$

which agrees with the result of (B3). However, we can easily see in this example that it is slightly easier to use loop matrix $\overleftrightarrow{\varepsilon}$ than to use Poincaré incidence matrix $\overleftrightarrow{\epsilon}$ because in the former case we deal with an $L \times L$ matrix ($L \equiv$ total number of independent loops), while in the latter case we deal with a $(N-1) \times (N-1)$ matrix ($N \equiv$ total number of vertices in the graph). Therefore, a choice between the two ways of calculation depends on whether

$$L > N - 1 \quad \text{or} \quad L < N - 1. \quad \text{(B6)}$$

The following trivial examples show not only the extreme cases of favoring one method over another but also illustrate the "duality" between the two graphs.

Consider an r-sided polygon graph of Fig. 6. It is trivial that

$$\overleftrightarrow{\varepsilon} = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \quad \text{(B7)}$$

where $\overleftrightarrow{\varepsilon}$ has r elements 1. Thus,

$$\det(\overleftrightarrow{\varepsilon}^{T}\overleftrightarrow{\epsilon}) = r. \quad \text{(B8)}$$

On the other hand, if one wants to use Poincaré incidence matrix, the computation is not simple at all. For instance, let us consider $r = 5$:

$$\overleftrightarrow{\epsilon} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{(B9)}$$

which leads to, by deleting the last column of $\overleftrightarrow{\epsilon}$, say,

$$\det(\overleftrightarrow{\epsilon}^{\natural T}\overleftrightarrow{\epsilon}^{\natural}) = \begin{vmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix} = 5. \quad \text{(B10)}$$

This is certainly a ridiculous way of doing a trivial job.

The last example is sort of "dual" to that of Fig. 6. Consider now the graph given in Fig. 7.

In this case, using Poincaré incidence matrix, we get

$$\overleftrightarrow{\epsilon} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & -1 \end{pmatrix}, \quad \text{(B11)}$$

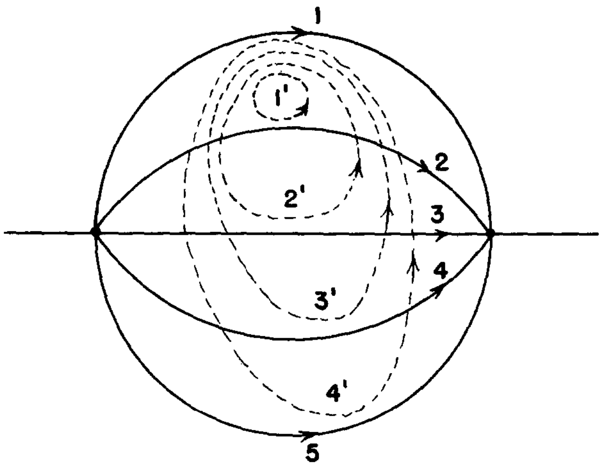


FIG. 8. An example.

an $r \times 2$ matrix, i.e.,

$$\overset{\leftrightarrow}{\epsilon}^{\natural} = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \qquad \text{(B12)}$$

with r elements, which trivially leads to

$$\det(\overset{\leftrightarrow}{\epsilon}^{\natural T} \overset{\leftrightarrow}{\epsilon}^{\natural}) = r. \qquad \text{(B13)}$$

On the other hand, it is not simple to evaluate it with loop matrix method. For instance, consider the case of $r = 5$; then for the loops drawn in Fig. 8 we have

$$\overset{\leftrightarrow}{\varepsilon} = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \qquad \text{(B14)}$$

Therefore,

$$\det(\overset{\leftrightarrow}{\varepsilon}^{T} \overset{\leftrightarrow}{\varepsilon}) = \begin{vmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{vmatrix} = 5,$$

which is again a roundabout way of getting the trivial result (B13).

[1] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. **18**, 1 (1961).
[2] Y. Chow, J. Math. Phys. **5**, 1255 (1964).
[3] Y. Chow and D. J. Kleitman, Progr. Theoret. Phys. (Kyoto) **32**, 950 (1964).
[4] Y. Chow, J. Math. Phys. **7**, 1158 (1966).
[5] K. Symanzik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958).
[6] A. A. Loganov, I. T. Todorov, and N. A. Chernikov, Zh. Eksp. Teor. Fiz. **42**, 1285 (1962) [Sov. Phys. JETP **15**, 891 (1962)].
[7] I. T. Todorov, Doctoral dissertation, Joint Institute of Nuclear Research, Report P-1205, Dubna, 1963, and literature cited there.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Nonlinear Light Propagation in a Resonant Medium and Causality*

R. C. T. DA COSTA†
Department of Physics and Astronomy, The University of Rochester, Rochester, New York

(Received 6 November 1969; Revised Manuscript 23 March 1970)

The propagation of an optical pulse through a system of two-level atoms embedded in a host medium of constant refractive index n is considered. It is shown that the values of the solutions of the nonlinear system of coupled Maxwell and Schrödinger equations at time t at a point z, in the McCall–Hahn approximation, depend only on data contained in the interval $[z - c/nt, z]$ at $t = 0$. Thus, information cannot be transmitted through this medium with velocity greater than c/n. Apparent violations of this property in the case of self-induced transparency are explained.

I. INTRODUCTION

The problem of how causality requirements are satisfied in the propagation of light through a material medium has been considered since the beginning of this century. In 1914, Sommerfeld and Brillouin[1] investigated the propagation of an electromagnetic wavetrain with a sharp front through a medium with a refractive index given by the Lorentz dispersion formula. They showed that, in spite of the fact that both the phase and the group velocity can exceed c in a

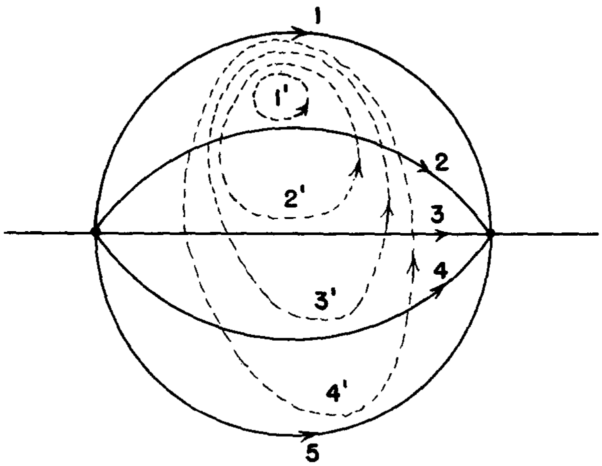


FIG. 8. An example.

an $r \times 2$ matrix, i.e.,

$$\overleftrightarrow{\epsilon}^{\natural} = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \qquad \text{(B12)}$$

with r elements, which trivially leads to

$$\det (\overleftrightarrow{\epsilon}^{\natural T} \overleftrightarrow{\epsilon}^{\natural}) = r. \qquad \text{(B13)}$$

On the other hand, it is not simple to evaluate it with loop matrix method. For instance, consider the case of $r = 5$; then for the loops drawn in Fig. 8 we have

$$\overleftrightarrow{\varepsilon} = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \qquad \text{(B14)}$$

Therefore,

$$\det (\overleftrightarrow{\varepsilon}^{T} \overleftrightarrow{\varepsilon}) = \begin{vmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{vmatrix} = 5,$$

which is again a roundabout way of getting the trivial result (B13).

[1] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. **18**, 1 (1961).
[2] Y. Chow, J. Math. Phys. **5**, 1255 (1964).
[3] Y. Chow and D. J. Kleitman, Progr. Theoret. Phys. (Kyoto) **32**, 950 (1964).
[4] Y. Chow, J. Math. Phys. **7**, 1158 (1966).
[5] K. Symanzik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958).
[6] A. A. Loganov, I. T. Todorov, and N. A. Chernikov, Zh. Eksp. Teor. Fiz. **42**, 1285 (1962) [Sov. Phys. JETP **15**, 891 (1962)].
[7] I. T. Todorov, Doctoral dissertation, Joint Institute of Nuclear Research, Report P-1205, Dubna, 1963, and literature cited there.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Nonlinear Light Propagation in a Resonant Medium and Causality*

R. C. T. DA COSTA†
Department of Physics and Astronomy, The University of Rochester, Rochester, New York

(Received 6 November 1969; Revised Manuscript 23 March 1970)

The propagation of an optical pulse through a system of two-level atoms embedded in a host medium of constant refractive index n is considered. It is shown that the values of the solutions of the nonlinear system of coupled Maxwell and Schrödinger equations at time t at a point z, in the McCall–Hahn approximation, depend only on data contained in the interval $[z - c/nt, z]$ at $t = 0$. Thus, information cannot be transmitted through this medium with velocity greater than c/n. Apparent violations of this property in the case of self-induced transparency are explained.

I. INTRODUCTION

The problem of how causality requirements are satisfied in the propagation of light through a material medium has been considered since the beginning of this century. In 1914, Sommerfeld and Brillouin[1] investigated the propagation of an electromagnetic wavetrain with a sharp front through a medium with a refractive index given by the Lorentz dispersion formula. They showed that, in spite of the fact that both the phase and the group velocity can exceed c in a region of anomalous dispersion, the front of the wave-train cannot propagate faster than c. A far-reaching extension of this result was obtained by Kramers[2] and Kronig,[3] who derived a general condition that must be satisfied by the refractive index of any *linear* medium, the dispersion relation,[4] in order that the requirements due to causality be satisfied.

The results of the recent investigations on the propagation of intense laser pulses in resonant media[5–12] suggest that the causality problem should be reexamined in this new domain. In fact, the interaction between the intense laser field and the resonant medium is described by an essentially nonlinear system of coupled Maxwell–Schrödinger equations, so that the usual causality criterion expressed by the Kramers–Kronig dispersion relation, which is valid only for linear media, cannot be applied.

It has been found[5,6,9] that, in media containing excited atoms, there seems to exist, under certain conditions, the possibility of pulse propagation with velocities greater than c. This would not occur for pulses with sharp wavefronts, but rather for pulses with tails that (though very weak) would extend to infinity.

Due to the absence of sharp fronts, the fact that the bulk of such a pulse would travel faster than c does not constitute a clear cut violation of causality. In fact, it was suggested by Basov *et al.*[6] that this effect can be explained in terms of stimulated emission induced by the weak leading edge of the pulse (cf. Sec. IV below).

With the discovery of self-induced transparency by McCall and Hahn,[7,8] the possibility was raised that such "fast" pulses might propagate without change of shape, as if the medium were transparent. Again, these self-transparent pulses, as well as other "fast" solutions found by Eberly[9] and Crisp,[10] would not have sharp wavefronts, so that it is not *a priori* clear whether or not their existence would constitute a violation of causality.

It should be emphasized that, even for pulses without sharp fronts, the problem of causality can still be discussed. What is required is that influence shall not be propagated faster than c. Whether or not this requirement is satisfied can be determined quite generally by investigating the domain of dependence (Ref. 13, pp. 209, 438) of the solution at a given space–time point on data given at previous times (e.g., for the ordinary wave equation, this domain is the backward light cone).

To the best of our knowledge, it has not been proved that all solutions of the nonlinear equations employed in the theory of self-induced transparency satisfy the causality condition in the above sense. On the other hand, it is not inconceivable that causality-violating solutions might exist, in view of the nonrelativistic character of the Schrödinger equation, as well as the use of several approximations in the derivation of the equations. It seems worthwhile, therefore, to investigate this problem. The discussion of causality for a nonlinear system can also be of some methodological interest.

The basic equations of the theory of self-induced transparency are reviewed in Sec. II. The domain of dependence of the solutions is derived in Sec. III. The result is that all solutions of the equations are indeed causal. In Sec. IV, this result is applied to self-transparent "fast" pulses, providing a justification of Basov's argument in this case. Finally, some comments are made about the reasons for the validity of the obtained results.

II. THE McCALL–HAHN EQUATIONS

Let us consider an infinite nondispersive host medium with refractive index n, in which is embedded a uniform distribution of N two-level atoms per unit volume. Let the atomic energy levels be $\pm\frac{1}{2}\hbar\omega_0$ and let p be the magnitude of the transverse components of the transition electric dipole moment. The electric field is taken as a slightly perturbed, circularly polarized plane wave propagating in the z direction,

$$\mathbf{E}(z,t) = \mathcal{E}(z,t)\{\cos[\omega t - kz + \phi(z)]\hat{\mathbf{x}} + \sin[\omega t - kz + \phi(z)]\hat{\mathbf{y}}\}, \quad (1)$$

where both $\mathcal{E}$ and ϕ are assumed to be slowly varying functions of their arguments and ϕ is assumed to be time independent, so that frequency modulation and pulling effects are disregarded.

Let u and v be the dispersive and absorptive components, respectively, of the transverse electric polarization density associated with the two-level atoms, and let W be their energy density. These quantities can be combined together to define a pseudopolarization vector[14,15]

$$\mathbf{P} = u\hat{\mathbf{x}} + v\hat{\mathbf{y}} + (\kappa/\omega_0)W\hat{\mathbf{z}},$$

where $\kappa = 2p/\hbar$. In a frame of reference (x', y', z) rotating about the z axis with angular velocity ω, this vector satisfies the torque equation[14,15]

$$\frac{d\mathbf{P}}{dt} = \mathbf{P} \times [\kappa\mathcal{E}\hat{\mathbf{x}}' + (\omega - \omega_0)\hat{\mathbf{z}}]. \quad (2)$$

The effect of inhomogeneous broadening on the system of two-level atoms is taken into account by introducing a normalized atomic line-shape function

$g(\gamma)$, $\gamma = \omega_0 - \omega$, with the polarization now given by

$$P_+(z, t) = P_x(z, t) + iP_y(z, t) = \int_{-\infty}^{\infty} g(\gamma)[u(\gamma, z, t) + iv(\gamma, z, t)] \times \exp\{i[\omega t - kz + \phi(z)]\}\, d\gamma.$$

By combining the above results with Maxwell's equations and neglecting higher-order terms (slowly varying envelope and phase approximation), one obtains the following set of equations[7,8]:

$$\frac{\partial \mathcal{E}}{\partial z} + \frac{n}{c}\frac{\partial \mathcal{E}}{\partial t} = \frac{2\pi\omega}{nc}\int_{-\infty}^{\infty} v(\gamma, z, t) g(\gamma)\, d\gamma, \quad (3)$$

$$k\frac{\partial \phi}{\partial z}\mathcal{E} = -2\pi\frac{\omega^2}{c^2}\int_{-\infty}^{\infty} u(\gamma, z, t) g(\gamma)\, d\gamma, \quad (4)$$

$$\frac{\partial u}{\partial t} = -\gamma v - \frac{u}{T_2}, \quad \frac{\partial v}{\partial t} = \gamma u + \frac{\kappa^2}{\omega}\mathcal{E}W - \frac{v}{T_2},$$

$$\frac{\partial W}{\partial t} = -\omega v \mathcal{E} - \frac{W - W_0}{T_1}, \quad (5)$$

where we have introduced phenomenological Bloch-type damping terms, associated with the relaxation times T_1 and T_2.

Equations (3)–(5) are the McCall–Hahn equations.

III. PROOF OF CAUSALITY

As mentioned in Sec. I, the problem of causality can be reduced quite generally to the determination of the domain of dependence of the solutions to the above equations. Equations (3) and (5) form a system of integral and partial differential equations for the unknowns $\mathcal{E}$, u, v, and W, whereas Eq. (4) plays the role of a subsidiary condition, since ϕ has been assumed to be time independent.

To simplify the treatment, we approximate the integral in (3) by a sum of m terms, where m can be taken arbitrarily large:

$$\frac{\partial \mathcal{E}}{\partial z} + \frac{n}{c}\frac{\partial \mathcal{E}}{\partial t} = \frac{2\pi\omega}{nc}\sum_{l=1}^{m} v(\gamma_l, z, t) g(\gamma_l)\Delta\gamma_l.$$

We also introduce the new unknowns

$$u_l(z, t) = u(\gamma_l, z, t), \quad v_l(z, t) = v(\gamma_l, z, t),$$
$$W_l(z, t) = W(\gamma_l, z, t), \quad l = 1, 2, \cdots, m.$$

We can then write (3) and (5) as a system of $3m + 1$ semilinear[13] partial differential equations in $3m + 1$ unknowns:

$$\frac{\partial \mathcal{E}}{\partial z} + \frac{n}{c}\frac{\partial \mathcal{E}}{\partial t} = \frac{2\pi\omega}{nc}\sum_{l=1}^{m} v_l g_l \Delta\gamma_l, \quad (6)$$

$$\frac{\partial u_l}{\partial t} = -\gamma_l v_l - \frac{u_l}{T_2}, \quad (7)$$

$$\frac{\partial v_l}{\partial t} = \gamma_l u_l + \frac{\kappa^2}{\omega}\mathcal{E}W_l - \frac{v_l}{T_2}, \quad (8)$$

$$\frac{\partial W_l}{\partial t} = -\omega_0 \mathcal{E} v_l - \frac{W - W_0}{T_1}, \quad (9)$$

where the only nonlinear terms are those in $\mathcal{E}W_l$ and $\mathcal{E}v_l$.

As was mentioned in the Introduction, the nonlinearity of this system prevents us from applying the well-known causality criteria associated with dispersion relations. However, it is possible to determine the domain of dependence by a technique similar to that employed in the uniqueness proof for quasilinear systems (Ref. 13, p. 448).

Let $\mathcal{E}_1$, $u_{l,1}$, $v_{l,1}$, $W_{l,1}$, and $\mathcal{E}_2$, $u_{l,2}$, $v_{l,2}$, $W_{l,2}$ be two solutions of the system which, at $t = 0$, coincide at all points of the closed interval $[z_0, z_1]$. The new unknowns

$$\mathcal{E}^* = \mathcal{E}_2 - \mathcal{E}_1, \quad u_l^* = u_{l,2} - u_{l,1},$$
$$v_l^* = v_{l,2} - v_{l,1}, \quad W_l^* = W_{l,2} - W_{l,1} \quad (10)$$

then satisfy the linear equations (6) and (7), together with the equations

$$\frac{\partial v_l^*}{\partial t} = \gamma_l u_l^* + \frac{\kappa^2}{\omega_0}(\mathcal{E}_2 W_l^* + W_{l,1}\mathcal{E}^*) - \frac{v_l^*}{T_2},$$

$$\frac{\partial W^*}{\partial t} = -\omega_0(\mathcal{E}_2 v_l^* + v_{l,1}\mathcal{E}^*) - \frac{W_l^*}{T_1}, \quad (11)$$

which are also linear in the above unknowns.

Considering $\mathcal{E}_2$, $v_{l,1}$, and $W_{l,1}$ as known functions of (z, t), we see that the unknowns (10) are solutions of a completely hyperbolic linear homogeneous system[13] satisfying the initial conditions

$$\mathcal{E}^*(z, 0) = u_l^*(z, 0) = v_l^*(z, 0) = W_l^*(z, 0) = 0$$

for all z in $[z_0, z_1]$. It follows (Ref. 13, p. 445) that $\mathcal{E}^*$, u_l^*, v_l^*, and W_l^* are identically zero in a region of the upper (z, t) plane bounded by the z axis and by two characteristic curves passing, respectively, through z_0 and z_1.

Since the characteristic curves for the system (6), (7), and (11) are given by

$$z = C_1 \quad \text{and} \quad z - (c/n)t = C_2 \quad (12)$$

(where C_1 and C_2 are constants), the above region is the right triangle bounded by the z axis and by the straight lines

$$z = z_1, \quad z = z_0 + (c/n)t.$$

We conclude that the values of the solution of (6)–(9) in this triangle are uniquely determined by the initial data in the interval $[z_0, z_1]$. The domain of dependence of a given point (z^*, t^*) is, therefore, the angular region of the (z, t) plane

$$z^* + (c/n)(t - t^*) \leq z \leq z^*, \quad t \leq t^*. \tag{13}$$

Since this region is contained within the backward light cone with vertex at (z^*, t^*), the proof of causality is completed.

IV. THE SELF-TRANSPARENT SOLUTIONS

In the treatment of self-induced transparency,[7,8] solutions of the McCall–Hahn equations of the form $\mathcal{E} = \mathcal{E}(z - Vt)$ have been found, corresponding to an envelope shape which is preserved and propagated with velocity V. For a hyperbolic secant-shaped pulse with all the atoms initially $(t \to -\infty)$ in the ground state, it is found that $V < c/n$. This retardation arises from the continuous absorption of energy by the atoms from the leading edge of the pulse and its later reemission into the trailing edge in such a way that the pulse shape is preserved; ultimately $(t \to \infty)$, all the atoms return to their ground state.

If all the atoms are initially in the excited state, the converse process takes place: The leading edge of the pulse gives rise to stimulated emission by the atoms, and the energy is later reabsorbed from the field to form the trailing edge. This gives rise to a velocity $V > c/n$ (and, eventually, greater than c). Other shape-preserving solutions with $V > c/n$ have also been found.[9,10] It must be emphasized that spontaneous emission, as well as other damping terms included in (5), are neglected in the treatments that led to these results.

Without going into the question of the stability of such "fast" solutions, that would involve a detailed examination of the effect of neglected terms, it follows from the results of Sec. III that no violation of causality is involved. The existence of solutions of the form $\mathcal{E}(z - Vt)$ does not imply a causal relation between the points (z, t) and $(z - Vt', t - t')$. In the case of hyperbolic-secant pulses with $V > c/n$, for example, the domain of dependence (13) of the peak of the pulse at time t does not contain the peak of the pulse at any previous time, but only a part of its leading edge.

What propagates with velocity V, in this case, is only the pulse shape and not the energy. If we consider two times t and t' $(t > t')$, at which the pulse is peaked at z and z', respectively, it is not the energy around z that propagates to z' during the time interval $t - t'$; rather, the excitation energy that was already contained in the atoms around z' is triggered by the weak leading edge of the pulse and released into the field to form the peak at t'. This effect, as ought to be expected, depends on the pulse width, the "propagation velocity" V approaching c/n as the width goes to zero.

An electric field pulse with a sharp wavefront, i.e., identically vanishing for $z > Vt$, might be considered to define the velocity of propagation of a signal. In this case, we always have $V = c/n$, as may readily be seen from the fact that characteristic curves [cf. (12)] are the only branch curves for a differential system,[13] i.e., the curves along which two different solutions can be joined together.

Another point that deserves some comment is the independence of the values of the solutions at (z, t) with respect to data given at $t' < t$, $z' > z$, which is related to the characteristics $z = C_2$ in (12). This is a consequence of the slowly varying envelope approximation, in which contributions from backscattered radiation[16] are neglected, as well as of the assumed absence of interaction among the atoms (other than through the common radiation field). This feature, therefore, is not likely to appear in more general treatments.

A final comment about the reasons why causality requirements are actually met in this model is in order. The nonrelativistic character of the Schrödinger equation does not lead to any limiting velocity. However, this property is not manifested when the atoms are treated as pointlike particles, interacting only through the common radiation field. In this sense, therefore, the equations are causal not so much in spite of the approximations made in their derivation, but rather because of them.

ACKNOWLEDGMENTS

The author wishes to thank Professor H. M. Nussenzveig for many helpful discussions and suggestions and Professor M. F. Kaplon for his hospitality at the Department of Physics of the University of Rochester. A fellowship from the Brazilian National Research Council is also gratefully acknowledged.

* This work was supported in part by the U.S. Office of Naval Research, Contract No. NO0014-67-A-0398-0005.

† On leave of absence from Pontificia Universidade Catolica, Rio de Janeiro.

[1] Cf. L. Brillouin, *Wave Propagation and Group Velocity* (Academic, New York, 1960).
[2] H. A. Kramers, Atti Congr. Intern. Fis. (Como) **2**, 545 (1927).
[3] R. de L. Kronig, J. Opt. Soc. Am. **12**, 547 (1926).
[4] J. Toll, Phys. Rev. **104**, 1760 (1956).
[5] F. T. Arecchi and R. Bonifacio, IEEE J. Quantum Electron. **1**, 169 (1965).
[6] N. G. Basov, P. V. Ambartsumyan, V. S. Zuev, P. G. Kryukov, and V. S. Letokhov, Dokl. Akad. Nauk SSSR **165**, 58 (1965) [Sov. Phys. Dokl. **10**, 1039 (1966)]; Zh. Exsp. Teor. Fiz. **50**, 23 (1966) [Sov. Phys. JETP **23**, 16 (1966)].
[7] S. L. McCall and E. L. Hahn, Phys. Rev. Letters **18**, 908 (1967).
[8] S. L. McCall and E. L. Hahn, Phys. Rev. **183**, 457 (1969).
[9] J. H. Eberly, Phys. Rev. Letters **22**, 760 (1969).
[10] M. D. Crisp, Phys. Rev. Letters **22**, 820 (1969).
[11] F. A. Hopf and M. O. Scully, Phys. Rev. **179**, 399 (1969).
[12] A. Icsevgy and W. E. Lamb, Jr., Phys. Rev. **185**, 517 (1969).
[13] R. Courant and D. Hilbert, Methods of Mathematical Physics (Interscience, New York, 1962), Vol. II.
[14] R. P. Feynman, F. L. Vernon, and R. W. Hellwarth, J. Appl. Phys. **28**, 49 (1957).
[15] I. D. Abella, N. A. Kurnit, and S. R. Hartmann, Phys. Rev. **141**, 391 (1966).
[16] M. D. Crisp, Opt. Commun. **1**, 59 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Labeling and Multiplicity Problem in the Reduction $U(n+m) \downarrow U(n) \times U(m)$

JOUKO MICKELSSON
Research Institute for Theoretical Physics, University of Helsinki, Helsinki, Finland

(Received 27 January 1970)

A labeling for the basis vectors in a general unitary irreducible representation of $U(n+m)$ is introduced with the Casimir operators of the subgroup $U(n) \times U(m)$ diagonal. The multiplicities are calculated in the reduction $U(n+m) \downarrow U(n) \times U(m)$ for some special cases.

1. INTRODUCTION

The irreducible representations of the unitary and orthogonal groups have already been constructed explicitly twenty years ago by Gel'fand and Zetlin,[1] who introduced the triangular Gel'fand–Zetlin patterns for the labeling of the basis vectors. In the case of $U(n)$, this labeling is based on the Weyl branching rule.[2] In Ref. 1, the matrix elements of the generators of the corresponding Lie algebras are given without derivations. These results were later derived by Nagel and Moshinsky[3] and by Baird and Biedenharn[4] for $U(n)$, using boson operator techniques. For the group $SO(n)$, this has been done by Pang and Hecht[5] and Wong.[6]

Common to all these papers is that the problem is treated using the canonical chain of subgroups

$$U(n) \supset U(n-1) \supset \cdots \supset U(1),$$
$$SO(n) \supset SO(n-1) \supset \cdots \supset SO(2),$$

respectively, and diagonalizing the Casimir operators of these subgroups. During the past years it has become important to have explicit reductions also in the cases $U(n+m) \downarrow U(n) \times U(m)$ and $SO(n+m) \downarrow SO(n) \times SO(m)$, i.e., the Casimir operators of the subgroup $U(n) \times U(m)$ and $SO(n) \times SO(m)$ are diagonalized. In particle physics, especially popular have been the choices $U(2n) \downarrow U(n) \times U(n)$, $n = 2, 3, 6$, and corresponding noncompact forms, e.g., $U(3, 3) \downarrow U(3) \times U(3)$. In this paper, we study the general case $U(n+m) \downarrow U(n) \times U(m)$.

In Sec. 2, we give a complete set of labels for the basis vectors in the reduction $U(n+m) \downarrow U(n) \times U(m)$. Using this set of labels, we derive formulas (Sec. 3) for the multiplicities of the UIR's of $U(n) \times U(m)$ in a UIR of $U(n+m)$ in the following cases: (a) $U(n+2) \downarrow U(n) \times U(2)$ and (b) $\lambda_4 = \lambda_5 = \cdots = \lambda_{n+m} = 0$ (the λ_i are the row lengths in the corresponding Young diagram equaling the highest weights). We complete Sec. 3 with a discussion of the physically interesting case $U(6) \downarrow U(3) \times U(3)$.[7]

2. A FORMAL BASIS FOR THE REDUCTION $U(n+m) \downarrow U(n) \times U(m)$

In the case $U(n) \supset U(n-1) \supset \cdots \supset U(1)$, the labels for the basis vectors are given by a set of nonnegative integers m_{ij}, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, i$, with the following properties[1,2]:

(i) the numbers m_{kj}, $j = 1, 2, \cdots, k$, are the highest weights for a UIR of the subgroup $U(k)$;
(ii) $m_{k+1i} \geq m_{ki} \geq m_{k+1i+1}$.

[1] Cf. L. Brillouin, *Wave Propagation and Group Velocity* (Academic, New York, 1960).
[2] H. A. Kramers, Atti Congr. Intern. Fis. (Como) **2**, 545 (1927).
[3] R. de L. Kronig, J. Opt. Soc. Am. **12**, 547 (1926).
[4] J. Toll, Phys. Rev. **104**, 1760 (1956).
[5] F. T. Arecchi and R. Bonifacio, IEEE J. Quantum Electron. **1**, 169 (1965).
[6] N. G. Basov, P. V. Ambartsumyan, V. S. Zuev, P. G. Kryukov, and V. S. Letokhov, Dokl. Akad. Nauk SSSR **165**, 58 (1965) [Sov. Phys. Dokl. **10**, 1039 (1966)]; Zh. Exsp. Teor. Fiz. **50**, 23 (1966) [Sov. Phys. JETP **23**, 16 (1966)].
[7] S. L. McCall and E. L. Hahn, Phys. Rev. Letters **18**, 908 (1967).
[8] S. L. McCall and E. L. Hahn, Phys. Rev. **183**, 457 (1969).
[9] J. H. Eberly, Phys. Rev. Letters **22**, 760 (1969).
[10] M. D. Crisp, Phys. Rev. Letters **22**, 820 (1969).
[11] F. A. Hopf and M. O. Scully, Phys. Rev. **179**, 399 (1969).
[12] A. Icsevgy and W. E. Lamb, Jr., Phys. Rev. **185**, 517 (1969).
[13] R. Courant and D. Hilbert, Methods of Mathematical Physics (Interscience, New York, 1962), Vol. II.
[14] R. P. Feynman, F. L. Vernon, and R. W. Hellwarth, J. Appl. Phys. **28**, 49 (1957).
[15] I. D. Abella, N. A. Kurnit, and S. R. Hartmann, Phys. Rev. **141**, 391 (1966).
[16] M. D. Crisp, Opt. Commun. **1**, 59 (1969).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Labeling and Multiplicity Problem in the Reduction $U(n+m) \downarrow U(n) \times U(m)$

JOUKO MICKELSSON
Research Institute for Theoretical Physics, University of Helsinki, Helsinki, Finland

(Received 27 January 1970)

A labeling for the basis vectors in a general unitary irreducible representation of $U(n+m)$ is introduced with the Casimir operators of the subgroup $U(n) \times U(m)$ diagonal. The multiplicities are calculated in the reduction $U(n+m) \downarrow U(n) \times U(m)$ for some special cases.

1. INTRODUCTION

The irreducible representations of the unitary and orthogonal groups have already been constructed explicitly twenty years ago by Gel'fand and Zetlin,[1] who introduced the triangular Gel'fand–Zetlin patterns for the labeling of the basis vectors. In the case of $U(n)$, this labeling is based on the Weyl branching rule.[2] In Ref. 1, the matrix elements of the generators of the corresponding Lie algebras are given without derivations. These results were later derived by Nagel and Moshinsky[3] and by Baird and Biedenharn[4] for $U(n)$, using boson operator techniques. For the group $SO(n)$, this has been done by Pang and Hecht[5] and Wong.[6]

Common to all these papers is that the problem is treated using the canonical chain of subgroups

$$U(n) \supset U(n-1) \supset \cdots \supset U(1),$$
$$SO(n) \supset SO(n-1) \supset \cdots \supset SO(2),$$

respectively, and diagonalizing the Casimir operators of these subgroups. During the past years it has become important to have explicit reductions also in the cases $U(n+m) \downarrow U(n) \times U(m)$ and $SO(n+m) \downarrow SO(n) \times SO(m)$, i.e., the Casimir operators of the subgroup $U(n) \times U(m)$ and $SO(n) \times SO(m)$ are diagonalized. In particle physics, especially popular have been the choices $U(2n) \downarrow U(n) \times U(n)$, $n = 2, 3, 6$, and corresponding noncompact forms, e.g., $U(3, 3) \downarrow U(3) \times U(3)$. In this paper, we study the general case $U(n+m) \downarrow U(n) \times U(m)$.

In Sec. 2, we give a complete set of labels for the basis vectors in the reduction $U(n+m) \downarrow U(n) \times U(m)$. Using this set of labels, we derive formulas (Sec. 3) for the multiplicities of the UIR's of $U(n) \times U(m)$ in a UIR of $U(n+m)$ in the following cases: (a) $U(n+2) \downarrow U(n) \times U(2)$ and (b) $\lambda_4 = \lambda_5 = \cdots = \lambda_{n+m} = 0$ (the λ_i are the row lengths in the corresponding Young diagram equaling the highest weights). We complete Sec. 3 with a discussion of the physically interesting case $U(6) \downarrow U(3) \times U(3)$.[7]

2. A FORMAL BASIS FOR THE REDUCTION $U(n+m) \downarrow U(n) \times U(m)$

In the case $U(n) \supset U(n-1) \supset \cdots \supset U(1)$, the labels for the basis vectors are given by a set of nonnegative integers m_{ij}, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, i$, with the following properties[1,2]:

(i) the numbers m_{kj}, $j = 1, 2, \cdots, k$, are the highest weights for a UIR of the subgroup $U(k)$;
(ii) $m_{k+1\,i} \geq m_{ki} \geq m_{k+1\,i+1}$.

These numbers can be arranged in a pattern (GZ pattern[1]),

$$\begin{bmatrix} m_{n1} & m_{n2} & \cdots & m_{nn} \\ m_{n-1\,1} & m_{n-1\,2} & \cdots & m_{n-1\,n-1} \\ \cdot & & & \cdot \\ & \cdot & \cdot & \\ & m_{21} & m_{22} & \\ & & m_{11} & \end{bmatrix},$$

such that every number is greater than or equal to the number above right and less than or equal to the number above left. Now we would like to have the same kind of pattern in the general case $U(n+m) \downarrow U(n) \times U(m)$. $U(n+m)$ operates in a $(n+m)$-dimensional complex vector space; $U(n)$ transforms the first n components and $U(m)$ the last m components of an $(n+m)$-component vector. To begin with, we have two small GZ patterns, namely, those associated with the UIR's of $U(n)$ and $U(m)$. Then we have $n+m$ labels furnished by the highest weights of the group $U(n+m)$. But we need more labels, because, in general, a UIR of $U(n) \times U(m)$ occurs more than once in a UIR of $U(n+m)$.

Let $[l_i]$, $i = 1, 2, \cdots, n$ (respectively $[j_i]$, $i = 1, 2, \cdots, m$) be the row lengths in the Young diagrams associated with the UIR's of $U(n)$ [$U(m)$, respectively] and $[\lambda_i]$, $i = 1, 2, \cdots, n+m$, for $U(n+m)$. As shown in Ref. 8, the multiplicity $m_{lj\lambda}$ giving the number of times the representation $[l_i] \times [j_i]$ of $U(n) \times U(m)$ occurs in $[\lambda_i]$ is equal to the multiplicity of $[\lambda_i]$ in the direct product

$$\Big[l_1, l_2, \cdots, l_n, \underbrace{0, 0, \cdots, 0}_{m\ \text{zeros}}\Big] \times \Big[j_1, j_2, \cdots, j_m, \underbrace{0, 0, \cdots, 0}_{n\ \text{zeros}}\Big]$$

of two UIR's of $U(n+m)$:

$$[l_i, 0] \times [j_i, 0] = \sum_{\lambda} m_{lj\lambda}[\lambda_i]. \tag{1}$$

Let us recall how the direct product of two irreducible representations of the unitary group is reduced.[9] One proceeds as follows: one fills the boxes in the first row of the Young diagram $[j_i]$ with letters a, the second row with b's, and so on. Then, one forms new Young diagrams from the diagram $[l_i]$ by adding first the boxes which contain a's, then the boxes filled by b's and so on, in every possible way subject to the following three conditions:

(i) the new diagram should be a regular Young diagram, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n+m}$;

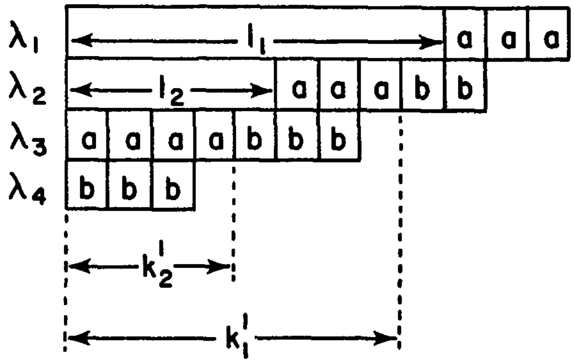


FIG. 1. Definition of multiplicity labels k_1^1 and k_2^1.

(ii) no letter should be repeated in the same column;

(iii) reading from right to left and from top to bottom in the new diagram, the number of a's should be $\geq$ the number of b's $\geq$ the number of c's $\cdots$ at every step.

Now we introduce the *multiplicity labels* k_j^i, $i = 1, 2, \cdots, m-1$, and $j = 1, 2, \cdots, n$. The number k_j^1 is defined as the length of the $(j+1)$th row in the intermediate diagram *after insertion of the letters a*. In a similar way, the number k_j^2 gives the length of the $(j+2)$th row after insertion of the b's, and so on. Because of (i)–(iii), we have some restrictions on the numbers k_j^i, in addition to the requirement that the row lengths in the new diagram should be $[\lambda_i]$. Let us first study the case[10] $U(4) \downarrow U(2) \times U(2)$ as an example and then generalize to arbitrary m and n. From (i), (ii), and the requirement mentioned above follow the inequalities

$$\begin{gathered} \lambda_2 \geq k_1^1 \geq \lambda_3 \geq k_2^1 \geq \lambda_4, \\ \lambda_1 \geq l_1 \geq k_1^1 \geq l_2 \geq k_2^1. \end{gathered} \tag{2}$$

Let us now arrange the numbers[11] λ_i, k_j^i, and l_i into a pattern in the following way (see Fig. 1):

$$\begin{bmatrix} \lambda_2 & \lambda_3 & \lambda_4 \\ \lambda_1 & k_1^1 & k_2^1 \\ & l_1 & l_2 \end{bmatrix}. \tag{3}$$

From the inequalities (2), a GZ-type condition follows, namely,that every number in the pattern (3) is less than or equal to the number above left, and greater than or equal to the number above it. Further, from (iii) (see Fig. 1) there follows

$$\begin{gathered} \lambda_2 - k_1^1 \leq \lambda_1 - l_1, \\ \lambda_2 + \lambda_3 - k_1^1 - k_2^1 \leq \lambda_1 + k_1^1 - l_1 - l_2. \end{gathered} \tag{4}$$

Now let us generalize to arbitrary n and m. We have the following pattern:

$$\begin{bmatrix} \lambda_m & \lambda_{m+1} & \lambda_{m+2} & \cdots & \lambda_{m+n} \\ \lambda_{m-1} & k_1^{m-1} & k_2^{m-1} & \cdots & k_n^{m-1} \\ \lambda_{m-2} & k_1^{m-2} & k_2^{m-2} & \cdots & k_n^{m-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \lambda_1 & k_1^1 & k_2^1 & \cdots & k_n^1 \\ & l_1 & l_2 & \cdots & l_n \end{bmatrix}. \quad (5)$$

Again, from (i) and (ii) we can derive a "betweenness condition." Next, we introduce the notation

$$S_j^i = \lambda_i + \sum_{p=1}^{j-1} k_p^i - \sum_{p=1}^{j} k_p^{i-1}; \quad i = 1, 2, \cdots, m,$$
$$j = 1, 2, \cdots, n, \quad (6)$$
$$k_i^0 \equiv l_i \quad \text{and} \quad k_i^m \equiv \lambda_{m+i}.$$

In other words, S_j^i is the sum of j first numbers in the $(i+1)$th row from the bottom minus the sum of numbers immediately below right. Corresponding to (4), we have now the conditions

$$S_j^{i+1} \leq S_j^i, \quad i = 1, 2, \cdots, m-1; j = 1, 2, \cdots, n. \quad (7)$$

The highest weights of the subgroup $U(m)$ are given by

$$j_i = S_n^i + k_n^i. \quad (8)$$

Let us summarize: To every possible way of constructing the diagram $[\lambda_i]$ from $[l_i]$ and $[j_i]$, through the procedure explained immediately after Eq. (1), there exists one and only one way to fill the pattern (5) subject to the above mentioned conditions and vice versa; from this follows

(a) the pattern (5) together with the GZ patterns for the subgroups $U(n)$ and $U(m)$ furnishes a solution to the labeling problem in the reduction[12]

$$U(m+n) \downarrow U(m) \times U(n);$$

(b) the multiplicity $m_{lj\lambda}$ is just the number of ways to fill up (5) with $[l_i]$, $[j_i]$, and $[\lambda_i]$ fixed.

3. SOME SPECIAL CASES

A. $U(n+2) \downarrow U(n) \times U(2)$

In this case, there is only one "multiplicity row" in our pattern:

$$\begin{bmatrix} \lambda_2 & \lambda_3 & \lambda_4 & \cdots & \lambda_{n+2} \\ \lambda_1 & k_1^1 & k_2^1 & \cdots & k_n^1 \\ & l_1 & l_2 & \cdots & l_n \end{bmatrix}. \quad (9)$$

For the representation $[l_i]$ of $U(n)$ to appear in $[\lambda_i]$, one can easily see that the conditions

$$\lambda_i \geq l_i \geq \lambda_{i+2} \quad (10)$$

are necessary and sufficient. The conditions for j_i (with the l_i fixed) are more complicated. Note that j_2 is not independent: From (8) there follows

$$j_2 = \sum \lambda_i - \sum l_i - j_1. \quad (11)$$

We have

$$(j_1)_{\max} = \lambda_1 + \sum_{i=1}^{n} (\min\{l_i, \lambda_{i+1}\} - l_i). \quad (12)$$

One can also derive a lower bound for j_1, but, because of (7), it is very complicated and is therefore not repeated here. According to (b) and (7), $m_{lj\lambda}$ is given by

$$m_{lj\lambda} = \sum_{k_1{}^1=\max\{l_2,\lambda_3,-S_1{}^1+\lambda_2\}}^{\min\{l_1,\lambda_2\}} \; \sum_{k_2{}^1=\max\{l_3,\lambda_4,S_1{}^2-S_2{}^1+\lambda_3\}}^{\min\{l_2,\lambda_3\}} \cdots \sum_{k_{n-1}{}^1=\max\{l_n,\lambda_{n+1},S_{n-2}{}^2-S_{n-1}{}^1+\lambda_n\}}^{\min\{l_{n-1},\lambda_n\}} \; \sum_{k_n{}^1=\max\{l_1,\lambda_{n+2},S_{n-1}{}^2-S_n{}^1+\lambda^{1\prime}{}_{n+1}\}}^{\min\{l_n,\lambda_{n+1}\}}. \quad (13)$$

$(j_i)_{\min}$ is the smallest value for which $m_{lj\lambda} > 0$.

B. $\lambda_4 = \lambda_5 = \cdots = \lambda_{n+m} = 0$

Now the *nonzero corner* of the pattern (5) is simply

$$\begin{bmatrix} \lambda_3 & & & \\ \lambda_2 & k_1^2 & & \\ \lambda_1 & k_1^1 & k_2^1 & \\ & l_1 & l_2 & l_3 \end{bmatrix}. \quad (14)$$

Because of (8) only one of the k_j^i is independent; we may take this as $k_1^1 \equiv x$. From (8), it also follows that

$$\sum \lambda_i = \sum l_i + \sum j_i. \quad (15)$$

Using (8), we read the "betweenness conditions" as

$$l_3 \leq \lambda_2 + \lambda_3 - j_2 - j_3 - x \leq \min\{\lambda_3 - j_3, l_2\},$$
$$\max\{l_2, \lambda_3 - j_3\} \leq x \leq \min\{\lambda_2, l_1\}, \quad \lambda_1 \geq l_1. \quad (16)$$

From (7), it follows that

$$\lambda_2 - x \geq j_3, \quad \lambda_1 - l_1 \geq \lambda_2 - x,$$
$$\lambda_1 - l_1 - l_2 \geq j_2 - x. \quad (17)$$

From (16) and (17), we can reduce the multiplicity:

$$m_{lj\lambda} = \min\{l_1, \lambda_2 - j_3, \lambda_2 + \lambda_3 - j_2 - j_3 - l_3\}$$
$$- \max\{l_2, \lambda_3 - j_3, \lambda_2 + l_1 - \lambda_1, l_1 - \lambda_1 + l_2 + j_2, \lambda_2 - j_2, l_3 + j_3 - \lambda_3 + l_1\} + 1. \quad (18)$$

If $m_{lj\lambda} \leq 0$, that means that the representation $[j_i] \times [l_i]$ does not occur in $[\lambda_i]$.[13]

C. $U(6)U \downarrow (3) \times U(3)$

In this case, the calculations are already very tedious. Finding the multiplicities is equivalent to solving a large set of inequalities, namely,(7) and the "betweenness conditions." That is why it is appropriate to use an automatic data machine, especially when handling a big class of representations. Using the Elliott 803 computer, we calculated the multiplicities for all representations with $\lambda_1 \leq 3$ and $\lambda_6 = 0$ [if we restrict our attention to the subgroup $SU(6)$, we can always put $\lambda_6 = 0$]. There are 55 such representations. In only ten cases were there $m_{lj\lambda}$ bigger than one. These are (the last number is the dimension)

$$(3, 2, 1, 1, 1, 0; 384), \quad (3, 2, 2, 2, 1, 0; 384),$$
$$(3, 2, 2, 1, 1, 0; 540), \quad (3, 3, 2, 2, 1, 0; 840),$$
$$(3, 2, 1, 1, 0, 0; 840), \quad (3, 2, 1, 0, 0, 0; 896),$$
$$(3, 3, 3, 2, 1, 0; 896), \quad (3, 2, 2, 1, 0, 0; 1050),$$
$$(3, 3, 2, 1, 1, 0; 1050), \quad (3, 3, 2, 1, 0, 0; 1960).$$

4. CONCLUSION

In this paper, we have solved the labeling problem in $U(n+m) \downarrow U(n) \times U(m)$. The matrix elements of the generators remain to be calculated. One method would be to modify somehow the boson operator calculus used in Refs. 3 and 4. The difficulty is that the success of this method in the case $U(n) \supset U(n-1)$ is intimately connected with the fact that this reduction is multiplicity free; in the case $U(n+m) \downarrow U(n) \times U(m)$, there exists a large amount of arbitrariness in defining the lowering operators for the highest weights of the subgroup $U(n) \times U(m)$. Perhaps the only sensible way is to try to guess the correct form of the matrix elements and then prove the commutation relations.

The knowledge of matrix elements in the compact case is clearly important also in constructing representations of the noncompact groups $U(n, m)$. The question is: What representations of $U(n, m)$ can be achieved through analytic continuation from those of $U(n+m)$ in $U(n) \times U(m)$ basis?

We hope to discuss the cases $SO(n+m) \downarrow SO(n) \times SO(m)$ and $Sp(2n+2m) \downarrow Sp(2n) \times Sp(2m)$ in a future publication.

ACKNOWLEDGMENT

We wish to thank Professor P. Tarjanne for critical reading of the manuscript.

[1] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825, 1017 (1950).

[2] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methuen, London, 1931), p. 390.

[3] J. G. Nagel and M. Moshinsky, J. Math. Phys. **6**, 682 (1965).

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

[5] S. C. Pang and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).

[6] M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).

[7] Professor J. Niederle has called my attention to a paper by C. R. Hagen and A. J. MacFarlane [J. Math. Phys. **6**, 1366 (1965)] who treat in detail the cases $SU(6) \downarrow SU(4) \times SU(2)$ and $SU(6) \downarrow U(1) \times SU(2) \times SU(4)$, using a different approach.

[8] G. de B. Robinson, *Representation Theory of the Symmetric Group* (Edinburgh U.P., Edinburgh, 1961).

[9] D. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford U.P., London, 1950).

[10] In this case, the multiplicity has been solved by Racah [Rev. Mod. Phys. **21**, 494 (1949)].

[11] Note that

$$j_1 = (\lambda_1 + k_1^1 + k_2^1) - (l_1 + l_2),$$
$$j_2 = (\lambda_2 + \lambda_3 + \lambda_4) - (k_1^1 + k_2^1).$$

[12] If we put $n = m$ and $\lambda_{n+1} = \lambda_{n+2} = \cdots = \lambda_{n+m} = 0$, we also have a complete set of labels for the reduction of the direct product $[l_i] \times [j_i]$ of $U(n)$ representations. The pattern (5) reduces to (we have suppressed the zeros)

$$\begin{bmatrix} \lambda_n & & & & \\ \lambda_{n-1} & k_1^{n-1} & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \lambda_1 & k_1^1 & k_2^1 & \cdots & k_n^1 \\ & l_1 & l_2 & \cdots & l_n \end{bmatrix}.$$

Note that not all the k_j^i are independent; we have the constraints of Eq. (8). If we add to the left the labels m_{ij} inside the representation $[\lambda_i]$ and turn around the pattern, we have a pattern for the reduction of the direct product $[\lambda_i] \times [j_i]$. This kind of pattern has been given also by Baird and Biedenharn [J. Math. Phys. **5**, 1730 (1964)] for classification of tensor operators in $SU(n)$.

[13] According to Sec. 2, $m_{lj\lambda}$ is also the multiplicity of $[\lambda_i]$ in the direct product $[l_i] \times [j_i]$ of two representations of $U(3)$. Also note that, if $\lambda_3 = 0$, then $m_{lj\lambda} \leq 1$.

Einstein Spaces with Symmetry Groups*

R. P. KERR
Department of Mathematics, University of Texas at Austin, Texas 78712
AND
G. C. DEBNEY, JR.
Department of Mathematics, Virginia Polytechnic Institute, Blacksburg, Va. 24061

(Received 16 May 1969; Revised Manuscript Received 9 April 1970)

In this paper, we construct all possible groups of motion (symmetry groups) for empty Einstein spaces admitting a diverging, geodesic,and shear-free ray congruence. (Minkowski space is excluded throughout the discussion.) It is proved that any such Einstein space cannot admit a symmetry group with dimension greater than four. Although the field equations are not solved completely for spaces with groups of dimension one or two, a generalization of the Kerr spinning-mass solution is obtained from the 2-dimensional class. It is shown that all such spaces with 4-dimensional symmetry groups are well known: Schwarzschild, NUT (Newman, Unti, and Tamborino), and a particular hypersurface orthogonal Kerr–Schild metric. The only member of these spaces admitting a 3-dimensional symmetry group is a Petrov Type III hypersurface orthogonal metric.

1. INTRODUCTION

In a previous paper by Debney, Kerr, and Schild[1] (hereafter referred to as DKS), we derived the field equations for empty Einstein spaces containing a diverging shear-free ray congruence. In this paper, we study the possible groups of motion $\mathcal{S}$ which such a metric ($\mathfrak{s}$, say) admits. We find that, when the dimension of $\mathcal{S}$ is greater than two, it is possible to find all possible solutions but that, for lower dimensional symmetry groups, the field equations reduce to ordinary differential equations which we have not always been able to solve completely.

We follow the notation of DKS so that $\{\mathbf{e}_a\}\{\epsilon^a\}$ are dual bases in the tangent and cotangent planes, respectively, g_{ab} is the metric tensor, and the $\{\omega_{ab}\}$ are the connection forms

$$\omega_{ab} = g_{ac}\Gamma^c_{bd}\epsilon^d = -e_{a\mu;\nu}e^\mu_b\,dx^\nu. \tag{1.1}$$

It is assumed that $\{\mathbf{e}_a\}$ is a null basis, so that the metric is

$$(d\tau)^2 = 2\epsilon_1\epsilon_2 + 2\epsilon_3\epsilon_4. \tag{1.2}$$

Both ϵ_3 and ϵ_4 are real vectors, while ϵ_1 and ϵ_2 are complex conjugates $\epsilon_1 = \bar{\epsilon}_2$. (We use a bar to denote complex conjugation. With the dual basis $\{\epsilon^a\}$, we use the convention $\epsilon_a \equiv g_{ab}\epsilon^b$.)

The vector $\mathbf{e}_4$ is assumed to be a multiple Debever vector, and so the only independent nonzero components of the curvature tensor are

$$C^{(3)} = 2R_{4231},$$
$$C^{(2)} = 2^{\frac{1}{2}}(R_{1231} + R_{3431}), \tag{1.3}$$
$$C^{(1)} = 2R_{3131}.$$

[The notation here is that of Goldberg and Sachs[2] with the exceptions that (i) ϵ_3 and ϵ_4 are exchanged and (ii) numerical multipliers are not the same since there is always a freedom of choice of the basis for the space of 2-forms Λ^2(cotangent plane).] The proper orthochronous Lorentz transformations preserving the direction of $\mathbf{e}_4$ can be written as

$$\mathbf{e}_1^* = e^{-iB}(\mathbf{e}_1 + \gamma\mathbf{e}_4),$$
$$\mathbf{e}_3^* = e^{-A}(\mathbf{e}_3 - \bar{\gamma}\mathbf{e}_1 - \gamma\mathbf{e}_2 - \gamma\bar{\gamma}\mathbf{e}_4), \tag{1.4}$$
$$\mathbf{e}_4^* = e^{A}\mathbf{e}_4,$$

where A and B are real and γ is complex. [Throughout this paper, we endeavor to use Latin (Greek) letters for real (complex) functions.]

The Cartan structural formulas are

$$d\epsilon^a + \omega^a_c \wedge \epsilon^c = 0,$$
$$d\omega^a_b + \omega^a_c \wedge \omega^c_b = \tfrac{1}{2}R^a_{bcd}\epsilon^c \wedge \epsilon^d.$$

Since the g_{ab} are assumed constant throughout this paper, the ω_{ab} are antisymmetric:

$$dg_{ab} = 0, \quad \omega_{ab} = -\omega_{ba}.$$

As was seen in DKS, the most important of the Cartan equations is

$$d\omega_{42} + \omega_{42} \wedge (\omega_{12} + \omega_{34}) = C^{(3)}\epsilon_4 \wedge \epsilon_2. \tag{1.5}$$

The components of ω_{42} are the optical scalars of Sachs.[3] In particular,

$$\Gamma_{241} = z = \text{complex divergence},$$
$$\Gamma_{242} = \sigma = \text{shear},$$
$$\Gamma_{244} = \kappa = \text{geodesy}.$$

The Goldberg–Sachs theorem[2] states that, if $\mathfrak{s}$ is an empty Einstein space and $\mathbf{e}_4$ is a multiple Debever vector, then it must be geodesic and shear free. Consequently,

$$\omega_{42} = -z\epsilon_2 + \Gamma_{423}\epsilon_4, \tag{1.6}$$

and so, from Eqs. (1.5) and (1.6),

$$d\omega_{42} \wedge \omega_{42} = C^{(3)}\epsilon_4 \wedge \epsilon_2 \wedge \omega_{42} = 0.$$

From this, there exist two locally smooth complex functions ϕ and ζ such that

$$d\omega_{42} = -e^{\phi}\, d\zeta.$$

Under a tetrad transformation (1.4), ω_{42} transforms as

$$\omega_{42}^* = e^{A+iB}\omega_{42},$$

and so, by taking $A + iB + \phi = 0$, we can arrange that ω_{42} is a perfect differential:

$$\omega_{42} = -d\zeta. \qquad (1.7)$$

Also, since we are only investigating spaces for which the complex divergence is nonzero and since

$$\Gamma_{423}^* = e^{iB}(\Gamma_{423} - \bar{\gamma} z), \quad z^* = e^{A} z, \qquad (1.8)$$

it is possible to transform Γ_{423} to zero (by solving for γ), so that

$$\omega_{24} = d\zeta = z\epsilon_2. \qquad (1.9)$$

The group of transformations which preserve this condition must satisfy

$$\omega_{24}^* = e^{A+iB}\, d\zeta = d\zeta^*,$$

which gives

$$\zeta^* = \Phi(\zeta), \qquad (1.10)$$
$$e^{A+iB} = \partial_\zeta \Phi = \Phi_\zeta, \quad \gamma = 0,$$

where Φ is an analytic function of ζ.

The two functions ζ and $\bar{\zeta}$ are used as local coordinates. As was proved in DKS, it is possible to introduce a real function u satisfying

$$\mathbf{e}_3 u \equiv e_3{}^{\mu} u_{,\mu} = 1, \quad \mathbf{e}_4 u = 0.$$

The Debever vector ϵ_4 can then be written as

$$\epsilon_4 = du + \Omega\, d\zeta + \bar{\Omega}\, d\bar{\zeta}, \qquad (1.11)$$

where Ω is a smooth function. We do not use the fourth coordinate of DKS but, instead, that of Kerr,[4] $v = \mathrm{Re}\,(z^{-1})$, since its transformation properties are simpler. It was shown in Ref. 4 and DKS that the metric can be written as a function of Ω and another function μ (roughly speaking, it is the complex mass):

$$\epsilon_1 = (v + \Delta)\, d\zeta,$$
$$\epsilon_3 = dv - 2\,\mathrm{Re}\,\{[(v - \Delta)\dot{\Omega} + \eth\Delta]\, d\zeta\} + \mathrm{Re}\,(\bar{\eth}\dot{\Omega} + \mu z)\epsilon_4. \qquad (1.12)$$

The operator $\eth$ is given by

$$\eth = \partial_\zeta - \Omega\partial_u; \qquad (1.13)$$

a "dot" denotes differentiation with respect to u, e.g., $\dot{\Omega} = \partial_u\Omega$, and z and Δ are given by

$$z = (v + \Delta)^{-1},$$
$$\Delta = i\,\mathrm{Im}\,(\bar{\eth}\Omega) = -\bar{\Delta}. \qquad (1.14)$$

[$\eth \equiv D$ in DKS; i.e., $e_1{}^{\mu}\partial_\mu = z(\partial - \beta\partial_v)$, where $\epsilon_3 = dv + \beta\, d\zeta + \bar{\beta}\, d\bar{\zeta} + H\epsilon_4$. Hence, as an operator on functions independent of v, $\eth = z^{-1}\mathbf{e}_1$.] The v dependence of the metric is shown explicitly; in particular, both μ and Ω are independent of this coordinate:

$$\mu = \mu(\zeta, \bar{\zeta}, u), \quad \Omega = \Omega(\zeta, \bar{\zeta}, u).$$

These functions have to satisfy the field equations

$$\eth\mu = 3\dot{\Omega}\mu, \qquad (1.15a)$$
$$\mathrm{Im}\,(\mu - \bar{\eth}\,\bar{\eth}\eth\Omega) = 0, \qquad (1.15b)$$
$$\partial_u(\mu - \bar{\eth}\,\bar{\eth}\eth\Omega) = |\partial_u\eth\Omega|^2. \qquad (1.15c)$$

It was also shown in DKS that the tetrad is parallel-propagated along the Debever congruence $\mathbf{e}_4$, but we do not need this result. Finally, the independent components of the conformal tensor are

$$C^{(3)} = \mu z^3,$$
$$C^{(2)} = -(\bar{\eth}\partial_u\eth\Omega)z^2 + (\text{terms} = 0 \text{ if } C^{(3)} = 0),$$
$$C^{(1)} = (\partial_u\partial_u\eth\Omega)z + (\text{terms} = 0 \text{ if } C^{(3)} = C^{(2)} = 0), \qquad (1.16)$$

and so $\mathcal{S}$ is flat iff the following equations are satisfied:

$$R_{abcd} = 0 \Leftrightarrow \mu = \bar{\eth}\partial_u\eth\Omega = \partial_u\partial_u\eth\Omega = 0. \qquad (1.17)$$

Henceforth, it is assumed that $\mathcal{S}$ is nonflat, so that at least one of the conditions on the right of Eq. (1.17) is not satisfied.

2. THE GROUP $\mathcal{C}$

The coordinate system and tetrad is not defined uniquely and, in fact, cannot be if the manifold admits a symmetry. In this section, we determine the residual group of coordinate and tetrad transformations which preserve the coordinate conditions. Since we eventually restrict our attention to infinitesimal transformations, we only consider those which lie in the identity component, $\mathcal{C}$ (say).

Suppose $\{x^{*a}, \mathbf{e}_a^*\}$ is another set of coordinates and tetrad satisfying the conditions of the last section. It has already been seen that $\zeta^* = \Phi(\zeta)$ and that $\mathbf{e}_a^*$ is related to $\mathbf{e}_a$ by Eq. (1.4), with $\gamma = 0$ and $A + iB$ given by Eq. (1.10). The defining equation for u, Eq. (1.11), gives

$$\epsilon_4^* = du^* + \Omega^*\, d\zeta^* + \bar{\Omega}^*\, d\bar{\zeta}^* = |\Phi_\zeta|\, \epsilon_4. \qquad (2.1)$$

The general solution of this equation is $u^* = |\Phi_\zeta| (u + S)$, where S is a real function of $(\zeta, \bar{\zeta})$. From Eq. (1.8), $z^* = |\Phi_\zeta| z$, and so, since $v = \text{Re}(z^{-1})$, we have $v^* = |\Phi_\zeta|^{-1} v$. The transformation properties of Ω and μ can be computed from Eq. (2.1) and the invariance of $C^{(3)} = \mu z^3$. Finally, $\mathcal{C}$ is the group of transformations $x \to x^*$ with

$$\begin{aligned}
\zeta^* &= \Phi(\zeta),\\
u^* &= |\Phi_\zeta| (u + S(\zeta, \bar{\zeta})),\\
v^* &= |\Phi_\zeta|^{-1} v,\\
\epsilon_1^* &= (|\Phi_\zeta|/\Phi_\zeta)\epsilon_1,\\
\epsilon_3^* &= |\Phi_\zeta|^{-1} \epsilon_3,\\
\epsilon_4^* &= |\Phi_\zeta| \epsilon_4,
\end{aligned} \tag{2.2}$$

and the functions μ, Ω, and Δ transform as follows:

$$\begin{aligned}
\mu^* &= |\Phi_\zeta|^{-3} \mu,\\
\Omega^* &= (|\Phi_\zeta|/\Phi_\zeta)[\Omega - S_\zeta - \tfrac{1}{2}(\Phi_{\zeta\zeta}/\Phi_\zeta)(u + S)], \quad (2.3)\\
\Delta^* &= |\Phi_\zeta|^{-1} \Delta.
\end{aligned}$$

3. SYMMETRIES OF $\mathcal{\$}$

Let $\mathcal{S}$ be the identity component of the group of symmetries of $\mathcal{\$}$. If we interpret these as coordinate transformations, rather than point transformations, then $\mathcal{S}$ is the set of transformations $x \to x^*$ for which

$$g^*_{\mu\nu}(x^*) = g_{\mu\nu}(x^*). \tag{3.1}$$

It is clear that any transformation of $\mathcal{S}$ must preserve the coordinate and tetrad conditions of Sec. 2, and so $\mathcal{S}$ must be a subgroup of $\mathcal{C}$; but the converse is not true.

Lemma 3.1: $\mathcal{S}$ is that subgroup of $\mathcal{C}$ for which

$$\begin{aligned}
\Omega^*(x^*) &= \Omega(x^*),\\
\mu^*(x^*) &= \mu(x^*).
\end{aligned} \tag{3.2}$$

Proof: Since $(d\tau)^2$ is given explicitly as a function of μ, Ω, and their derivatives, those members of $\mathcal{C}$ which satisfy Eq. (3.2) belong to $\mathcal{S}$. Conversely, $g_{\zeta v} = \Omega$, and so the first of Eq. (3.2) follows from Eq. (3.1). It follows from this that $\eth\dot{\Omega}$, and therefore Δ and z, are invariant. The second of Eq. (3.2) now follows from the invariance of g_{uu}.

Let us suppose that $x \to x^*(x, t)$ is a 1-parameter group of motions,

$$\begin{aligned}
\zeta^* &= \psi(\zeta; t),\\
u^* &= |\psi_\zeta| [u + T(\zeta, \bar{\zeta}; t)],\\
v^* &= |\psi_\zeta|^{-1} v.
\end{aligned} \tag{3.3}$$

Since $x^*(0) = x$, the initial values of ψ and T are

$$\psi(\zeta; 0) = \zeta, \quad T(\zeta, \bar{\zeta}; 0) = 0.$$

The corresponding infinitesimal transformation, $\mathbf{K} = K^\mu \partial/\partial x^\mu$, is given by

$$K^\mu = \left[\frac{\partial x^{*\mu}}{\partial t}\right]_{t=0}. \tag{3.4}$$

If we define

$$\alpha(\zeta) = \left.\frac{\partial \psi}{\partial t}\right|_{t=0},$$

$$R(\zeta, \bar{\zeta}) = \left.\frac{\partial T}{\partial t}\right|_{t=0},$$

then the first part of the following lemma follows from Eqs. (3.3) and (3.4).

Lemma 3.2: A vector $\mathbf{K}$ is an infinitesimal transformation of the group $\mathcal{S}$ iff it can be written as

$$\mathbf{K} = \alpha\partial_\zeta + \bar{\alpha}\partial_{\bar{\zeta}} + \text{Re}(\alpha_\zeta)(u\partial_u - v\partial_v) + R\partial_u, \tag{3.5}$$

where α and R satisfy the equations

$$\text{(I)} \quad R_\zeta + \tfrac{1}{2}\alpha_{\zeta\zeta} u + \mathbf{K}\Omega + \tfrac{1}{2}(\alpha_\zeta - \bar{\alpha}_{\bar{\zeta}})\Omega = 0, \tag{3.6a}$$

$$\text{(III)} \quad \mathbf{K}\mu + 3\,\text{Re}(\alpha_\zeta)\mu = 0. \tag{3.6b}$$

Proof: From Lemma 3.1, and the general theory of Killing vectors, $\mathbf{K}$ is a Killing vector iff it satisfies

$$\frac{d}{dt}[\mu^*(x^*) - \mu(x^*)] = 0, \tag{3.7a}$$

$$\frac{d}{dt}[\Omega^*(x^*) - \Omega(x^*)] = 0. \tag{3.7b}$$

The lemma then follows from Eq. (2.2) by a straightforward calculation. For instance, from Eq. (3.7a),

$$\begin{aligned}
0 &= \frac{d}{dt}[|\psi_\zeta|^{-3} \mu(x) - \mu(x^*)]|_{t=0}\\
&= -3\,\text{Re}(\alpha_\zeta) \cdot \mu - \frac{\partial x^{*\nu}}{\partial t}\frac{\partial \mu}{\partial x^\nu}\\
&= -3\,\text{Re}(\alpha_\zeta) \cdot \mu - \mathbf{K}\mu.
\end{aligned}$$

If Eq. (3.6a) is differentiated with respect to u and $\dot{R} = \dot{\alpha} = 0$ used, we have

$$\text{(II)} \quad \tfrac{1}{2}\alpha_{\zeta\zeta} + \mathbf{K}\dot{\Omega} + \alpha_\zeta\dot{\Omega} = 0, \tag{3.6a'}$$

which may then be substituted back into Eq. (3.6a). In order to reduce the Killing equations to a standard form, let $\beta = \alpha_\zeta$, so that the unknowns are $y^A = (\alpha, \bar{\alpha}, \beta, \bar{\beta}R)$. From Eqs. (3.6a) and (3.6a'), the Killing equations are equivalent to

$$\begin{aligned}
\alpha_\zeta = \beta, \quad \beta_\zeta &= -2(\mathbf{K}\dot{\Omega} + \beta\dot{\Omega}), \quad R_\zeta = Q(y^A, x),\\
\alpha_{\bar{\zeta}} &= 0, \quad \beta_{\bar{\zeta}} = 0, \quad R_{\bar{\zeta}} = \bar{Q},\\
\alpha_u &= 0, \quad \beta_u = 0, \quad R_u = 0,
\end{aligned} \tag{3.8}$$

$$Q = \mathbf{K}(u\dot{\Omega} - \Omega) + \tfrac{1}{2}(\beta - \bar{\beta})(u\dot{\Omega} - \Omega) - R\dot{\Omega}.$$

Since the y^A, and Ω and μ, are independent of v, the Killing equations are essentially 3 dimensional, and Eqs. (3.6a) and (3.6a′) are equivalent to

$$\partial_\mu y^A = Q^A{}_\mu(y^B; x),$$

subject to the linear constraint (3.6b). The $Q^A{}_\mu$ are given in Eq. (3.8), and are linear in the y^A. These equations are then in standard form, and their first integrability conditions are given by

$$Q^A{}_{[\mu,\nu]} = 0. \tag{3.9}$$

The higher-order equations follow by differentiating Eqs. (3.6b) and (3.9) and substituting for $\partial_\mu y^A$ from Eq. (3.8). In order to compute these, the following commutation relations are used:

$$[\partial_u, \mathbf{K}] = \mathrm{Re}\,(\alpha_\zeta)\partial_u, \quad [\eth, \mathbf{K}] = \alpha_\zeta \eth + (\cdots)\partial_v, \tag{3.10}$$

where the coefficient of ∂_v is irrelevant, since all functions considered are independent of v. The Killing equations and first integrability conditions are

$$\text{(I)} \quad R_\zeta + \mathbf{K}(\Omega - u\dot\Omega) + \tfrac{1}{2}(\alpha_\zeta - \bar\alpha_{\bar\zeta})(\Omega - u\dot\Omega) + R\dot\Omega = 0, \tag{3.11}$$

$$\text{(II)} \quad \alpha_{\zeta\zeta} + 2\mathbf{K}\dot\Omega + \alpha_\zeta\dot\Omega = 0, \tag{3.12}$$

$$\text{(III)} \quad \mathbf{K}\mu + 3\,\mathrm{Re}\,(\alpha_\zeta)\mu = 0, \tag{3.13}$$

$$\text{(IVa)} \quad \mathbf{K}\dot\mu + 4\,\mathrm{Re}\,(\alpha_\zeta)\dot\mu = 0, \tag{3.14a}$$

$$\text{(IVb)} \quad \mathbf{K}(\bar\eth\dot\Omega) + 2\,\mathrm{Re}\,(\alpha_\zeta)\bar\eth\dot\Omega = 0, \tag{3.14b}$$

$$\text{(IVc)} \quad \mathbf{K}\Delta + \mathrm{Re}\,(\alpha_\zeta)\Delta = 0, \tag{3.14c}$$

$$\text{(IVd)} \quad \mathbf{K}\ddot\Omega + \tfrac{1}{2}(3\alpha_\zeta + \bar\alpha_{\bar\zeta})\ddot\Omega = 0. \tag{3.14d}$$

The dimension of the group $\mathcal{S}$ is $(5 - s)$ where 5 is the number of y^A and s is the number of independent integrability conditions [including (3.13)]. A partial result is

Lemma 3.3: The dimension of the group $\mathcal{S}$ satisfies the inequality

$$\dim(\mathcal{S}) \le 5 - (\text{number of known integrability conditions}).$$

In particular, $\dim(\mathcal{S}) = 5$ iff the first-order integrability conditions are identically zero, i.e.,

$$\dim(\mathcal{S}) = 5 \Leftrightarrow \ddot\Omega = \bar\eth\dot\Omega = \mu = \Delta = 0.$$

This gives $\Omega = \rho(\zeta)u + \sigma(\zeta, \bar\zeta)$, and so

$$\partial_u \bar\eth\Omega = (\rho_\zeta - \rho^2)(\zeta).$$

Comparing this with Eq. (1.17), we see that the space is flat, which contradicts the assumption about $\mathcal{E}$.

Lemma 3.4: If $\mathcal{E}$ is a nonflat, algebraically special empty Einstein space with diverging Debever congruence, then

$$\dim(\mathcal{S}) \le 4.$$

We need the transformation properties of $\mathbf{K}$ under an element of $\mathcal{C}$, henceforth called a (Φ, S)-transformation. If (α^*, R^*) are the transformed (α, R) for $\mathbf{K}$, then Eqs. (2.1) and (3.5) give

$$\alpha^* = \Phi_\zeta \alpha, \quad R^* = |\Phi_\zeta|\,[R - \mathrm{Re}\,(\alpha_\zeta)\cdot S + \mathbf{K}S]. \tag{3.15}$$

We observe that, if $\mathbf{K}$ is a particular Killing vector, then the differential equations $R^* = 0$ and $\alpha^* = 1$ can be solved for (Φ, S) if $\alpha \ne 0$. If $\alpha = 0$, then the Killing vector has the simple form $\mathbf{K} = R\partial_u$,

Lemma 3.5: If K is a particular Killing vector, then the coordinate system can be chosen so that it has one of the following two canonical forms:

$$\text{(i)}\ \mathbf{K} = \partial_\zeta + \partial_{\bar\zeta} \quad \text{or} \quad \text{(ii)}\ \mathbf{K} = R\partial_u.$$

Furthermore, a vector of one type cannot be transformed to the other.

4. KILLING VECTORS OF THE TYPE $R\partial_u$

Let us prove the following theorem.

Lemma 4.1: $\mathcal{E}$ admits a Killing vector of the type $\mathbf{K} = e^{-p}\partial_u$ iff

$$\ddot\Omega = \dot\Delta = \dot\mu = 0. \tag{4.1}$$

Furthermore, $p = p(\zeta, \bar\zeta)$.

Proof: If $e^{-p}\partial_u$ is a Killing vector, then Eqs. (3.12), (3.13), and (3.14c) give Eq. (4.1). Conversely, if Eq. (4.1) is satisfied, then (3.11) and (3.13), with $\mathbf{K} = e^{-p}\partial_u$, reduce to

$$p_\zeta = \dot\Omega, \quad \dot p = 0, \tag{4.2}$$

whose integrability conditions are just

$$\ddot\Omega = 0, \quad \bar\eth\dot\Omega - \eth\dot{\bar\Omega} = p_{\zeta\bar\zeta} - p_{\bar\zeta\zeta} = 0;$$

that is, $\ddot\Omega = \dot\Delta = 0$, and so $\mathcal{E}$ admits a Killing vector of the required type.

If $e^{-p}\partial u$ is a Killing vector of $\mathcal{E}$, then it is convenient to introduce a slightly different coordinate system. We define

$$s = e^p u, \quad r = e^{-p} v = e^{-p}\,\mathrm{Re}\,(z^{-1}), \tag{4.3}$$

and then the metric reduces to the simpler form

$$\tfrac{1}{2}(d\tau)^2 = (r^2 + d^2)(e^{2p}\, d\zeta\, d\bar{\zeta}) + [dr + i(d_{\bar{\zeta}}\, d\bar{\zeta} - d_{\zeta}\, d\zeta)](\kappa) + \{R^{(2)} + \mathrm{Re}\,(m/(r + id))\}\kappa^2, \tag{4.4}$$

where

$$\begin{aligned} \kappa &= e^p \epsilon_4 = ds + \Lambda\, d\zeta + \bar{\Lambda}\, d\bar{\zeta}, \\ \Lambda &= e^p(\Omega - p_{\zeta} u), \\ d &= -i\Delta e^{-p} = e^{-2p}\,\mathrm{Im}\,(\Lambda_{\bar{\zeta}}), \\ m &= \mu e^{-3p}, \end{aligned} \tag{4.5}$$

and $R^{(2)}$ is the 2-curvature of the metric tensor $e^{2p}\, d\zeta\, d\bar{\zeta}$,

$$R^{(2)} = e^{-2p} p_{\zeta\bar{\zeta}}.$$

From Eq. (4.2), Λ (and therefore the metric) is independent of s, and so is invariant under the Killing vector $\partial_s = e^{-p}\partial_u$, as it should be. We can now tighten the allowed coordinate transformations $\mathcal{C}$ so that this independence with respect to s is preserved. If $\hat{\mathcal{C}}$ is the group of transformations $(s, r, \zeta) \to (s^*, r^*, \zeta^*)$ preserving all coordinate conditions, then

$$e^{-p^*}(ds^* + \Lambda^*\, d\zeta^* + \bar{\Lambda}^*\, d\bar{\zeta}^*) = |\Phi_{\zeta}|\, e^{-p}(ds + \Lambda\, d\zeta + \bar{\Lambda}\, d\bar{\zeta}), \tag{4.6}$$

with $\zeta^* = \Phi(\zeta)$. Since $\dot{\Lambda}^* = \dot{\Lambda} = 0$, this equation can be integrated to give $s^* = C_0(s + A)$, where C_0 is a constant and $A = A(\zeta, \bar{\zeta})$. Hence, from Eq. (4.6),

$$e^{p^*} = |C_0 \Phi_{\zeta}^{-1}|\, e^p, \tag{4.7}$$

and $\hat{\mathcal{C}}$ is given by

$$\hat{\mathcal{C}}: \quad \begin{aligned} \zeta^* &= \Phi(\zeta), \\ s^* &= C_0(s + A), \\ r^* &= C_0^{-1} r. \end{aligned} \tag{4.8}$$

The functions m and Λ transform as

$$m^* = C_0^{-3} m, \quad \Lambda^* = C_0 \Phi_{\zeta}^{-1}(\Lambda - A_{\zeta}). \tag{4.9}$$

As in Sec. 3, if $\mathbf{K}$ is a Killing vector, then it must be the tangent vector to a 1-parameter subgroup of $\mathcal{C}$. We prove the following lemma exactly as we proved Lemma 3.2.

Lemma 4.2: If the manifold $\mathcal{M}$ admits a Killing vector of the type $e^{-p}\partial_u$, then coordinates $(\zeta, \bar{\zeta}, s, r)$ can be introduced so that the metric is given by Eq. (4.4). Furthermore, $\mathbf{K}$ is a Killing vector iff it has the form

$$\mathbf{K} = \alpha\partial_{\zeta} + \bar{\alpha}\partial_{\bar{\zeta}} + a_0(s\partial_s - r\partial_r) + T\partial_s, \tag{4.10}$$

where a_0 is a constant and $T = e^p R$, and if it satisfies the following Killing equations:

$$\text{(II}') \qquad \mathbf{K}p + \mathrm{Re}\,(\alpha_{\zeta}) = a_0, \tag{4.11}$$

$$\text{(I)} \qquad T_{\zeta} + \mathbf{K}\Lambda + (-a_0 + \alpha_{\zeta})\Lambda = 0, \tag{4.12}$$

$$\text{(III)} \qquad \mathbf{K}m + 3a_0 m = 0. \tag{4.13}$$

Proof: It is easily seen that $g^*_{\mu\nu}(x^*) = g_{\mu\nu}(x^*)$ gives $p^*(x^*) = p(x^*)$ and similar equations for Λ and m. If we differentiate these with respect to the parameter t and let $a_0 = \partial_t C_0$, then the theorem follows.

Lemma 4.2 can be derived directly from Lemma 3.2, using (4.11) as a first integral of (3.6a′). From Eq. (4.10), we have

$$[\partial_s, \mathbf{K}] = a_0 \partial_s,$$
$$[\partial_{\zeta}, \mathbf{K}] = \alpha_{\zeta}\partial_{\zeta} + T_{\zeta}\partial_s,$$

which is a commutation relation between two Killing vectors.

We need the field equations for this metric. Substituting Λ and m into Eq. (1.15), we reduce them to

$$\partial_{\bar{\zeta}} m = 0, \tag{4.14a}$$

$$(e^{-2p} p_{\zeta\bar{\zeta}})_{\zeta\bar{\zeta}} = 0, \tag{4.14b}$$

$$\mathrm{Im}\,(m) = e^{-2p}\, d_{\zeta\bar{\zeta}} - 2e^{-2p} p_{\zeta\bar{\zeta}}\, d. \tag{4.14c}$$

From Eq. (4.14a), m is an analytic function of ζ alone, and, from Eq. (4.14b), the 2-curvature $R^{(2)}$ is a harmonic function. The last field equation can be solved for d and, therefore, for Λ.

We now come to the fundamental theorem of this paper. This is needed to find all $\mathcal{M}$ with high-dimensional symmetry groups.

Theorem 4.1: If $\dim(\mathcal{S}) > 2$, then $\mathcal{M}$ admits a special Killing vector of type $e^{-p}\partial_u$.

Proof (by interminable contradiction!) *:* Let us suppose that $\dim(\mathcal{S}) \geq 3$ and so

$$\text{number of independent integrability conditions} \leq 2, \tag{4.15}$$

but that $\mathcal{M}$ does not satisfy Eq. (4.1) and so does not admit a Killing vector of the required type. Consider the set of integrability conditions obtained from Eq. (3.14d) by successive differentiation with respect to u:

$$\begin{aligned} \mathbf{K}\omega + [\alpha_{\zeta} + \mathrm{Re}\,(\alpha_{\zeta})]\omega &= 0, \\ \mathbf{K}\dot{\omega} + [\alpha_{\zeta} + 2\,\mathrm{Re}\,(\alpha_{\zeta})]\dot{\omega} &= 0, \\ \mathbf{K}\ddot{\omega} + [\alpha_{\zeta} + 3\,\mathrm{Re}\,(\alpha_{\zeta})]\ddot{\omega} &= 0, \end{aligned} \tag{4.16}$$

where $\omega = \dot{\Omega}$. We constantly need to consider equations of this type as linear constraints on the unknowns $[R + u\,\mathrm{Re}\,(\alpha_{\zeta}), \alpha, \bar{\alpha}, \alpha_{\zeta}, \bar{\alpha}_{\bar{\zeta}}]$. From Eq. (4.15), the rank of the matrix of the coefficients of these unknowns in Eqs. (4.16) must be less than three. From the coefficients of $(R + u\,\mathrm{Re}\,(\alpha_\zeta), \alpha_\zeta, \bar{\alpha}_{\bar\zeta})$, we have

$$\begin{vmatrix} \dot\omega & \omega & 0 \\ \ddot\omega & \dot\omega & \dot\omega \\ \dddot\omega & \ddot\omega & 2\ddot\omega \end{vmatrix} = 0.$$

This can be rewritten as

$$\text{(a)}\quad \dot\omega = 0 \quad\text{or}\quad \text{(b)}\quad \frac{\partial^2}{\partial u^2}\left(\frac{\omega}{\dot\omega}\right) = 0. \tag{4.17}$$

We prove that, in fact, $\omega = 0$. First, consider the case where $\dot\omega = 0$, so that the Eq. (4.16) becomes

$$\alpha\omega_\zeta + \bar\alpha\omega_{\bar\zeta} + (\tfrac{3}{2}\alpha_\zeta + \tfrac{1}{2}\bar\alpha_{\bar\zeta})\omega = 0,$$
$$\alpha\bar\omega_\zeta + \overline{\alpha\omega}_{\bar\zeta} + (\tfrac{1}{2}\alpha_\zeta + \tfrac{3}{2}\bar\alpha_{\bar\zeta})\bar\omega = 0. \tag{4.18}$$

Differentiating the first of these with respect to ζ and using Eq. (3.6a′), we have

$$\omega^2 R + (\text{function of } \Omega, \alpha, \alpha_\zeta) = 0. \tag{4.19}$$

Since the integrability conditions (4.18) and (4.19) must be dependent, the determinant of the coefficients of $(R, \alpha_\zeta, \bar\alpha_{\bar\zeta})$ is zero, and so $\omega^3\bar\omega = 0$, i.e., $\omega = 0$. Now, let us suppose that $\dot\omega \neq 0$ so that (4.17b) is satisfied:

$$\omega = \ddot\Omega = \gamma(u + \beta)^\sigma, \tag{4.20}$$

where β, γ, and σ are functions of $(\zeta, \bar\zeta)$ and

$$\gamma \neq 0, \quad \sigma \neq 0. \tag{4.21}$$

Let $\beta = a + ib$ and perform an (Φ, S)-transformation with $(\Phi(\zeta), S) = (\zeta, a)$. From Eq. (2.2), $\ddot\Omega^* = \ddot\Omega$, and so $\omega^*(x^*) = \omega(x) = C\gamma(u^* + ib)^\sigma$. Consequently, we can assume that β is pure imaginary, and $\omega = \gamma(u + ib)^\sigma$. By equating to zero the various powers of u in the first of Eq. (4.16), this equation splits into

$$R = 0, \tag{4.22a}$$
$$\mathbf{K}\sigma = 0, \tag{4.22b}$$
$$\mathbf{K}b - \mathrm{Re}\,(\alpha_\zeta)b = 0, \tag{4.22c}$$
$$\mathbf{K}\gamma + [\alpha_\zeta + \mathrm{Re}\,(\alpha_\zeta)(\sigma + 1)]\gamma = 0. \tag{4.22d}$$

For the coefficients of $(R, \alpha_\zeta, \bar\alpha_{\bar\zeta})$ in Eqs. (4.22a) and (4.22d) and its complex conjugate, we have

$$b\gamma = 0,$$

and so, since $\gamma \neq 0$, $b = \sigma + 2 = 0$. Hence,

$$\ddot\Omega = \gamma u^{-2}, \quad \Omega = -\gamma(1 + \log u) + \kappa u + 2. \tag{4.23}$$

Substituting $R = 0$ into (3.11) gives

$$\mathbf{K}(\Omega - u\dot\Omega) + \tfrac{1}{2}(\alpha_\zeta - \bar\alpha_{\bar\zeta})(\Omega - u\dot\Omega) = 0. \tag{4.24}$$

Since $\Omega - u\dot\Omega = \gamma \log u + \tau$, this reduces to

$$\mathbf{K}\gamma + \tfrac{1}{2}(\alpha_\zeta - \bar\alpha_{\bar\zeta})\gamma = 0,$$
$$\mathbf{K}\tau + \tfrac{1}{2}(\alpha_\zeta - \bar\alpha_{\bar\zeta})\tau + \mathrm{Re}\,(\alpha_\zeta)\gamma = 0. \tag{4.25}$$

The dependence of Eqs. (4.22a) and (4.25) immediately gives $\gamma = 0$, and so we have $\omega = 0$. Hence, dim $(\mathcal{S}) > 2$ gives

$$\Omega = \kappa u + \tau.$$

We now prove that $\dot\Delta = 0$. Contrariwise, assume that $\dot\Delta \neq 0$. Equations (3.12) and (3.14d) become

$$\mathbf{K}\kappa + \alpha_\zeta\kappa = -\tfrac{1}{2}\alpha_{\zeta\zeta}, \tag{4.26}$$
$$\mathbf{K}\kappa_{\bar\zeta} + 2\,\mathrm{Re}\,(\alpha_\zeta)\kappa_{\bar\zeta} = 0. \tag{4.27}$$

Taking the ζ-derivative of Eq. (4.27) and using Eq. (4.26), we obtain

$$\mathbf{K}\kappa_{\zeta\bar\zeta} - 2\kappa_{\bar\zeta}\mathbf{K}\kappa + 2\alpha_\zeta(\kappa_{\zeta\bar\zeta} - \kappa\kappa_{\bar\zeta}) + \bar\alpha_{\bar\zeta}\kappa_{\zeta\bar\zeta} = 0. \tag{4.28}$$

Since the determinant of the coefficients of

$$[R + u\,\mathrm{Re}\,(\alpha_\zeta), \alpha_\zeta, \bar\alpha_{\bar\zeta}]$$

in Eqs. (3.14c), (4.24), and (4.25) is zero,

$$\dot\Delta(\kappa_{\zeta\bar\zeta} - 2\kappa\kappa_{\bar\zeta})\kappa_{\bar\zeta} = 0,$$

where $2\dot\Delta = \kappa_{\bar\zeta} - \bar\kappa_\zeta$. Hence, since $\dot\Delta \neq 0$, then

$$(\log \kappa_{\bar\zeta})_\zeta = 2\kappa. \tag{4.29}$$

Similarly, from the dependence of Eqs. (3.14c), (4.25), and its complex conjugate, we have

$$\kappa_{\bar\zeta}\bar\kappa_{\zeta\bar\zeta} = \bar\kappa_\zeta\kappa_{\zeta\bar\zeta}$$

so that

$$(\log \kappa_{\bar\zeta})_{\bar\zeta} = (\log \bar\kappa_\zeta)_{\bar\zeta}.$$

From this equation and Eq. (4.29),

$$2\kappa_{\bar\zeta} = (\log \kappa_{\bar\zeta})_{\zeta\bar\zeta} = (\log \bar\kappa_\zeta)_{\bar\zeta\zeta} = 2\bar\kappa_\zeta,$$

and so $\dot\Delta = 0$ and

$$2\Delta = \kappa_{\bar\zeta} - \bar\kappa_\zeta = 0.$$

This is the necessary and sufficient condition for the existence of a real $p(\zeta, \bar\zeta)$ such that $\kappa = p_\zeta$ and

$$\Omega = p_\zeta u + \tau.$$

The field equation (1.15c) reduces to

$$e^{-2p}\ddot\mu = (e^{-2p}p_{\zeta\bar\zeta})_{\zeta\bar\zeta}, \tag{4.30}$$

and so $\ddot\mu = 0$. The u-derivative of Eq. (1.15a) then gives

$$\dot\mu_\zeta = 4p_\zeta\dot\mu. \tag{4.31}$$

Setting the determinant of the coefficients of

$$(R + u\,\mathrm{Re}\,(\alpha_\zeta), \alpha_\zeta, \bar\alpha_{\bar\zeta})$$

to zero in Eqs. (3.13) and (3.14), setting the ζ-derivative of Eq. (3.14) equal to zero, and using $\ddot\mu = 0$, we get

$$\dot\mu^2(\dot\mu_\zeta - 4p_\zeta\dot\mu) = 0,$$

and so

$$\dot{\mu}_\zeta = 4p_\zeta \dot{\mu}. \tag{4.32}$$

Substituting Eqs. (4.31) and (4.32) into Eq. (3.14) gives

$$\alpha p_\zeta + \bar{\alpha} p_{\bar{\zeta}} + \mathrm{Re}\,(\alpha_\zeta) = 0 \tag{4.33}$$

provided that $\dot{\mu} \neq 0$. From the dependence of Eqs. (3.13), (4.27), and (4.33) with $\kappa = p_\zeta$, we have

$$\dot{\mu}(p_{\zeta\zeta\bar{\zeta}} - 2p_{\zeta\bar{\zeta}}p_\zeta) = 0,$$

and so

$$\dot{\mu}[e^{-2p}p_{\zeta\bar{\zeta}}]_\zeta = 0.$$

From Eq. (4.30) this gives $\dot{\mu} = 0$, and so we have finally proved that, if $\dim(\mathcal{S}) > 2$, then $\dot{\Omega} = \dot{\mu} = \dot{\Delta} = 0$ and therefore, from Lemma 4.1, $\mathcal{g}$ admits a Killing vector of the type $e^{-p}\partial_u$.

5. SPACES WITH $\dim(\mathcal{S}) = 3, 4$

From Theorem 4.1, these spaces must admit a special Killing vector of the type $e^{-p}\partial_u = \partial_s$. The metric is given by Eq. (4.4) and the Killing equations are given in Lemma 4.2. Differentiating Eq. (4.11) with respect to ζ and using $da_0 = 0$ give Eq. (3.12). We list the equations to be satisfied by $\mathbf{K}$, together with their first integrability conditions:

$$\text{(I)} \quad T_\zeta + \mathbf{K}\Lambda + (-a_0 + \alpha_\zeta)\Lambda = 0, \tag{5.1}$$

$$\text{(II)} \quad \tfrac{1}{2}\alpha_{\zeta\zeta} + \mathbf{K}p_\zeta + \alpha_\zeta p_\zeta = 0, \tag{5.2}$$

$$\text{(II}'\text{)} \quad \mathbf{K}p + \mathrm{Re}\,(\alpha_\zeta) = a_0, \tag{5.3}$$

$$\text{(III)} \quad \mathbf{K}m + 3a_0 m = 0, \tag{5.4}$$

$$\text{(IVb)} \quad \mathbf{K}(e^{-2p}p_{\zeta\bar{\zeta}}) + 2a_0(e^{-2p}p_{\zeta\bar{\zeta}}) = 0, \tag{5.5}$$

$$\text{(IVc)} \quad \mathbf{K}d + a_0 d = 0. \tag{5.6}$$

This is a complete set of differential equations in the *six* unknowns $(T, \alpha, \bar{\alpha}, \alpha_\zeta, \bar{\alpha}_{\bar{\zeta}}, a_0)$, since $da_0 \equiv 0$, and so, since we have assumed that $\dim(\mathcal{S}) > 2$, there can be at most two more conditions independent of (5.3).

From the field equation (4.14a), m is an analytic function of ζ, and so Eq. (5.4) becomes $\alpha m_\zeta + 3a_0 m = 0$. If $m_\zeta \neq 0$, this can be solved for α as a function of a_0, and so the number of Killing vectors is at most two (corresponding to the initial values of T and a_0 at a given point). This means that

$$\dim(\mathcal{S}) > 2 \Rightarrow m = m_0 = \text{const.} \tag{5.7}$$

We shall consider three separate cases, depending on the character of the 2-curvature $R^{(2)}$. This curvature transforms as follows under an element of $\hat{\mathcal{C}}$ [see Eq. (4.7)]:

$$R^{(2)*} = C_0^{-2}R^{(2)}. \tag{5.8}$$

Since C_0 is a nonzero constant and since $R^{(2)}$ satisfies Eq. (4.14b) and therefore is a harmonic function, the Killing problem separates into three cases,

$$R^{(2)} = 0,$$
$$dR^{(2)} = 0, \quad R^{(2)} \neq 0,$$

and

$$dR^{(2)} \neq 0. \tag{5.9}$$

We shall treat them separately.

Case I: $R^{(2)} = 0$, and so $p_{\zeta\bar{\zeta}} = 0$, that is, $2p = \mathrm{Re}\,[\sigma(\zeta)]$, where σ is an analytic function of ζ. From Eq. (4.7), $e^{p*} = C_0\,|\sigma/\Phi_\zeta|$, and so p can be transformed to 0 by choosing $\Phi_\zeta = C_0\sigma(\zeta)$. We shall suppose this done. From Eq. (4.5), $\Omega = \Lambda(\zeta, \bar{\zeta})$, and so $\partial_u \eth\Omega = 0$. From Eq. (1.17) this means that m $(= m_0)$ is a nonzero constant. Equation (5.4) now gives

$$a_0 = 0,$$

and so (5.3) becomes $\mathrm{Re}\,(\alpha_\zeta) = 0$. The complete solution of this equation is $\alpha = ib_0\zeta + \alpha_0$, where α_0 and b_0 are both constants, b_0 being real. If we differentiate the remaining first integrability condition (5.6) with respect to ζ, we have

$$Kd_\zeta + \alpha_\zeta d_\zeta = 0.$$

This, together with its complex conjugate and Eq. (5.6) itself, gives three independent constraints on the five independent variables ($a_0 = 0$!), and so $\dim(\mathcal{S}) < 3$, unless $d_\zeta = 0$. Hence $d = d_0$ (a constant), and the last of the field equations (4.14c) gives $m_0 = \bar{m}_0$. From Eq. (4.5), $\Lambda = id_0\bar{\zeta} + A_\zeta$ where $A(\zeta, \bar{\zeta})$ is real and can be eliminated by the transformation $s \to s^* = s + A$. [See Eq. (4.9).] Finally, since $p = 0$,

$$\mu = m_0, \quad \Omega = id_0\bar{\zeta}, \quad \eth\Omega = 0.$$

The last of these equations implies that $\mathcal{g}$ is a Kerr–Schild metric.[5] If we introduce new coordinates $(\eta, \bar{\eta}, \omega, r)$ given by

$$\eta = \zeta(r + id_0), \quad \omega = s - \zeta\bar{\zeta}r,$$

then the metric reduces to

$$(d\tau)^2 = 2(d\eta\, d\bar{\eta} + d\omega\, dr) + 2\left(\frac{m_0 r}{r^2 + d_0^2}\right)\left(d\omega + \frac{\eta\, d\bar{\eta}}{r + id_0} + \frac{\bar{\eta}\, d\eta}{r - id_0} - \frac{\eta\bar{\eta}r}{r^2 + d_0^2}\right)^2. \tag{5.10}$$

The constants m_0 and d_0 are not invariants and, in fact, can be transformed to one of the following two cases:

$$d_0 = 0, \quad m_0 = 1, \tag{5.11a}$$

$$d_0 = 1, \quad m_0 \neq 0 \quad \text{(an invariant)}. \tag{5.11b}$$

The dimension of $\mathcal{S}$ is 4.

Case II: $dR^{(2)} = 0$, $R^{(2)} \neq 0$: From Eq. (5.8) we see that we can transform $R^{(2)}$ to ± 1:

$$R^{(2)} = R_0 = \pm 1. \tag{5.12}$$

The 2-metric $(e^{2p}\, d\zeta\, d\bar{\zeta})$ is that for a sphere or pseudosphere and so the coordinates ζ can be chosen so that

$$e^{-p} = \zeta\bar{\zeta} - R_0. \tag{5.13}$$

From Eq. (5.5) we see that $a_0 = 0$. Equation (5.3) can now be solved completely for α, giving

$$\alpha = \alpha_0\zeta^2 + ib_0\zeta - \bar{\alpha}_0 R_0,$$

where α_0 and b_0 are constants, b_0 being real. Also Eq. (5.6) and its derivatives with respect to ζ and $\bar{\zeta}$ give three independent integrability conditions, unless $d = d_0$ (a constant). The last field equation (4.14c) then gives

$$\mathrm{Im}\,(m_0) = -2R_0 d_0.$$

The remaining field variable Λ can be found from Eq. (4.5). Using the coordinate freedom of Eq. (4.8), in particular, that $s^* = s + A$, we can reduce Λ to $-id_0\bar{\zeta}/R_0(\zeta\bar{\zeta} - R_0)$. Finally, the metrics with four Killing vectors (KV) reduce to

$$\tfrac{1}{2}(d\tau)^2 = (r^2 + d_0^2)(\zeta\bar{\zeta} - R_0)^{-2}\, d\zeta\, d\bar{\zeta} + dr(\kappa) + \left(R_0 + \mathrm{Re}\left(\frac{m_0}{r + id_0}\right)\right)(\kappa)^2, \tag{5.14}$$

where $\kappa = ds + [id_0/R_0(\zeta\bar{\zeta} - R_0)](\zeta\, d\bar{\zeta} - \bar{\zeta}\, d\zeta)$,

$$\mathrm{Im}\,(m_0) = -2R_0 d_0.$$

There are two classes of metrics here:

$$m_0 \text{ is real} \Rightarrow \text{Schwarzschild},^6 \tag{5.15a}$$

$$m_0 \text{ is complex} \Rightarrow \text{NUT}.^7 \tag{5.15b}$$

Both of these are known to have four Killing vectors.

Case III: $dR^{(2)} \neq 0$: From Eq. (4.14b) this gives

$$e^{-2p}p_{\zeta\bar{\zeta}} = 2\,\mathrm{Re}\,[F(\zeta)].$$

If we transform the ζ coordinate, $\zeta^* = F(\zeta)$, and then drop the asterisk, the 2-curvature reduces to

$$e^{-2p}p_{\zeta\bar{\zeta}} = \zeta + \bar{\zeta}. \tag{5.16}$$

The complete solution of Eq. (5.5) is

$$\alpha = -2a_0\zeta + ib_0, \tag{5.17}$$

where a_0 and b_0 are real constants. Equation (5.1) and its complex conjugate can now be solved for T. Since this can only introduce one further constant of integration, in addition to a_0 and b_0, we see that $\dim(\mathcal{S}) \leq 3$. In order for the dimension to be 3, none of the remaining integrability conditions can be independent of Eq. (5.5). Since Eq. (5.4) is independent of Eq. (5.5), the complex mass must be zero unless $m = m(\zeta) = 0$. Also, if Eq. (5.17) is substituted into Eq. (5.2) and the coefficients of a_0 and b_0 equated to zero, the resultant equations can be solved for e^{-p}, giving $e^{-p} = A(\zeta + \bar{\zeta})^{\frac{3}{2}}$. Substituting this into Eq. (5.16), we find $A = (\frac{2}{3})^{\frac{1}{2}}$ and

$$e^{-p} = (\tfrac{2}{3})^{\frac{1}{2}}(\zeta + \bar{\zeta})^{\frac{3}{2}}. \tag{5.18}$$

Since Eq. (5.6) cannot be independent of Eq. (5.5) d^2 must be proportional to $R^{(2)}$, i.e., $d = d_0(\zeta + \bar{\zeta})^{\frac{1}{2}}$, where d_0 is a real constant. Substituting this into the last of the field equations (4.14c), we have $d_0 = 0$ and so $d = 0$. From Eq. (4.5) $\Lambda = A_\zeta$, where A is a real function of $(\zeta, \bar{\zeta})$. It can be eliminated by a transformation of $\hat{\mathcal{C}}$ with $(\Phi, A, C_0) = (\zeta, A, 1)$, and so we have

$$m = \Lambda = 0 = d = 0.$$

Finally, if we substitute $\zeta = \frac{4}{3}(x + iy)$, then

$$(d\tau)^2 = r^2x^{-3}(dx^2 + dy^2) + 2\, dr\, ds + \tfrac{3}{2}x(ds)^2. \tag{5.19}$$

This is a Robinson–Trautman[8] metric of Petrov type III with a 3-dimensional symmetry group.

6. SPACES WITH 2-DIMENSIONAL SYMMETRY GROUPS

From Lemma 3.5 there are two distinct possibilities for $\dim(\mathcal{S}) = 2$:

$$\textit{Case I:} \quad \exists \text{ a KV of the form } \mathbf{K}_1 = e^{-p}\partial_u,$$

$$\textit{Case II:} \quad \nexists \text{ a KV of the form } \mathbf{K}_1 = e^{-p}\partial_u. \tag{6.1}$$

Since there are only two independent KV, it will be convenient to consider all possible Lie groups separately.

Case I: We shall write $\mathbf{K}_1 = e^{-p}\partial_u = \partial_s$ and use the form of the metric in Eqs. (4.4) and (4.5). From Lemma 4.2, any other KV ($\mathbf{K}_2$, say) is given by Eq. (4.10),

$$\mathbf{K}_1 = \partial_s, \quad K_2 = \alpha\partial_\zeta + \bar{\alpha}\partial_{\bar{\zeta}} + a_0(s\partial_s - r\partial_r) + T\partial_s, \tag{6.2}$$

and the group structure constants can be computed directly, using Eq. (5.3),

$$[\mathbf{K}_1, \mathbf{K}_2] = a_0\mathbf{K}_1. \tag{6.3}$$

There are four distinct possibilities, according to whether the commutator is zero, or not,

$$[\mathbf{K}_1, \mathbf{K}_2] = 0, \tag{6.4a}$$

$$[\mathbf{K}_1, \mathbf{K}_2] = (1)\mathbf{K}_1 \tag{6.4b}$$

and whether the two-metric $e^{+2p}\,d\zeta\,d\bar{\zeta}$ has constant curvature, or not,

$$e^{-2p}p_{\zeta\bar{\zeta}} = R_0 \text{ (constant)}, \tag{6.5a}$$

$$e^{-2p}p_{\zeta\bar{\zeta}} = 2\,\mathrm{Re}\,[F(\zeta)], \quad F_\zeta \neq 0. \tag{6.5b}$$

We shall consider these separately.

Case 1: Eqs. (6.4a) and (6.5a). $\mathbf{K}_2$ can be reduced to the following form by a transformation of coordinates [see Eq. (4.8)]:

$$\mathbf{K}_2 = i(\zeta\partial_\zeta - \bar{\zeta}\partial_{\bar{\zeta}}). \tag{6.6}$$

The allowed coordinate transformation left for ζ is

$$\zeta^* = \phi(\zeta) = \beta_0\zeta^{\gamma_0}, \tag{6.7}$$

where β_0 and γ_0 are complex constants. Using Lemma 4.2, we reduce the Killing equations to

$$\begin{aligned} \mathbf{K}_2 p &= 0,\\ \mathbf{K}_2 m &= 0,\\ T_\zeta + \mathbf{K}_2\Lambda + i\Lambda &= 0. \end{aligned} \tag{6.8}$$

The first of these gives

$$p = p(R), \quad R = |\zeta|, \tag{6.9}$$

while the second, together with the field equation (4.14a), gives $m = \mu_0$ (a constant). Using Eqs. (6.5a) and (6.9), we can show that

$$e^{-p} = a_0R^{\alpha_0} - R_0R^{2-\alpha_0}/a_0(\alpha_0 - 1)^2.$$

Using Eq. (6.7) with $\zeta^* = a_0(\alpha_0 - 1)\zeta^{\alpha_0 - 1}$, we may transform this to

$$e^{-p} = \zeta\bar{\zeta} - R_0 = R^2 - R_0,$$

where we have removed all asterisks after the transformation. Equation (5.6) now reduces to $\mathbf{K}_2 d = 0$, and so $d = d(R)$. Substituting this into Eq. (4.5) gives $\Lambda = i\bar{\zeta}B(R) + A(\zeta, \bar{\zeta})_\zeta$, where A and B are both real functions. The first of these functions can be transformed to zero by replacing s by $(s + A)$, and so, without loss of generality,

$$\Lambda = i\bar{\zeta}B(R), \quad B = \bar{B}.$$

The last field equation (4.14c) can be reduced to

$$\left(R\frac{d}{dR}\right)^2\left(R^3e^p\frac{d}{dR}(e^{-2p}B)\right) = 8e^{3p}\,\mathrm{Im}\,(\mu_0).$$

The complete solution to this is

$$\begin{aligned} \Lambda = i\bar{\zeta}e^{2p}[-\tfrac{1}{2}\,\mathrm{Im}\,(\mu_0)R^2 + C_1 \\ + C_2(2\log R + R_0R^{-2}) + C_3(R^2 + R_0^2R^{-2})], \\ e^{-p} = \zeta\bar{\zeta} - R_0, \\ m = \mu_0. \end{aligned} \tag{6.10}$$

A simple calculation shows that

$$\partial_u\eth\Omega = p_{\zeta\zeta} - p_\zeta^2 = 0,$$

and so, from Eq. (1.17), the space is flat unless $\mu_0 \neq 0$. If Λ is substituted into Eq. (4.4), this gives a 6-parametric solution as a function of μ_0 (complex), the $\{C_i\}_{i=1}^3$, and R_0. These constants are not independent invariants. If $R_0 \neq 0$, then it can be transformed to ± 1 by the transformation

$$\zeta^* = |R_0|^{-\frac{1}{2}}\zeta, \quad s^* = |R_0|^{\frac{1}{2}}s, \quad r^* = |R_0|^{-\frac{1}{2}}r.$$

In the second case where $R_0 = 0$, μ_0 can be transformed to $C_0^{-3}\mu_0 = \mu^* \neq 0$ and so either Re (μ_0) or Im (μ_0) can be taken to be $+1$, leaving *four* independent parameters.

Case 2: Eqs. (6.4a) and (6.5b). If we let $\zeta^* = F(\zeta)$, then Eq. (6.5b) can be transformed to

$$e^{-2p}p_{\zeta\bar{\zeta}} = \zeta + \bar{\zeta}, \tag{6.11}$$

and so, from Eq. (5.5), the second Killing vector must be

$$\mathbf{K}_2 = i(\partial_\zeta - \partial_{\bar{\zeta}}).$$

From Eq. (5.3), $p = p(x)$, where $x = \zeta + \bar{\zeta}$, and so Eq. (6.11) gives

$$\frac{d^2p}{dx^2} = xe^{2p}. \tag{6.12}$$

Unfortunately, the only known solution of this equation is

$$e^{-p} = (\tfrac{2}{3})^{\frac{1}{2}}x^{\frac{3}{2}}. \tag{6.13}$$

From the Killing equations and Eq. (4.14a), $\Lambda = \Lambda(x)$ and $m = m_0$ (a constant). The real part of Λ can be eliminated by a $\hat{\mathfrak{C}}$ transformation with $A = A(x)$. Equations (4.5) and (4.14) can then be solved for Λ:

$$-i\Lambda = a_0x^{-\frac{5}{2}}\sinh\tfrac{1}{2}(13)^{\frac{1}{2}}(x - x_0) + \tfrac{3}{4}\,\mathrm{Im}\,(m_0)x^{-3}, \tag{6.14}$$

where a_0 is an arbitrary real constant. There is a further solution $\Lambda = ib_0$, with b_0 a real constant which can be added to Eq. (6.14); but this can be eliminated by a $\hat{\mathfrak{C}}$ transformation with $s^* = s + ib_0(\zeta - \bar{\zeta})$. This is not the complete solution for Eqs. (6.4a) and (6.5b), but Eq. (6.13) is the only known solution to Eq. (6.12).

Case 3: Eqs. (6.4b) and (6.5a). Eq. (3.14b), together with $a_0 = 1$ gives $R^{(2)} = 0$, and so $p_{\zeta\bar\zeta} = 0$, that is, $p = \mathrm{Re}\,[\psi(\zeta)]$. From Eqs. (4.7) and (4.8), there is a $\hat{\mathfrak{C}}$ transformation with $\Phi_\zeta = \psi$ for which $p^* = 0$, and so we shall take $p = 0$. From Eq. (5.3), $\mathrm{Re}\,(\alpha_\zeta) = 1$, and so $\alpha = \alpha_0(\zeta - \beta_0)$, where α_0 and β_0 are constants and $\mathrm{Re}\,(\alpha_0) = 1$. If we remember that $\mathbf{K}_1$ and $\mathbf{K}_2$ are not invariantly defined vectors, then it can be seen that the residual group of transformations is

$$\mathbf{K}_1^* = \partial_s^* = C_0^{-1}\mathbf{K}_1, \quad \mathbf{K}_2^* = \mathbf{K}_2 + e_0\mathbf{K}_1,$$
$$\zeta^* = \Phi(\zeta), \quad r^* = C_0^{-1}r, \quad s^* = C_0(s + A). \tag{6.15a}$$

From Eq. (4.7) and $p = p^* = 0$, we see that $|\Phi_\zeta| = |C_0|$, and so

$$\zeta^* = \Phi(\zeta) = e^{id_0}C_0(\zeta - \zeta_0). \tag{6.15b}$$

The functions α and T transform as follows:

$$\alpha^* = e^{id_0}C_0\alpha, \quad T^* = C_0(T + \alpha A_\zeta + \bar\alpha A_{\bar\zeta} + a_0 s);$$

so, by choosing ζ_0 correctly, β_0 can be transformed to zero,

$$\alpha = \alpha_0\zeta, \quad \mathrm{Re}\,(\alpha_0) = 1,$$

where α_0 is invariant.

From the field equations (4.14a) and (4.14c), $m = 2\rho_\zeta$, where $\rho = \rho(\zeta)$, and so

$$\mathrm{Im}\,(\Lambda_{\bar\zeta})_{\zeta\bar\zeta} = 2\,\mathrm{Im}\,(\rho_\zeta) = 2\,\mathrm{Im}\,(\bar\zeta\rho)_{\zeta\bar\zeta}.$$

This is immediately integrable, giving

$$\mathrm{Im}\,(\Lambda_{\bar\zeta} - 2\bar\zeta\rho - \sigma) = 0,$$

where $\sigma = \sigma(\zeta)$. This can be rewritten as

$$\mathrm{Im}\,(\Lambda - \bar\zeta^2\rho - \bar\zeta\sigma)_{\bar\zeta} = 0$$

so that $\Lambda = \bar\zeta^2\rho + \bar\zeta\sigma + A_\zeta$, where A is a real function of $(\zeta, \bar\zeta)$. This can be eliminated by letting $s^* = s + A$, and so we shall assume

$$\Lambda = \bar\zeta^2\rho + \bar\zeta\sigma.$$

Of the basic Killing equations (5.1)–(5.4), Eq. (5.2) is already satisfied and Eq. (5.1) can be solved for T provided that its integrability condition (5.6) is satisfied. This means that $\mathfrak{S}$ admits a 2-parameter group (at least), provided that Eqs. (5.4) and (5.6) are satisfied, i.e.,

$$\alpha_0\zeta\rho_{\zeta\zeta} + 3\rho_\zeta = 0,$$
$$\mathrm{Im}\,[\bar\zeta(2\alpha_0\zeta\rho_\zeta + 2\bar\alpha_0\rho + 2\rho) + (\alpha_0\zeta\sigma_\zeta + \sigma)] = 0,$$

which have as complete solutions

$$\sigma = 2\bar\gamma_0\zeta + \lambda_0\zeta^{-1/\alpha_0},$$
$$\rho = \gamma_0 + m_0\zeta^{-3/(\alpha_0+1)},$$

where γ_0, λ_0, and m_0 are all complex constants. The first of these can be eliminated be letting $s^* = s + 2\,\mathrm{Re}\,(\gamma_0\bar\zeta^2\zeta)$, and so

$$\Lambda = \lambda_0\bar\zeta\zeta^{-1/\alpha_0} + m_0\bar\zeta^2\zeta^{-3/(\alpha_0+1)},$$
$$m = 2m_0(1 - 3/\alpha_0)\zeta^{-3/\alpha_0}, \tag{6.16}$$
$$p = 0, \quad \mathrm{Re}\,(\alpha_0) = 1.$$

The remaining coordinate freedom can be used to reduce m_0 to a *real* constant. The complete metric for type (6.4b) and (6.5a) is given by Eq. (4.4) with the above Λ, m, and p. It is a function of four real constants, with a symmetry group given by

$$s^* = C_0(s + s_0), \quad r^* = C_0^{-1}r,$$
$$\zeta^* = C_0^{\alpha_0}\zeta,$$

where C_0 and s_0 are real.

Case 4: Eqs. (6.4b) and (6.5b). Just as for case 2, Eq. (6.5b) can be transformed to

$$e^{-2p}p_{\zeta\bar\zeta} = \zeta + \bar\zeta, \tag{6.17}$$

and so, from Eq. (5.5) with $a_0 = 1$, $\alpha = -2(\zeta + ie_0)$. The constant e_0 can be eliminated by $\zeta^* = \zeta + ie_0$, and so, without loss of generality,

$$\alpha = -2\zeta. \tag{6.18}$$

From Eq. (5.3),

$$e^{-2p} = e^{-2P}(\zeta + \bar\zeta)^3, \tag{6.19}$$

where P is a function of the angle θ,

$$P = P(\theta), \quad \zeta = Re^{i\theta}. \tag{6.20}$$

Substituting this into Eq. (6.17)

$$e^{-2P}\left(\cos^2\theta\,\frac{d^2P}{d\theta^2} + \frac{3}{2}\right) = 1. \tag{6.21}$$

Again, the only known solution to this equation is that of Eq. (6.12), $e^{-2P} = \frac{2}{3}$,

$$e^{-2p} = \tfrac{2}{3}(\zeta + \bar\zeta)^3. \tag{6.22}$$

From Eq. (5.6), $d = (\zeta + \bar\zeta)^{\frac{1}{2}}D(\theta)$, where D is a real function. From Eq. (4.5)

$$\Lambda = i(\zeta + \bar\zeta)^{-1}L(\theta),$$

where an additional term of the type A_ζ has been eliminated by the transformation $s^* = s + A$. The functions L and D are related by

$$\tfrac{1}{2}\sin 2\theta\dot{L} - L = e^{2p}D. \tag{6.23}$$

From Eq. (5.4) and (4.14a),

$$m = \mu_0 \zeta^{\frac{3}{2}}, \tag{6.24}$$

where μ_0 is a complex constant. Finally, the last field equation (4.14c) reduces to

$$\text{Im}\,[\mu_0(1 + e^{-2i\theta})^{-\frac{3}{2}}] = e^{-2P}\cos^2\theta\,\frac{d^2D}{d\theta^2} - [2 + \tfrac{1}{4}e^{-2p}]D. \tag{6.25}$$

We do not have any solutions to these equations.

Case II: These metrics do not admit a Killing vector of the type $e^{-p}\partial_u$ and so not all of $\{\ddot{\Omega}, \dot{\Delta}, \dot{\mu}\}$ can be zero. There are two distinct possibilities for the symmetry group.

Case 1: Eq. (6.4a). $[\mathbf{K}_1, \mathbf{K}_2] = 0$: By means of the transformation (3.15) one of these Killing vectors, $\mathbf{K}_1$ say, can be reduced to $\partial_\zeta + \partial_{\bar\zeta}$. Since it is assumed to commute with any other Killing vector $\mathbf{K}_2$ [with corresponding (α_2, R_2)], it follows that $(\partial_\zeta + \partial_{\bar\zeta})\alpha_2 = (\partial_\zeta + \partial_{\bar\zeta})R_2 = 0$, and so α_2 is a constant, which without loss of generality can be taken to be i. R_2 can be transformed to zero by (3.15) without disturbing the canonical form for $\mathbf{K}_1$ and so the Killing vectors reduce to Re (∂_ζ) and Im (∂_ζ). The metric is independent of $(\zeta, \bar\zeta)$ and so can only be a function of (u, v). The field equations (1.15) reduce to

$$\mu = \mu_0 \dot{\Omega}^{-3},$$

$$\partial_u\left\{4\mu_0\dot{\Omega}^{-3} + 2\dot{\Omega}\frac{d}{du}\left[\dot{\Omega}\frac{d^2}{du^2}(\Omega^2)\right]\right\} = \left|\frac{d^2}{du^2}(\Omega^2)\right|^2, \tag{6.26}$$

$$\text{Im}\left\{4\mu_0\dot{\Omega}^{-3} + 2\dot{\Omega}\frac{d}{du}\left(\dot{\Omega}\frac{d^2}{du^2}(\Omega^2)\right)\right\} = 0.$$

Since $\mathcal{S}$ is assumed to be nonflat, either μ_0 or $(d^2/du^2)(\Omega^2)$ is nonzero. We do not know any solutions to these equations.

Case 2: Eq. (6.4b). $[\mathbf{K}_1, \mathbf{K}_2] = \mathbf{K}_1$: Again, we take $\mathbf{K}_1 = \partial_\zeta + \partial_{\bar\zeta}$. If we take $\mathbf{K}_2$ to be a second Killing vector [with corresponding (α_2, R_2)], then we must have

$$\alpha_{2,3} = 1, \quad (\partial_\zeta + \partial_{\bar\zeta})R_2 = 0, \tag{6.27}$$

so that $\alpha = \zeta + is_0$, where s_0 is a real constant which can be eliminated by $\zeta^* = \zeta + is_0$ without affecting the canonical form for $\mathbf{K}_1$. Similarly, the second of Eqs. (6.27) allows R_2 to be eliminated by Eq. (3.15), and so the Killing vectors can be reduced to

$$\mathbf{K}_1 = \partial_\zeta + \partial_{\bar\zeta}, \quad \mathbf{K}_2 = \zeta\partial_\zeta + \bar\zeta\partial_{\bar\zeta} + u\partial_u - v\partial_v.$$

From Lemma 3.2, the Killing equations can be integrated immediately to give

$$\Omega = \Omega(t),$$
$$\mu = u^3\nu(t),$$

where $t = u/\text{Im}\,(\zeta)$. The field equations (4.14) reduce to ordinary differential equations in $\Omega(t)$ and $\nu(t)$. However, these are even worse than the equations for case 1, and we do not have any solutions for them.

7. SPACES WITH 1-DIMENSIONAL SYMMETRY GROUPS

As with some of the 2-dimensional groups, there is not much that can be said about this case. It is always possible to put the single Killing vector into one of the following canonical forms,

$$\mathbf{K} = e^{-p}\partial_u,$$
$$\mathbf{K} = \partial_\zeta + \partial_{\bar\zeta},$$

as we saw in Lemma 3.5. We have already investigated case I in Sec. 4 and found that the field equations reduce to Eqs. (4.14). In case II the metric is independent of Re (ζ), but this does not simplify the field equations to any appreciable degree.

8. CONCLUSION

In this paper we have been able to construct all possible groups of motions for empty Einstein spaces admitting a diverging, geodesic, and shear-free ray congruence. However, we have been unable to solve the field equations (1.15) completely when the dimension of the group of motions is one or two. The spaces with 4-dimensional groups are those of Scharzschild and NUT (Newman, Unti, and Tamburino), and the metric of Eq. (5.10). The only metric with dim $(\mathcal{S}) = 3$ is that of Eq. (5.19). The main results of the section on 2-dimensional symmetry groups are the metrics given in Eq. (6.10), (6.13), (6.14), and (6.16).

* This work was supported in part by the Aerospace Research Laboratories, Office of Aerospace Research, U.S. Air Force, under Contract AF 33-615-1029, and by a grant from the National Science Foundation, NSF Grant No. GP-8868.

[1] G. C. Debney, R. P. Kerr, and A. Schild, J. Math. Phys. **10**, 1842 (1969).
[2] J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. **22**, 13 (1962).
[3] R. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).
[4] R. P. Kerr, Phys. Rev. Letters **11**, 237 (1963).
[5] R. P. Kerr and A. Schild, *Proceedings of Symposia in Applied Mathematics* (Amer. Math Soc., Providence, R.I., 1965), Vol. XVII, p. 199.
[6] K. Schwarzschild, S.-B. Preuss, Akad. Wiss., Berlin, 189 (1916).
[7] E. Newmann, L. Tamburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[8] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962).

Scalar-Tensor Theory in Canonical Form. II. Exact Solutions for a Static Spherically Symmetric Charged Mass Distribution*†

EDWARD T. TOTON‡
Physics Department, University of Maryland, College Park, Maryland

(Received 17 February 1970)

In this part variational techniques are used to generate the differential equations determining the static fields of a charged mass distribution. The equations are solved external to the source by the Hamilton-Jacobi method under the conditions of spherical symmetry. Solutions for a charged mass shell are investigated, the rigorous solution for the electron in general relativity is given and, by limit arguments, it is shown that the same external solution arises for the electron in the scalar-tensor theory.

1. INTRODUCTION

In Paper I variational arguments were used to discuss the self-energy of the electron in the scalar-tensor theory of gravity.[1] The model chosen for the electron was that of a δ-function source with charge e and bare mass m_0. By casting the scalar-tensor theory into canonical form, there arose the true Hamiltonian of the system. By solving the constraint equations, one could then obtain the energy of the system for the configuration of static fields characteristic of the electron. However, it was also true that the static configuration was not simply given by setting all independent excitations of the fields present equal to zero because of the *scalar* nature of the scalar field. No gauge condition exists on the scalar, and thus no constraints arise to suggest a decomposition of the field into that describing waves and that part relating directly to the source. Although gravity and electromagnetic waves can be excluded from the system by setting the independent excitations of the metric field and the electromagnetic field equal to zero, the scalar field that corresponds to the static configuration of the system must be included.

The static configuration of the electron is most properly defined by the minimization of the energy of the system with respect to all independent excitations. This corresponds to a minimization of the true Hamiltonian. But, with all waves eliminated, we are left with a minimization of the energy with respect to the scalar field alone. The following calculations show this minimization procedure for a charged mass distribution. Exact solutions are then obtained for a spherically symmetric distribution, the solutions being given outside the mass.

2. ENERGY MINIMIZATION CONDITIONS

According to the choice of coordinates for the canonical form of the scalar-tensor theory of Paper I and with a desire for the static solution, the momentum constraints are identically satisfied, and the energy constraint becomes[1]

$$g^{\frac{1}{2}}\,{}^3R = -8\chi\nabla^2\chi = (\tfrac{3}{2}+\omega)\chi^2(\nabla \ln \phi)^2 + \frac{(E^{\mathrm{L}})^2}{2\chi^2} + \frac{\rho_m(x)}{\phi^{\frac{1}{2}}}, \quad (1)$$

where ϕ is as yet an unspecified function of position, ρ_m is the mass density, E^{L} is the longitudinal part of the electric field and is determined by solving

$$E^{i\mathrm{L}}_{,i} = \rho_e(x) \quad \text{(the charge density)}, \quad (2)$$

and where χ is related to the metric through $g_{ij} = \chi^4\delta_{ij}$. The energy of a system can be written[2]

$$m = \oint dS_i(g_{ij,j} - g_{jj,i}) \quad (3)$$

which, because of the particular form of our metric, can be further written[3]

$$m = -8\oint dS_i\chi^3\chi_{,i} = -8\oint dS_i\chi_{,i} = -8\int d^3x\nabla^2\chi. \quad (4)$$

Here we have used the boundary condition that space becomes Minkowskian at infinity. M is a functional of the scalar ϕ because of the dependence of χ on ϕ in Eq. (1). The maintenance of this constraint throughout the variation implies the necessity of a Lagrange multiplier. We thus prepare m for variation in the following manner.

If the constraint equation is divided by χ and multiplied by an arbitrary function $\lambda(x)$, an integration over all space yields

$$-8\int d^3x\lambda\nabla^2\chi = \int d^3x\left((\tfrac{3}{2}+\omega)\chi\lambda(\nabla \ln \phi)^2 + \frac{\lambda(E^{\mathrm{L}})^2}{2\chi^3} + \frac{\lambda\rho_m}{\phi^{\frac{1}{2}}\chi}\right). \quad (5)$$

An identity can be written as

$$\int d^3x(-8\lambda\nabla^2\chi) = \int d^3x(8\chi\nabla^2\lambda - 8\lambda\nabla^2\chi) - 8\int d^3x\chi\nabla^2\lambda, \quad (6)$$

which by Gauss' theorem yields

$$\int d^3x(-8\lambda\nabla^2\chi) = 8\int d\mathbf{s}\cdot(\chi\nabla\lambda - \lambda\nabla\chi) - 8\int d^3x\chi\nabla^2\lambda, \quad (7)$$

the surface integral being taken at infinity. With the assumption that $\lambda(x)$ assumes the nonzero value λ_o at infinity and with the requirement that, in every case, χ assumes the value unity at infinity, Eq. (7) can be written

$$\int d^3x(-8\lambda\nabla^2\chi) = 8\int d\mathbf{s}\cdot\nabla\lambda - 8\lambda_0\int d\mathbf{s}\cdot\nabla\chi - 8\int d^3x\chi\nabla^2\lambda. \quad (8)$$

By Gauss' theorem

$$\int d\mathbf{s}\cdot\nabla\chi = \int d^3x\nabla^2\chi = -\tfrac{1}{8}m.$$

Thus from (8) we find

$$\lambda_0 m = \int d^3x(-8\lambda\nabla^2\chi + 8\chi\nabla^2\lambda) - 8\int d\mathbf{s}\cdot\nabla\lambda. \quad (9)$$

Finally, if we substitute for $\nabla^2\chi$ obtained from the constraint relation (1), Eq. (9) becomes

$$\lambda_0 m = \int d^3x\left(8\chi\nabla^2\lambda + (\tfrac{3}{2}+\omega)\chi\lambda(\nabla\ln\phi)^2 + \frac{(E^{\mathrm{L}})^2\lambda}{2\chi^3} + \frac{\lambda\rho_m}{\phi^{\frac{1}{2}}\chi}\right) - \int d\mathbf{s}\cdot\nabla\lambda. \quad (10)$$

The first variation of m is thus given by

$$\lambda_0\delta m = \int d^3x\left[\left(8\nabla^2\lambda + (\tfrac{3}{2}+\omega)\lambda(\nabla\ln\phi)^2 - \frac{3\lambda(E^{\mathrm{L}})^2}{2\chi^4} - \frac{\lambda\rho_m}{\phi^{\frac{1}{2}}\chi^2}\right)\delta\chi + \left(-2(\tfrac{3}{2}+\omega)\nabla\cdot(\chi\lambda\nabla\ln\phi) - \frac{\lambda\rho_m}{\phi^{\frac{1}{2}}\chi}\right)\frac{\delta\phi}{\phi}\right]. \quad (11)$$

If the Lagrange multiplier is adjusted at every point in space so that the coefficient of $\delta\chi$ is zero, then, because $\delta\phi$ is arbitrary, the condition of extremized mass gives us

$$\nabla\cdot(\chi\lambda\nabla\ln\phi) + \lambda\rho_m/(3+2\omega)\phi^{\frac{1}{2}}\chi = 0, \quad (12a)$$

where λ is determined from

$$\frac{8\nabla^2\lambda}{\lambda} + (\tfrac{3}{2}+\omega)(\nabla\ln\phi)^2 - \frac{3(E^{\mathrm{L}})^2}{2\chi^4} - \frac{\rho_m}{\phi^{\frac{1}{2}}\chi^2} = 0 \quad (12b)$$

and the constraint is rewritten

$$\frac{8\nabla^2\chi}{\chi} + (\tfrac{3}{2}+\omega)(\nabla\ln\phi)^2 + \frac{(E^{\mathrm{L}})^2}{2\chi^4} + \frac{\rho_m}{\phi^{\frac{1}{2}}\chi^2} = 0. \quad (12c)$$

A solution of these coupled nonlinear differential equations will then give us the static minimum-energy configuration for a charged mass distribution.

3. SOLUTION OUTSIDE A SPHERICALLY SYMMETRIC CHARGED MASS DISTRIBUTION

Outside the charged mass distribution exhibiting spherical symmetry, it is possible to obtain exact solutions of Eqs. (12). In this region the mass density is zero and the solution of the electric constraint [Eq. (2)] is such that $(E^{\mathrm{L}})^2 = \alpha^2 r^{-4}$, where $\alpha^2 = (e/4\pi)^2$. The three equations can then be written

$$\frac{d}{dr}\left(r^2\chi\lambda\frac{d\ln\phi}{dr}\right) = 0, \quad (13a)$$

$$\frac{8}{r^2\lambda}\frac{d}{dr}\left(r^2\frac{d\lambda}{dr}\right) + (\tfrac{3}{2}+\omega)\left(\frac{d\ln\phi}{dr}\right)^2 - \frac{3\alpha^2}{2r^4\chi^4} = 0, \quad (13b)$$

$$\frac{8}{r^2\chi}\frac{d}{dr}\left(r^2\frac{d\chi}{dr}\right) + (\tfrac{3}{2}+\omega)\left(\frac{d\ln\phi}{dr}\right)^2 + \frac{\alpha^2}{2r^4\chi^4} = 0. \quad (13c)$$

The solution of (13a) can be written, owing to the positive behavior of $(\frac{3}{2}+\omega)$,

$$\frac{d\ln\phi}{dr} = \frac{-A}{(\frac{3}{2}+\omega)^{\frac{1}{2}}r^2\chi\lambda}. \quad (14)$$

When Eq. (13a) is substituted into Eqs. (13b) and (13c), we are left to solve

$$\frac{8}{r^2\lambda}\frac{d}{dr}\left(r^2\frac{d\lambda}{dr}\right) + \frac{A^2}{r^4\chi^2\lambda^2} - \frac{3\alpha^2}{2r^4\chi^4} = 0, \quad (15a)$$

$$\frac{8}{r^2\chi}\frac{d}{dr}\left(r^2\frac{d\chi}{dr}\right) + \frac{A^2}{r^4\chi^2\lambda^2} + \frac{\alpha^2}{2r^4\chi^4} = 0. \quad (15b)$$

The particular behavior of the electric field in r allows us to take special advantage of the substitution

$r = z^{-1}$. For, then, Eqs. (15) become

$$\frac{8}{\lambda}\frac{d^2\lambda}{dz^2} + \frac{A^2}{\chi^2\lambda^2} - \frac{3\alpha^2}{2\chi^4} = 0, \tag{16a}$$

$$\frac{8}{\chi}\frac{d^2\chi}{dz^2} + \frac{A^2}{\chi^2\lambda^2} + \frac{\alpha^2}{2\chi^4} = 0. \tag{16b}$$

These coupled equations are of second order, and the independent argument z appears nowhere explicitly. Because of this it is possible to employ the device of considering these equations to be dynamical Lagrangian equations for some mechanical system. The variables χ and λ are interpreted as the coordinates of the system; z is taken as the time. Once a Lagrangian is obtained, the corresponding Hamiltonian can be found. It is quite likely that the Lagrangian will not depend on z so that the Hamiltonian will represent a first integral of Eqs. (16). But, more importantly, the Hamilton–Jacobi equation formed from it is, in fact, separable; full solutions can thus be immediately generated for Eqs. (16).

A Lagrangian which effects this is

$$\mathcal{L} = 8\frac{d\lambda}{dz}\frac{d\chi}{dz} + \frac{A^2}{\chi\lambda} - \frac{\alpha^2\lambda}{2\chi^3}. \tag{17}$$

By defining $P_\chi = \partial\mathcal{L}/\partial\chi'$ and $P_\lambda = \partial\mathcal{L}/\partial\lambda'$, we find that $P_\chi = 8\lambda'$ and $P_\lambda = 8\chi'$. By the standard methods of constructing a Hamiltonian from a Lagrangian, we thus find

$$\mathcal{H} = \tfrac{1}{8}P_\chi P_\lambda - \frac{A^2}{\chi\lambda} + \frac{\alpha^2\lambda}{2\chi^3}. \tag{18}$$

According to the Hamilton–Jacobi theory, if a generator of canonical transformations of the form $\mathcal{S}(q_i, P_i, t)$ is introduced in such a way that it satisfies the equation

$$\mathcal{H}[q_i, p_i] + \frac{\partial\mathcal{S}}{\partial t} = 0, \tag{19a}$$

where

$$p_i = \frac{\partial\mathcal{S}}{\partial q_i}, \tag{19b}$$

then $\mathcal{S}$ generates transformations to new variables, the P_i and the Q_i ($= \partial\mathcal{S}/\partial P_i$), which are constant in time since the new Hamiltonian—the right-hand side of Eq. (19a)—is identically zero. The fact that the Q_i and P_i are constants allow us to invert these relations to obtain the q_i as functions of time.

Corresponding to Eq. (18), we find the Hamilton–Jacobi equation to be

$$\frac{1}{8}\frac{\partial S}{\partial\chi}\frac{\partial S}{\partial\lambda} - \frac{A^2}{\chi\lambda} + \frac{\alpha^2\lambda}{2\chi^3} + \frac{\partial S}{\partial z} = 0. \tag{20}$$

It is standard to assume that $\mathcal{S}$ can be written in the form $\mathcal{S} = W(\chi, \lambda) - \beta z$, where β is an arbitrary constant. The characteristic equation is then

$$\frac{\partial W}{\partial\chi}\frac{\partial W}{\partial\lambda} - \frac{8A^2}{\chi\lambda} + \frac{4\alpha^2\lambda}{\chi^3} = 8\beta. \tag{21}$$

This equation becomes more transparent when expressed in terms of two new variables u and v, where $u = \chi\lambda$ and $v = \lambda/\chi$. Then (21) becomes, by transposing terms,

$$u^2\left(\frac{\partial W}{\partial u}\right)^2 - 8A^2 - 8\beta u = -4\alpha^2 v^2 + v^2\left(\frac{\partial W}{\partial v}\right)^2. \tag{22}$$

This equation can be separated into two ordinary differential equations on the assumption that $W = F(u) + G(v)$ for, then, the left side of (22) depends on u alone and the right side on v alone. Calling the separation constant k^2, we can write the solutions as

$$F(u) = \int \frac{1}{u}(k^2 + 8A^2 + 8\beta u)^{\frac{1}{2}}\, du, \tag{23a}$$

$$G(u) = \int \frac{1}{v}(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}}\, dv. \tag{23b}$$

In practice, there is an arbitrariness in the generating function, owing to the possibility of minus signs when extracting the square root to achieve (23). This arbitrariness, if it leads to a multiplicity of sets of solutions $\{\chi(z), \lambda(z)\}$, can be eliminated by choosing the one that leads to physically admissible solutions. At this stage, therefore, we write the full generator as

$$\mathcal{S} = \epsilon F(\chi\lambda) + \delta G(\lambda/\chi) - \beta z, \tag{24}$$

where $\epsilon = \pm 1$ and $\delta = \pm 1$.

Two constants of integration appear in $\mathcal{S}$. These quantities, both β and k^2, constitute the new momenta; differentiation of $\mathcal{S}$ with respect to them define the new coordinates which are themselves constants. Thus

$$\frac{\partial\mathcal{S}}{\partial\beta} = Q_\beta = \epsilon\frac{\partial F}{\partial\beta} + \delta\frac{\partial G}{\partial\beta} = z, \tag{25a}$$

$$\frac{\partial\mathcal{S}}{\partial k^2} = Q_{k^2} = \epsilon\frac{\partial F}{\partial k^2} + \delta\frac{\partial G}{\partial k^2}. \tag{25b}$$

These two relations allow us to solve for χ and λ as functions of z.

Consider Eq. (25a). G is independent of β; from the definition of F,

$$\frac{\partial F}{\partial\beta} = 4\int (k^2 + 8A^2 + 8\beta u)^{-\frac{1}{2}}\, du = \frac{1}{\beta}(k^2 + 8A^2 + 8\beta u)^{\frac{1}{2}}. \tag{26}$$

When inserted into (25a), it becomes

$$Q = \frac{\epsilon}{\beta}(k^2 + 8A^2 + 8\beta\chi\lambda)^{\frac{1}{2}} - z. \qquad (27)$$

It is convenient to introduce the boundary conditions $\chi \to 1$ and $\lambda \to 1$ as $r \to \infty$, i.e., as $z \to 0$. This is just the requirement that space become Minkowskian infinitely far from the localized mass distribution.[4] With a rearrangement of terms, Eq. (27) becomes

$$\chi\lambda = 1 + \tfrac{1}{4}\epsilon(k^2 + 8A^2 + 8\beta)^{\frac{1}{2}}z + \tfrac{1}{8}\beta z^2. \quad (28)$$

It will also prove to be convenient to introduce two new constants in place of A and β, $a = (k^2 + 8A^2 + 8\beta)^{\frac{1}{2}}$ and $b = (k^2 + 8A^2)^{\frac{1}{2}}$. In terms of these (28) becomes

$$\chi\lambda = 1 + \tfrac{1}{4}\epsilon az + \tfrac{1}{64}(a^2 - b^2)z^2. \qquad (29)$$

Consider the second relation (25b). According to the definitions of F and G,

$$\frac{\partial F}{\partial k^2} = \frac{1}{2}\int \frac{1}{u}(b^2 + 8\beta u)^{-\frac{1}{2}}\,du$$
$$= \frac{1}{2b}\ln\left|\frac{(b + 8\beta u)^{\frac{1}{2}} - b}{(b + 8\beta u)^{\frac{1}{2}} + b}\right|, \qquad (30a)$$
$$\frac{\partial G}{\partial k^2} = \frac{1}{2}\int \frac{1}{v}(k^2 + 4\alpha^2 v^2)^{-\frac{1}{2}}\,dv$$
$$= \frac{1}{4k}\ln\left|\frac{(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}} - k}{(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}} + k}\right|, \qquad (30b)$$

k being the positive root of k^2. By substituting these expressions into (25b), combining the logarithms, and exponentiating, we can obtain

$$\frac{(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}} - k}{(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}} + k} = B_0'\left|\frac{(b^2 + 8\beta u)^{\frac{1}{2}} + b}{(b^2 + 8\beta u)^{\frac{1}{2}} - b}\right|^{2k\epsilon\delta/b}. \quad (31)$$

Because of the relationship (29), this can be rewritten by absorbing a constant into B_0':

$$\frac{(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}} - k}{(k^2 + 4\alpha^2 v^2)^{\frac{1}{2}} + k} = B_0\left|\frac{1 + \frac{1}{8}\epsilon(a - b)z}{1 + \frac{1}{8}\epsilon(a + b)z}\right|^{2k\epsilon\delta/b}. \quad (32)$$

At this stage B_0 can readily be evaluated by recognizing that, at $z = 0$, $v\ (= \lambda/\chi) = 1$:

$$B_0 = \frac{(k^2 + 4\alpha^2)^{\frac{1}{2}} - k}{(k^2 + 4\alpha^2)^{\frac{1}{2}} + k}. \qquad (33)$$

It is here worth noting that the left side of Eq. (32) is always greater than or equal to zero, but is also always less than unity for any value of v. This is not, in general, true for the right-hand side. Consistency can, however, be guaranteed by the choice $\epsilon = \delta = +1$. For then $1 + \frac{1}{8}(a + b)z$ is always a positive number and appears in the denominator while $1 + \frac{1}{8}(a - b)z$ appears in the numerator and is always less than the denominator.

If we define $B(z)$ according to

$$B(z) = B_0\left|\frac{1 + \frac{1}{8}(a - b)z}{1 + \frac{1}{8}(z + b)z}\right|^{2k/b}, \qquad (34)$$

then Eq. (32) can be inverted to yield

$$v = \lambda/\chi = (k/\alpha)B^{\frac{1}{2}}(z)[1 - B(z)]^{-1}, \qquad (35)$$

where α is the positive root of α^2.

The scalar field can be determined from the differential Eq. (14):

$$\frac{d\ln\phi}{dz} = \frac{A}{(\frac{3}{2} + \omega)^{\frac{1}{2}}\chi\lambda} = \frac{64A(\frac{3}{2} + \omega)^{-\frac{1}{2}}}{64 + 16az + (a^2 - b^2)z^2}. \qquad (36)$$

The second relation arises because of Eq. (29). An immediate integration yields

$$\phi = \phi_0\left|\frac{1 + \frac{1}{8}(a + b)z}{1 + \frac{1}{8}(a - b)z}\right|^{4A/b(\frac{3}{2}+\omega)^{\frac{1}{2}}}. \qquad (37)$$

Here ϕ_0 is an integration constant which can be taken to be the boundary value of ϕ as $r \to \infty$, i.e., at $z = 0$.

A word on the sign of A is necessary. Consider the full differential equation for ϕ,

$$\nabla\cdot(\chi\lambda\nabla\ln\phi) + \lambda\rho_m(x)/(3 + 2\omega)\phi^{\frac{1}{2}}\chi = 0. \quad (38)$$

An integration over all space yields by Gauss' theorem

$$\int ds\cdot\chi\lambda\nabla\ln\phi = -\int d^3x\,\frac{\lambda\rho_m(x)}{(3 + 2\omega)\phi^{\frac{1}{2}}\chi}, \quad (39)$$

the surface integral being taken at infinity. Because (36) is true outside the source, Eq. (39) implies that

$$A = \frac{1}{8\pi(\frac{3}{2} + \omega)^{\frac{1}{2}}}\int d^3x\,\frac{\lambda\rho_m(x)}{\phi^{\frac{1}{2}}\chi}. \qquad (40)$$

Since each of the quantities appearing in the integrand is positive, so is A.

Finally, a comparison of Eqs. (29) and (35) enables us to write the solutions for χ and λ:

$$\chi(z) = [1 + \tfrac{1}{4}az + \tfrac{1}{64}(a^2 - b^2)z^2]^{\frac{1}{2}} \times (\alpha/k)^{\frac{1}{2}}B^{-\frac{1}{4}}(z)[1 - B(z)]^{\frac{1}{2}}, \quad (41a)$$
$$\lambda(z) = [1 + \tfrac{1}{4}az + \tfrac{1}{64}(a^2 - b^2)z^2]^{\frac{1}{2}} \times (k/\alpha)^{\frac{1}{2}}B^{\frac{1}{4}}(z)[1 - B(z)]^{-\frac{1}{2}}. \quad (41b)$$

Equations (37) and (41a, b) comprise the general exterior solution to the spherically symmetric charged mass distribution. The boundary conditions $\chi \to 1$, $\lambda \to 1$ as $r \to \infty$ have been imposed to ensure Minkowskian space at infinity while $\phi \to \phi_0$ as $r \to \infty$. The arbitrary constants a, b, and k are determined by the actual distribution of charge and mass constituting the source.

4. THE MASS IN TERMS OF THE PARAMETERS a, b, AND k

According to Eq. (4) the energy or mass of a system can be recognized in the coefficient of r^{-1} in an asymptotic expansion of $\chi - \chi(r) \sim 1 + m/32\pi r$. Consider, therefore, an expansion of χ in z about $z = 0$, χ being given by (41a). To first order in z, the first factor in χ can be written

$$[1 + \tfrac{1}{4}az + \tfrac{1}{64}(a^2 - b^2)z^2]^{\frac{1}{2}} \simeq 1 + \tfrac{1}{8}az. \quad (42)$$

A first-order expansion of $B(z)$ yields

$$B(z) \simeq B_0(1 - \tfrac{1}{2}kz). \quad (43)$$

Inserting each of these expressions into $\chi(z)$ and retaining terms to first order, we find

$$\chi(z) = 1 + \tfrac{1}{8}z[a + k(1 + B_0)/(1 - B_0)]. \quad (44)$$

According to the definition of B_0 [Eq. (33)], simple algebraic manipulations reveal that

$$(1 + B_0)/(1 - B_0) = (1/k)(k^2 + \alpha^2)^{\frac{1}{2}}. \quad (45)$$

The expansion (44) thus shows that the mass of the system is expressed as

$$m = 4\pi[a + (k^2 + 4\alpha^2)^{\frac{1}{2}}]. \quad (46)$$

5. THE CHARGED SYSTEM IN THE ABSENCE OF THE SCALAR FIELD

The general solution for a charged system reduces to a particularly simple form in the absence of the scalar field. In the above equations a constant scalar field can be mathematically achieved by demanding that A equal zero. A glance at Eq. (37) then shows that the scalar field everywhere assumes its value given at $r \to \infty$. Furthermore, $B(z)$ can be written

$$B(z) = B_0\left(\frac{1 + \frac{1}{8}(a - k)z}{1 + \frac{1}{8}(a + k)z}\right)^2, \quad (47)$$

since $b = k$ when $A = 0$. The first factor in $\chi(z)$ can always be factored according to

$$[1 + \tfrac{1}{4}az + \tfrac{1}{64}(a^2 - b^2)z^2]^{\frac{1}{2}} = [1 + \tfrac{1}{8}(a - b)z]^{\frac{1}{2}}[1 + \tfrac{1}{8}(a + b)z]^{\frac{1}{2}}. \quad (48)$$

By employing each of these expressions in $\chi(z)$, an algebraic reduction allows us to write

$$\chi^2(z) = \{1 + \tfrac{1}{8}[a + (k^2 + 4\alpha^2)^{\frac{1}{2}}]z\}^2 - \tfrac{1}{16}\alpha^2 z^2. \quad (49)$$

Thus, with Eq. (46) and the fact that $\alpha = |e|/4\pi$, we find that $\chi^2(z)$ can be written

$$\chi^2(z) = (1 + mz/32\pi)^2 - (ez/16\pi)^2, \quad (50)$$

where e is the total charge of the system.[5]

6. THE CHARGED MASS SHELL

The mass shell is characterized by essentially replacing $\rho_m(x)$ by $m_0\delta(r - \epsilon)/4\pi r^2$ where m_0 is the bare mass and ϵ is the radius of the shell. It is further characterized by the absence of an electric field in its interior, since the charge is assumed to be distributed uniformly over the surface of the shell. The general solutions for χ, λ, and ϕ, valid for the region exterior to the shell, are the solutions given by Eqs. (37) and (41); the interior solutions can be obtained from the Hamilton–Jacobi equation in which α is taken to be zero, for this effectively imposes the requirement of zero electric field in the interior region. A glance at Eq. (23a) shows that $F(u)$ is unmodified by this requirement; an equation of the form $\chi\lambda = A + Bz + Cz^2$ will thus be obtained. Because the interior of the shell is matter-free, all interior solutions must be well behaved. In particular, the fields must remain finite as $r \to 0$, i.e., as $z \to \infty$. The product $\chi\lambda$ is then constant and reference to Eq. (36) shows that ϕ diverges exponentially as $z \to \infty$ unless it, too, is a constant. Equations (12b) and (12c) then show that both χ and λ satisfy Laplace's equation everywhere within the shell. The only admissible solutions are χ, λ, and ϕ being constants in the interior; the requirement of continuity of the fields thus demands that

$$\chi(r) = \chi(\epsilon), \quad r \le \epsilon, \quad (51a)$$

$$\lambda(r) = \lambda(\epsilon), \quad r \le \epsilon, \quad (51b)$$

$$\phi(r) = \phi(\epsilon), \quad r \le \epsilon. \quad (51c)$$

The undetermined constants appearing in the exterior solutions for χ, λ, and ϕ can now be evaluated by demanding that these solutions and (51) constitute the complete solution of the full differential equations (12) for the shell. Specifically, the presence of the δ function enables us to evaluate by how much the derivative of each field must change as the surface of the shell is crossed. If each of the source equations is integrated across $r = \epsilon$, we find

$$8r^2 \frac{d\chi(r > \epsilon)}{dr}\bigg|_{r=\epsilon} - 8r^2 \frac{d\chi(r < \epsilon)}{dr}\bigg|_{r=\epsilon} = -\frac{m_0}{4\pi\phi^{\frac{1}{2}}(\epsilon^{-1})\chi(\epsilon^{-1})}, \quad (52a)$$

$$8r^2 \frac{d\lambda(r > \epsilon)}{dr}\bigg|_{r=\epsilon} - 8r^2 \frac{d\lambda(r < \epsilon)}{dr}\bigg|_{r=\epsilon} = +\frac{m_0}{4\pi\phi^{\frac{1}{2}}(\epsilon^{-1})\chi(\epsilon^{-1})}, \quad (52b)$$

$$r^2\chi\lambda \frac{d\ln\phi(r > \epsilon)}{dr}\bigg|_{r=\epsilon} - 8r^2\chi\lambda \frac{d\ln\phi(r < \epsilon)}{dr}\bigg|_{r=\epsilon} = -\frac{m_0\lambda(\epsilon^{-1})}{4\pi(3 + 2\omega)\phi^{\frac{1}{2}}(\epsilon^{-1})\chi(\epsilon^{-1})}. \quad (52c)$$

Here $F(\epsilon^{-1})$ means $F(z = \epsilon^{-1})$. Because the derivative of the scalar field and the electric field change by only a finite amount across $r = \epsilon$, the corresponding terms make no contribution to the discontinuity of the derivatives of χ and λ.

These relations can be written in a simple form, first, by recognizing that the second terms of each of the equations vanish because of the constancy of the fields in the interior and, secondly, by means of the transformation $r = z^{-1}$. We find the three relations for determining the three unknown constants to be

$$\left.\frac{d\ln\chi}{dz}\right|_{z=\epsilon^{-1}} = \frac{m_0}{32\pi\phi^{\frac{1}{2}}(\epsilon^{-1})\chi^2(\epsilon^{-1})}, \tag{53a}$$

$$\left.\frac{d\ln\lambda}{dz}\right|_{z=\epsilon^{-1}} = -\frac{m_0}{32\pi\phi^{\frac{1}{2}}(\epsilon^{-1})\chi^2(\epsilon^{-1})}, \tag{53b}$$

$$\left.\frac{d\ln\phi}{dz}\right|_{z=\epsilon^{-1}} = \frac{m_0}{4\pi(3+2\omega)\phi^{\frac{1}{2}}(\epsilon^{-1})\chi^2(\epsilon^{-1})}. \tag{53c}$$

7. THE ELECTRON IN GENERAL RELATIVITY

To restore general relativity, it is necessary to eliminate the effects of the scalar field. By choosing the constant A equal to zero, we force the scalar field to assume the constant value given by the boundary value ϕ_0 [see Eq. (37)]. To be compatible with Eqs. (53), we see that this simultaneously entails choosing ω to be infinity. Equation (53c) is then identically satisfied; Eq. (53a) can be written

$$\left.\frac{d\ln\chi^2}{dz}\right|_{z=\epsilon^{-1}} = \frac{m_0}{16\pi\phi_0^{\frac{1}{2}}\chi^2(\epsilon^{-1})} \tag{54}$$

or

$$\left.\frac{d\chi^2}{dz}\right|_{z=\epsilon^{-1}} = \frac{m_0}{16\pi\phi_0^{\frac{1}{2}}}. \tag{55}$$

Referring to Eq. (50), we find that

$$\left.\frac{d\chi^2}{dz}\right|_{\epsilon^{-1}} = 2\frac{m}{32\pi}\left(1+\frac{m}{32\pi\epsilon}\right) - 2\left(\frac{e}{16\pi}\right)^2\frac{1}{\epsilon}. \tag{56}$$

Combining (55) and (56) enables us to express the mass of the charged shell as

$$m = 16\pi\{-\epsilon + [\epsilon^2 + (e/8\pi)^2 + m_0\epsilon/8\pi\phi_0^{\frac{1}{2}}]^{\frac{1}{2}}\}. \tag{57}$$

The electron in general relativity is characterized by a point charge source and thus instead of a shell δ function a point δ function should be used. In the limit ϵ becomes zero, the solution of the static fields for the electron is obtained where $m = 2\,|e|$ from (57) and from (50):

$$\chi^2(z) = 1 + (|e|/8\pi)z, \tag{58}$$

while

$$\chi_2\,(r=0) = \chi^2(\epsilon^{-1}) = [1 + m(\epsilon)/32\pi\epsilon]^2 - (e/16\pi)^2\epsilon^{-2} \tag{59}$$

and diverges as ϵ^{-1} as ϵ approaches zero. This is just the solution of the self-energy problem as found by Arnowitt, Deser, and Misner.[5]

8. THE ELECTRON IN THE SCALAR-TENSOR THEORY

The relations (53) do not reduce to a tractable set of equations for the unknown constants appearing in the general solutions. However, some arguments can be made which support the idea of Paper I that the solution to the self-energy problem of the electron in the scalar–tensor theory is the same as in general relativity; that is, the function χ is unchanged from the general relativity form, the mass of the system is $2\,|e|$, and the scalar field is constant everywhere outside the source.

For our purposes it is more convenient to cast the relations (53) into different forms by use of the functional forms of χ, λ, and ϕ. Note, first, that Eqs. (34) and (37) allow us to write

$$B = B_0(\phi_0/\phi)^{(\frac{3}{2}+\omega)^{\frac{1}{2}}k/2A} \tag{60}$$

and that, in terms of the definitions of a and b, A is written as $\frac{1}{2}[\frac{1}{2}(b^2 - k^2)]^{\frac{1}{2}}$. B is thus a functional of ϕ, and a reference to Eqs. (41a,b) shows that χ and λ can be written

$$\chi(z) = U^{\frac{1}{2}}(z)V^{\frac{1}{2}}(z), \tag{61a}$$

$$\lambda(z) = U^{\frac{1}{2}}(z)V^{-\frac{1}{2}}(z), \tag{61b}$$

where

$$U(z) = 1 + \tfrac{1}{4}az + \tfrac{1}{64}(a^2 - b^2)z^2, \tag{62a}$$

$$V(z) = (\alpha/k)B^{-\frac{1}{2}}(1 - B). \tag{62b}$$

Because of (61a, b) we can write the derivatives of the logarithms of χ and λ as

$$\frac{d\ln\chi}{dz} = \frac{1}{2}\frac{d\ln U}{dz} + \frac{1}{2}\frac{d\ln V}{dz}, \tag{63a}$$

$$\frac{d\ln\lambda}{dz} = \frac{1}{2}\frac{d\ln U}{dz} - \frac{1}{2}\frac{d\ln V}{dz}. \tag{63b}$$

The addition of Eqs. (53a) and (53b) gives us the requirement that

$$\left.\frac{d\ln\chi\lambda}{dz}\right|_{\epsilon^{-1}} = 0. \tag{64}$$

Equations (63) show, then, that the derivative of U must vanish at $z = \epsilon^{-1}$ or

$$8a\epsilon = b^2 - a^2. \tag{65}$$

This shows that b is greater than a for the charged shell. We assume that in the limit of ϵ approaching zero (characterizing the electron) a does not vanish; b must then approach a as ϵ approaches zero.

We next consider Eqs. (53a) and (53c). From them we see that we can write

$$\left.\frac{d\ln\chi}{dz}\right|_{\epsilon^{-1}} = \frac{(\frac{3}{2}+\omega)}{4}\left.\frac{d\ln\phi}{dz}\right|_{\epsilon^{-1}}. \tag{66}$$

Because the derivative of U vanishes at ϵ^{-1}, this statement is equivalent to the condition

$$\frac{1}{2}\left.\frac{d\ln V}{dz}\right|_{\epsilon^{-1}} = \frac{(\frac{3}{2}+\omega)}{4}\left.\frac{d\ln\phi}{dz}\right|_{\epsilon^{-1}} \tag{67}$$

or

$$-\frac{(1+B)}{4(1-B)}\left.\frac{d\ln B}{dz}\right|_{\epsilon^{-1}} = \frac{(\frac{3}{2}+\omega)}{4}\left.\frac{d\ln\phi}{dz}\right|_{\epsilon^{-1}}. \tag{68}$$

Because of Eq. (60) it is true that

$$\left.\frac{d\ln B}{dz}\right|_{\epsilon^{-1}} = -\frac{(\frac{3}{2}+\omega)^{\frac{1}{2}}k}{2A}\left.\frac{d\ln\phi}{dz}\right|_{\epsilon^{-1}}. \tag{69}$$

Thus, by (68),

$$\left[\frac{k}{A}\frac{(1+B)}{(1-B)} - 2(\tfrac{3}{2}+\omega)^{\frac{1}{2}}\right]_{\epsilon^{-1}}\left.\frac{d\ln\phi}{dz}\right|_{\epsilon^{-1}} = 0. \tag{70}$$

The derivative of the logarithm of ϕ cannot vanish at $z=\epsilon^{-1}$. Straightforward differentiation of ϕ given by Eq. (37) yields

$$\left.\frac{d\ln\phi}{dz}\right|_{\epsilon^{-1}} = \frac{16\epsilon^2 b}{(8\epsilon+a)^2-b^2}\,\frac{4A}{b(\frac{3}{2}+\omega)^{\frac{1}{2}}}. \tag{71}$$

The bracketed term must therefore vanish. Solving for $B(\epsilon^{-1})$, we obtain

$$B(\epsilon^{-1}) = \frac{2A(\frac{3}{2}+\omega)^{\frac{1}{2}}-k}{2A(\frac{3}{2}+\omega)^{\frac{1}{2}}+k}. \tag{72}$$

$B(\epsilon^{-1})$ can be written, by use of Eq. (65),

$$B(\epsilon^{-1}) = B_0\left|\frac{8\epsilon a+a^2-ab}{8a+a^2+ab}\right|^{2k/b} = B_0\left|\frac{b-a}{b+a}\right|^{2k/b}, \tag{73}$$

and thus the second constraint on the three unknown constants can be written, instead of (72), as

$$\frac{b-a}{b+a} = \left|\frac{(k^2+4\alpha^2)^{\frac{1}{2}}+k}{(k^2+4\alpha^2)^{\frac{1}{2}}-k}\,\frac{2A(\frac{3}{2}+\omega)^{\frac{1}{2}}-k}{2A(\frac{3}{2}+\omega)^{\frac{1}{2}}+k}\right|^{b/2k}. \tag{74}$$

According to the assumption that a approaches a nonzero limit as ϵ approaches zero, we have shown that b must approach the limit of a. The left-hand side of Eq. (74) is less than unity and must also approach zero as ϵ approaches zero. This gives us the behavior of k. In order for the fraction on the right-hand side to vanish, k must satisfy in the limit

$$2A(\tfrac{3}{2}+\omega)^{\frac{1}{2}} = k$$

or

$$k = b[(7+2\omega)/(3+2\omega)]^{\frac{1}{2}}. \tag{75}$$

On first sight, the thought occurs that the limit of k approaching zero will also satisfy the limit of Eq. (74), for then the fraction on the right side is raised to higher powers as k approaches zero. However, for small k, the interior fraction becomes

$$\frac{(k^2+4\alpha^2)^{\frac{1}{2}}+k}{(k^2+4\alpha^2)^{\frac{1}{2}}-k}\,\frac{2A(\frac{3}{2}+\omega)^{\frac{1}{2}}-k}{2A(\frac{3}{2}+\omega)^{\frac{1}{2}}+k} \simeq 1 - k\left(\frac{4}{(3+2\omega)^{\frac{1}{2}}b}-\frac{1}{\alpha}\right), \tag{76}$$

and the limit is

$$\lim_{k\to0}\left|1-k\left(\frac{4}{(3+2\omega)^{\frac{1}{2}}b}-\frac{1}{\alpha}\right)\right|^{b/2k} = \exp\left(\frac{b}{2\alpha}-\frac{2}{(3+2\omega)^{\frac{1}{2}}}\right) \tag{77}$$

and is not zero, as the left side of (74) must be.

We thus see that, on the assumption that a approaches a nonzero limit as ϵ approaches zero, b approaches the same limit and k approaches the limit given by (75). Is this compatible with the last constraint to be extracted from Eqs. (53)?

Consider Eq. (53a) alone. We have already shown that the derivative of the logarithm of χ satisfies the following equation at $z=\epsilon^{-1}$:

$$\left.\frac{d\ln\chi}{dz}\right|_{\epsilon^{-1}} = -\frac{(1+B)}{4(1-B)}\left.\frac{d\ln B}{dz}\right|_{\epsilon^{-1}}.$$

According to (60) and (71), we can write

$$\left.\frac{d\ln B}{dz}\right|_{\epsilon^{-1}} = -\frac{(\frac{3}{2}+\omega)^{\frac{1}{2}}k}{2A}\left.\frac{d\ln\phi}{dz}\right|_{\epsilon-1} = -\frac{32ka^2\epsilon^2}{b^2(b^2-a^2)}. \tag{78}$$

Relation (65) was used to simplify the result in this equation.

Referring to Eqs. (61a) and (62), we find

$$\chi^2(\epsilon^{-1}) = \frac{b^2(b^2-a^2)}{64\epsilon^2a^2}\left(\frac{\alpha}{k}\right)B^{-\frac{1}{2}}(\epsilon^{-1})[1-B(\epsilon^{-1})], \tag{79}$$

again by simple algebraic reduction.

By substitution of these expressions into the relation (53a), we see that a number of cancellations arise with the result

$$\frac{[1+B(\epsilon^{-1})]\phi^{\frac{1}{2}}(\epsilon^{-1})}{B^{\frac{1}{2}}(\epsilon^{-1})}=\frac{m_0}{4\pi\alpha}. \quad (80)$$

Equation (73) shows that $B(\epsilon^{-1})$ approaches zero; the ratio of k to b must be a finite constant in the limit according to (75). If (80) is to be compatible with the initial assumption that a does not approach zero, then $\phi^{\frac{1}{2}}(\epsilon^{-1})$ must approach zero as fast as $B^{\frac{1}{2}}(\epsilon^{-1})$. But, according to (60) and (75),

$$\phi^{\frac{1}{2}}=\phi_0^{\frac{1}{2}}(B_0/B)^{1/(3+2\omega)} \quad (81)$$

and, in fact, *diverges* as ϵ approaches zero, since B approaches zero. This contradiction shows that a must approach zero in the limit. Equation (65) then shows that b must also approach zero and, because A is a positive quantity by (40) and $2A=[\frac{1}{2}(b^2-k^2)]^{\frac{1}{2}}$, k must also approach zero.

These arguments do not tell us the rates at which these quantities reach their limits, but the limits do tell us something about the exterior solutions for the electron in the scalar-tensor theory. According to Eq. (46), the mass of the charged shell is written

$$m=4\pi[a+(k^2+4\alpha^2)^{\frac{1}{2}}].$$

The above limits show that the mass of the electron is $m=8\pi\alpha=2\,|e|$ as is true in general relativity. Also, the external scalar field (37)

$$\phi=\phi_0\left|\frac{1+\frac{1}{8}(a+b)z}{1+\frac{1}{8}(a-b)z}\right|^{[2/(3+2\omega)^{\frac{1}{2}}](1-k^2/b^2)^{\frac{1}{2}}}$$

approaches ϕ_0 for any $z\neq\epsilon^{-1}$ because a and b approach zero and because k must always be less than b.

Note that the arbitrariness of the boundary value of the scalar field, the charge and bare mass, and Eq. (80) demand that the limit of a/b be nonzero. If the converse were true, the following would occur. The limit of zero k would demand that B_0 approach unity. At $z=\epsilon^{-1}$ we have shown (73) that

$$B(\epsilon^{-1})=B_0\left|\frac{1-a/b}{1-a/b}\right|^{2k/b}.$$

Since k must always be less than b, if the limit of a/b were zero, the limit of $B(\epsilon^{-1})$ would be unity also. In like manner, $\phi^{\frac{1}{2}}(\epsilon^{-1})$ is written, from (37) and (65), as

$$\phi^{\frac{1}{2}}(\epsilon^{-1})=\phi_0^{\frac{1}{2}}\left|\frac{1+a/b}{1-a/b}\right|^{[(1-k^2/b^2)/(3+2\omega)]^{\frac{1}{2}}} \quad (82)$$

and would approach $\phi_0^{\frac{1}{2}}$ in the limit. Equation (80) would then demand in the limit that

$$2\phi_0^{\frac{1}{2}}=m_0/4\pi\alpha,$$

which is, in general, not true.

We see, then, that the rigorous solution to the self-energy problem for the electron in the scalar-tensor theory consists of a constant scalar field equal to the boundary value at infinity everywhere outside the electron and that an essential singularity in the scalar develops at the electron itself; Eq. (82) shows that $\phi(\epsilon^{-1})$ is always greater than ϕ_0. The mass of the electron is $2\,|e|$ or, in more usual units, $G_0^{-\frac{1}{2}}(e^2/4\pi)^{\frac{1}{2}}$ and, for a constant external scalar field, the function χ must of necessity be given by (58), exactly as in general relativity.

ACKNOWLEDGMENTS

I wish to express my sincerest gratitude to Professor J. Weber for many helpful discussions throughout these calculations and to numerous friends, in particular to Dr. B. Moritz, Dr. D. Kaup, and Mr. V. Moncrief.

* Supported in part by the National Science Foundation and NASA grant NGR21-002-010.

† Based on the author's Ph.D. thesis, University of Maryland, 1969.

‡ Present address: Institute for Advanced Study, Princeton, New Jersey 08540.

[1] E. Toton, J. Math. Phys. **11**, 1714 (1970).

[2] L. Witten, *Gravitation: An Introduction To Current Research* (Wiley, New York, 1962), p. 246.

[3] R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **120**, 319 (1960).

[4] If in isotropic coordinates the line element is written in the form

$$ds^2=-N^2dt^2+\chi^4(dx^2+dy^2+dz^2),$$

then the equation for the scalar field becomes

$$\Box \ln\phi\rightarrow N^{-1}\chi^{-6}\nabla\cdot(N\chi^2\nabla\ln\phi)=0.$$

Comparison with (12a) outside the distribution yields $N=\lambda/\chi$ and thus the boundary conditiions $\chi\rightarrow 1$, $\lambda\rightarrow 1$ are appropriate for Minkowski space.

[5] See Ref. 3, p. 317.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Hyperplane Independence of the Scattering Amplitude*

GORDON N. FLEMING AND JAMES T. MILLER, JR.
Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 19 January 1970)

In the context of the formal theory of scattering the scattering amplitude on the energy–momentum shell is shown to be hyperplane independent even in the presence of hyperplane-dependent interaction "potentials." Also, a formal field-theoretic example of consistent hyperplane-dependent potentials is presented.

1. INTRODUCTION

In this paper, the hyperplane independence[1] of the scattering amplitude in the context of the formal theory of scattering is established. One motivation for this study is the apparently widespread notion that the explicit introduction of a timelike vector or similar device to characterize the conditions surrounding measurements must destroy the relativistic invariance of the theory regardless of how manifestly covariant the resulting formalism may appear.[2] This view is erroneous, and it occurs to one that the demonstration of the hyperplane independence of the scattering amplitude when the interaction *is* hyperplane dependent would do more to still the critics than the general, somewhat philosophical, arguments already in print.

Beyond this, a formal solution of the relations that must hold between the interaction parts of the Poincaré generators, and which play a crucial role in the aforementioned demonstration, is obtained, which is considerably more general than the solution provided by conventional local quantum field theory.

A source of confusion about hyperplane dependence is the question of the hypothetical status of the hyperplane dependence of the interaction parts of the Poincaré generators. On one hand, the quantitative details of the hyperplane dependence of the interaction parts are model dependent and, therefore, hypothetical to the extent that the model of the interactions being employed is hypothetical. On the other hand, the mere existence of hyperplane dependence is not hypothetical, for it is a consequence of the fact that the decomposition of the Poincaré generators into free and interaction parts is always carried out on, or with reference to, a particular spacelike hyperplane.[3] The more conventional view that such decomposition is carried out in, or with reference to, a particular inertial reference frame gives undue emphasis to processes, events, and analytical procedures referring to definite instants of time and ignores the equivalence of describing the instantaneous of any frame as the hyperplane-associated of a given frame.

2. PARTS OF THE HYPERPLANE-DEPENDENT INTERACTION

In an earlier paper,[3] it has been shown that, if the Poincaré generators P_μ and $M_{\mu\nu}$ are separated into free parts $P_\mu^{(0)}(\eta, \tau)$ and $M_{\mu\nu}^{(0)}(\eta, \tau)$ and interaction parts on the (η, τ) hyperplane, then the interaction parts have the form

$$P_\mu - P_\mu^{(0)}(\eta, \tau) = \eta_\mu V(\eta, \tau), \qquad (2.1)$$

$$M_{\mu\nu} - M_{\mu\nu}^{(0)}(\eta, \tau) = \eta_\mu U_\nu(\eta, \tau) - \eta_\nu U_\mu(\eta, \tau). \qquad (2.2)$$

The operators V and U_μ are scalar and vector operators, respectively, and, therefore,[4]

$$[P_\mu, V(\eta, \tau)] = -i\hbar\eta_\mu \frac{\partial V(\eta, \tau)}{\partial \tau}, \qquad (2.3a)$$

$$[M_{\mu\nu}, V(\eta, \tau)] = i\hbar(\eta_\nu\delta_\mu - \eta_\mu\delta_\nu)V(\eta, \tau), \qquad (2.3b)$$

$$[P_\mu, U_\nu(\eta, \tau)] = i\hbar\left((g_{\mu\nu} - \eta_\mu\eta_\nu)V(\eta, \tau) - \eta_\mu \frac{\partial U_\nu(\eta, \tau)}{\partial \tau}\right), \qquad (2.3c)$$

$$[M_{\mu\nu}, U_\lambda(\eta, \tau)] = i\hbar[g_{\nu\lambda}U_\mu(\eta, \tau) - g_{\mu\lambda}U_\nu(\eta, \tau) + (\eta_\nu\delta_\mu - \eta_\mu\delta_\nu)U_\lambda(\eta, \tau)]. \qquad (2.3d)$$

Furthermore, the "potentials" V and U_μ cannot be chosen independently, but, rather, must satisfy the nonlinear coupled equations[5]

$$[U_\mu, V] = i\hbar\left(\frac{\partial U_\mu}{\partial \tau} + \delta_\mu V\right) \qquad (2.4a)$$

and

$$[U_\mu, U_\nu] = i\hbar[(\delta_\mu - \eta_\mu)U_\nu - (\delta_\nu - \eta_\nu)U_\mu], \qquad (2.4b)$$

where the hyperplane dependence of the potentials is suppressed for convenience.

If a V and a U_μ can be found which satisfy all of the foregoing relations, then the free generators $P_\mu^{(0)}$ and $M_{\mu\nu}^{(0)}$, as well as the total generators P_μ and $M_{\mu\nu}$, satisfy the Lie algebra of the Poincaré group and together provide a description of interactions between the stable particles defined by the free generators in a relativistically covariant manner.

3. HYPERPLANE DEPENDENCE OF INITIAL AND FINAL STATES

Consider the asymptotically free incoming and outgoing states of scattering theory with asymptotic configurations α. Under a Poincaré transformation $\{\Lambda, a\}$, implemented by the unitary operator $U(\Lambda, a)$, we have

$$|\alpha'(\pm)\rangle \equiv U(\Lambda, a)\,|\alpha(\pm)\rangle = \mathcal{U}_\alpha^\beta(\Lambda, a)\,|\beta(\pm)\rangle, \quad (3.1)$$

where the repeated β symbol indicates a sum–integral over a complete set. In the specific case in which the vectors $|\alpha(\pm)\rangle$ are momentum–spin eigenvectors with the spin components or state indicated by tensor and/or spinor indices, we have

$$U(\Lambda, a)\,|P_1\alpha_1, P_2\alpha_2, \cdots, P_n\alpha_n(\pm)\rangle = e^{(i/\hbar)(p_1'+\cdots+P_n')a}\mathcal{U}^{\beta_1\cdots\beta_n}_{\alpha_1\cdots\alpha_n}(\Lambda)\,|P_1'\beta_1, \cdots, P_n'\beta_n(\pm)\rangle, \tag{3.2}$$

where $P_i' = \Lambda P_i$ and the matrix $\mathcal{U}$ belongs to a finite-dimensional representation of the homogeneous Lorentz group. The α_i here indicate the set of tensor and/or spinor indices used to denote the spin state of particle i. The important consideration is that, with the use of such indices, the transformation matrix is independent of the momenta P_i.[6] As is well known, eigenvectors of the z components of the particle spins or helicity eigenvectors do not share this feature.

Now, in the formal theory of scattering, one employs not only the asymptotic momentum–spin eigenvectors (or scattering states, as they are sometimes called) but the momentum–spin eigenvectors of the free generators as well. Since the free generators are defined on spacelike hyperplanes and differ from hyperplane to hyperplane in the presence of interactions, it follows that the corresponding momentum–spin eigenvectors will also be hyperplane dependent in the presence of interactions. In fact, if one does not employ the tensor–spinor formalism for spins but introduces helicity states, say, then hyperplane dependence of the states is present even for free systems.[7] For our purpose, it is desirable to keep the hyperplane dependence of the initial and final states of a collision entirely dynamical in origin. Only in that case is the scattering amplitude hyperplane independent.

Under the *free* Poincaré group, i.e., the group of operators $U^{(0)}(\Lambda, a; \eta 0)$ formed from the free generators $P^{(0)}_\mu(\eta, 0)$ and $M^{(0)}_{\mu\nu}(\eta, 0)$, the momentum–spin eigenvectors satisfy

$$U^{(0)}(\Lambda, a; \eta)\,|P_1\alpha_1, \cdots, P_n\alpha_n; \eta 0\rangle = e^{(i/\hbar)\Lambda(P_1+\cdots+P_n)a}\,\mathcal{U}^{\beta_1}_{\alpha_1}(\Lambda)\cdots\mathcal{U}^{\beta_n}_{\alpha_n}(\Lambda) \times |\Lambda P_1, \beta_1, \cdots, \Lambda P_n\beta_n; n0\rangle. \tag{3.3}$$

Under the full Poincaré group generated by the P_μ and $M_{\mu\nu}$, the corresponding equation is

$$U(\Lambda, a)\,|p_1\alpha_1, \cdots, P_n\alpha_n; n\tau\rangle = \exp\{(i/\hbar)[\Lambda(P_1+\cdots+P_n) - \Lambda\eta(P_1+\cdots+P_n)\eta]_\mu a^\mu\} \times \mathcal{U}^{\beta_1}_{\alpha_1}(\Lambda)\cdots\mathcal{U}^{\beta_n}_{\alpha_n}(\Lambda) \times |\Lambda P_1\beta_1, \cdots, \Lambda P_n\beta_n; \Lambda\eta, \tau + a\Lambda\eta\rangle. \tag{3.4}$$

Combining (3.3) and (3.4) and considering infinitesimal transformations, one finally obtains

$$V(\eta,\tau)\,|P_1\alpha_1, \cdots, P_n\alpha_n; \eta\tau\rangle = -\left(\eta(P_1+\cdots+P_n) + i\hbar\frac{\partial}{\partial\tau}\right)|P_1\alpha_1, \cdots, P_n\alpha_n; \eta\tau\rangle \tag{3.5}$$

and

$$U_\mu(\eta,\tau)\,|P_1\alpha_1, \cdots, P_n\alpha_n; \eta\tau\rangle = [i\hbar\delta_\mu + \tau(P_\mu - \eta_\mu\eta P)]\,|P_1\alpha_1, \cdots, P_n\alpha_n; n\tau\rangle. \tag{3.6}$$

4. THE SCATTERING AMPLITUDE

Denoting the initial state by $A = (P_1\alpha_1, \cdots, P_n\alpha_n)$ and the final state by

$$B = (P_1'\alpha_1', \cdots, P_n', \alpha_{n'}'),$$

we see that the relation between the S matrix and the scattering amplitude in the conventional instantaneous formalism is

$$\langle B(+) \mid A(-)\rangle = \delta(B, A) - 2\pi i\delta(P_{BO} - P_{AO})\,\langle B|\,T(P_{AO})\,|A\rangle, \tag{4.1}$$

$$T(P_{AO}) \equiv V + V(P_{AO} - P_O - iO)^{-1}V, \tag{4.2}$$

and the states $\langle B|$ and $|A\rangle$ refer to the time $t = 0$. The generalization to arbitrary hyperplane orientation is[1]

$$\langle B(+) \mid A(-)\rangle = \delta(B, A) - 2\pi i\delta(\eta P_B - \eta P_A)\,\langle B; \eta|\,T(\eta P_A, \eta)\,|A; \eta\rangle, \tag{4.3}$$

where

$$T(\eta P_A, \eta) = V(\eta) + V(\eta)(\eta P_A - \eta P - iO)^{-1}V(\eta), \tag{4.4}$$

$V(\eta) = V(\eta, 0)$, and $|A; \eta\rangle = |A; \eta 0\rangle$. The validity of (4.1) clearly requires that

$$\delta_\mu[\delta(\eta P_B - \eta P_A)\,\langle B; \eta|\,T(\eta P_A, \eta)\,|A; \eta\rangle] = 0. \tag{4.5}$$

Now[4]

$$\delta_\mu\delta(\eta P_B - \eta P_A) = [(P_B - P_A)_\mu - \eta_\mu\eta(P_B - P_A)]\delta'(\eta P_B - \eta P_A), \tag{4.6}$$

and, since

$$\langle B, \eta|\,T(\eta P_A, \eta)\,|A, \eta\rangle \propto \delta^3_\eta(P_B - P_A), \tag{4.7}$$

where

$$\delta^4(P_B - P_A) \equiv \delta(\eta P_B - \eta P_A)\delta^3_\eta(P_B - P_A), \tag{4.8}$$

it follows that the contribution to (4.5) from differentiating the δ function vanishes by itself. Hence, (4.5) is replaced by

$$\delta(\eta P_B - \eta P_A)\delta_\mu\,\langle B, \eta|\,T(\eta P_A, \eta)\,|A, \eta\rangle = 0. \tag{4.9}$$

For the rest of the calculation, it is convenient to introduce an arbitrary *complex* 4-momentum argument P into $T(\eta P, \eta)$, with $P_B \neq P \neq P_A$, in general. To distinguish between the c number ηP and the operator ηP, the symbol $H(\eta) \equiv \eta P$ is introduced for the latter. Thus,

$$T(\eta P,\eta)=V(\eta)+V(\eta)[\eta P-H(\eta)]^{-1}V(\eta). \quad (4.10)$$

From (3.6),

$$\delta_\mu\langle B,\eta|\,T(\eta P,\eta)\,|A,\eta\rangle=\langle B,\eta|\,\delta_\mu T(\eta P,\eta)+(i\hbar)^{-1}[T(\eta P,\eta),U_\mu(\eta)]\,|A,\eta\rangle, \quad (4.11)$$

and applying (2.3) and (2.4) to

$$\delta_\mu T=\delta_\mu V+\delta_\mu V(\eta P-H)^{-1}V+V(\eta P-H)^{-1}\delta_\mu V-V(\eta P-H)^{-1}(P_\mu-\eta_\mu\eta P-K_\mu)\times(\eta P-H)^{-1}V \quad (4.12)$$

yields

$$\begin{aligned}i\hbar\delta_\mu T&=([H,U_\mu]-[V,U_\mu])+([H,U_\mu]-[V,U_\mu])(\eta P-H)^{-1}V\\&\quad+V(\eta P-H)^{-1}([H,U_\mu]-[V,U_\mu])-V(\eta P-H)^{-1}(P_\mu-\eta_\mu\eta P-K_\mu)\times(\eta P-H)^{-1}V\\&=-[T,U_\mu]+[H,U_\mu]+[H,U_\mu](\eta P-H)^{-1}V+V(\eta P-H)^{-1}[H,U_\mu]\\&\quad+V[(\eta P-H)^{-1},U_\mu]V-V(\eta P-H)^{-1}(P_\mu-\eta_\mu\eta P-K_\mu)\times(\eta P-H)^{-1}V\\&=-[T,U_\mu]+[H,U_\mu]+[H,U_\mu](\eta P-H)^{-1}V+V(\eta P-H)^{-1}[H,U_\mu]\\&\quad+V(\eta P-H)^{-1}[H,U_\mu](\eta P-H)^{-1}V-V(\eta P-H)^{-1}(P_\mu-\eta_\mu\eta P-K_\mu)\times(\eta P-H)^{-1}V\\&=-[T,U_\mu]+(\eta P-H^{(0)})(\eta P-H)^{-1}\times[H,U_\mu](\eta P-H)^{-1}(\eta P-H^{(0)})\\&\quad-V(\eta P-H)^{-1}(P_\mu-\eta_\mu\eta P-K_\mu)\times(\eta P-H)^{-1}V, \end{aligned} \quad (4.13)$$

where K_μ is the operator $P_\mu-\eta_\mu\eta P$ and $H^{(0)}\equiv\eta P^{(0)}$. Substituting this into (4.11) yields

$$\begin{aligned}i\hbar\delta_\mu\langle B,\eta|\,T(\eta p,\eta)\,|A,\eta\rangle&=(\eta P-\eta P_B)\langle B,\eta|\,[(\eta P-H)^{-1},U_\mu]\,|A,\eta\rangle\\&\quad\times(\eta P-\eta P_A)-[(P-P_A)_\mu-\eta_\mu\eta(P-P_A)]\\&\quad\times\langle B,\eta|\,V(\eta P-H)^{-2}V\,|A,\eta\rangle.\end{aligned} \quad (4.14)$$

Now put

$$P_\mu=P_{A\mu}-i\epsilon_\mu. \quad (4.15)$$

Then

$$\begin{aligned}&\delta(\eta P_B-\eta P_A)\delta_\mu\langle B,\eta|\,T(\eta P_A-i\eta\epsilon,\eta)\,|A,\eta\rangle\\&=\delta(\eta P_B-\eta P_A)\{-(\eta\epsilon)^2\\&\times\langle B,\eta|\,[(\eta P_A-H-i\eta\epsilon)^{-1},U_\mu]\,|A,\eta\rangle\\&+i(\epsilon_\mu-\eta_\mu\eta\epsilon)\langle B,\eta|\,V(\eta P_A-H-i\eta\epsilon)^{-2}V\,|A,\eta\rangle\}.\end{aligned} \quad (4.16)$$

The limits inside the matrix elements as $\eta\epsilon\to 0+$ yield well-defined generalized functions in momentum space. Thus,

$$\begin{aligned}&\langle B,\eta|\,[(\eta P_A-H-i\eta\epsilon)^{-1},U_\mu]\,|A,\eta\rangle\\&\to\langle B,\eta|\,[\mathcal{P}/(\eta P_A-H)+i\pi\delta(\eta P_A-H),U_\mu]\,|A,\eta\rangle,\end{aligned} \quad (4.17)$$

and, from (2.3a) and (2.3c), is proportional to

$$\partial\delta_\eta^3(P_A-P_B)/\partial P_{A\mu}.$$

Similarly,

$$\begin{aligned}&\langle B,\eta|\,V(\eta P_A-H-i\eta\epsilon)^{-2}V\,|A,\eta\rangle\\&=-\langle B,\eta|\,V\left(\frac{d}{dy}(y-H-i\eta\epsilon)^{-1}\right)_{y=\eta P_A}V\,|A,\eta\rangle\\&\to\langle B,\eta|\,V[-\mathcal{P}/(\eta P_A-H)^2+i\pi\delta'(\eta P_A-H)]\,V\,|A,\eta\rangle,\end{aligned} \quad (4.18)$$

which is proportional to $\delta_\eta^3(P_A-P_B)$. Hence, in the limit $\eta\epsilon\to 0+$, Eq. (4.16) vanishes and (4.9) and (4.5) are satisfied.

5. SOLUTION OF THE COMMUTATION RELATIONS

Unless one can find examples of hyperplane-dependent potentials beyond the familiar cases provided by local quantum field theory, the foregoing considerations are academic, since one knows that conventional field theory does not lead to hyperplane dependence of the S matrix.[8] The difficulty, of course, is in satisfying all the commutation relations, involving the potentials, which occur in Sec. 2, i.e., Eqs. (2.3) and (2.4). These equations are special cases of commutation relations that were studied in an earlier paper[3] and a formal expression for their most general solution was then presented. This general solution, however, is sufficiently complicated to be of questionable practical value so that one is motivated to look for less general but simpler solutions.

A simple solution of a field-theoretic character exists which allows for bona fide hyperplane dependence at the field-theoretic level. The solution retains a formal character in that the problem is merely reduced to that of finding a single scalar hyperplane-dependent Hermitian field operator with a particular equal-hyperplane commutation relation satisfied.

Thus, consider an Hermitian field operator $\mathcal{V}(x,\eta)$ satisfying

$$U^{-1}(\Lambda,a)\mathcal{V}(\Lambda x+a,\Lambda n)U(\Lambda,a)=\mathcal{V}(x,\eta), \quad (5.1)$$

or

$$[P_\mu,\mathcal{V}(x,\eta)]=-i\hbar\partial_\mu\mathcal{V}(x,\eta), \quad (5.2a)$$

$$[M_{\mu\nu},\mathcal{V}(x,\eta)]=i\hbar(x_\nu\partial_\mu-x_\mu\partial_\nu+\eta_\nu\delta_\mu-\eta_\mu\delta_\nu)\times\mathcal{V}(x,\eta). \quad (5.2b)$$

In other words, the field is a scalar field. If the potentials V and U_μ are constructed from $\mathcal{V}$ via

$$V(\eta,\tau)\equiv\int d^4x\,\delta(\eta x-\tau)\mathcal{V}(x,\eta) \quad (5.3a)$$

and

$$U_\mu(\eta,\tau) \equiv \int d^4x\, \delta(nx-\tau)(x_\mu - \eta_\mu \eta x)\mathcal{V}(x,\eta), \tag{5.3b}$$

then Eqs. (2.3) are automatically satisfied, provided only that $\mathcal{V}(x,\eta)$ "vanishes" sufficiently rapidly in the spacelike directions orthogonal to η, so that integration by parts can be performed without regard to surface terms.

With this construction, the solution of the nonlinear relations (2.4) clearly requires a statement concerning the equal hyperplane commutation relation of $\mathcal{V}$ with itself. A solution is provided by the relation[9]

$$\delta(\eta x' - \eta x)[\mathcal{V}(x',\eta), \mathcal{V}(x,\eta)] = i\hbar[\partial^\mu \delta^4(x'-x) \times \delta_\mu \mathcal{V}(x',\eta) - \partial'^\mu \delta^4(x'-x)\delta_\mu \mathcal{V}(x,\eta)]. \tag{5.4}$$

Thus,

$$\delta_\mu V(\eta,\tau) = \int d^4x[\delta'(\eta x-\tau)(x_\mu - \eta_\mu \eta x)\mathcal{V}(x,\eta) + \delta(\eta x - \tau)\delta_\mu \mathcal{V}(x,\eta)] \tag{5.5}$$

and

$$\frac{\partial U_\mu(\eta,\tau)}{\partial \tau} = -\int d^4x\, \delta'(\eta x - \tau)(x_\mu - \eta_\mu \eta x)\mathcal{V}(x,\eta), \tag{5.6}$$

so that, from (2.4a),

$$i\hbar \int d^4x\, \delta(\eta x - \tau)\delta_\mu \mathcal{V}(x,\eta) = [U_\mu(\eta,\tau), V(\eta,\tau)]. \tag{5.7}$$

But

$$\begin{aligned}
[U_\mu(\eta,\tau), V(\eta,\tau)]
&= \int d^4x' d^4x\, \delta(\eta x' - \tau)\delta(\eta x - \tau) \times (x'_\mu - \eta_\mu \eta x')[\mathcal{V}(x',\eta), \mathcal{V}(x,\eta)] \\
&= \int d^4x' d^4x\, \delta(\eta x' - \tau)(x'_\mu - \eta_\mu \eta x') \times \delta(\eta x - \eta x')[\mathcal{V}(x',\eta), \mathcal{V}(x,\eta)] \\
&= i\hbar \int d^4x' d^4x\, \delta(\eta x' - \tau)(x'_\mu - \eta_\mu \eta x') \times [-\partial'^\nu \delta^4(x'-x)\delta_\nu \mathcal{V}(x,\eta)] \\
&= i\hbar \int d^4x\, \delta(\eta x - \tau)\delta_\mu \mathcal{V}(x,\eta),
\end{aligned} \tag{5.8}$$

where the first term on the right-hand side of (5.4) makes no contribution here. Similarly,

$$\begin{aligned}
\delta_\mu U_\nu(\eta,\tau) = \int d^4x\{&\delta'(\eta x - \tau)(x_\mu - \eta_\mu \eta x) \times (x_\nu - \eta_\nu \eta x)\mathcal{V}(x,\eta) \\
&- \delta(\eta x - \tau)[(g_{\mu\nu} - \eta_\mu \eta_\nu)\eta x \mathcal{V}(x,\eta) + \eta_\nu(x_\mu - \eta_\mu \eta x)\mathcal{V}(x,\eta)] \\
&+ \delta(\eta x - \tau)(x_\nu - \eta_\nu \eta x)\delta_\mu \mathcal{V}(x,\eta)\},
\end{aligned} \tag{5.9}$$

so that

$$\delta_\mu U_\nu(\eta,\tau) + \eta_\nu U_\mu(\eta,\tau) - \delta_\nu U_\mu(\eta,\tau) - \eta_\mu U_\nu(\eta,\tau) = \int d^4x\, \delta(\eta x - \tau)[(x_\nu - \eta_\nu \eta x)\delta_\mu - (x_\mu - \eta_\mu \eta x)\delta_\nu]\mathcal{V}(x,\eta). \tag{5.10}$$

But

$$\begin{aligned}
[U_\mu(\eta,\tau), U_\nu(\eta,\tau)]
&= \int d^4x' d^4x (x'_\mu - \eta_\mu \eta x')(x_\nu - \eta_\nu \eta x) \times \delta(\eta x' - \tau)\delta(\eta x - \tau)[\mathcal{V}(x',\eta), \mathcal{V}(x,\eta)] \\
&= i\hbar \int d^4x' d^4x\, \delta(\eta x' - \tau)(x'_\mu - \eta_\mu \eta x')(x_\nu - \eta_\nu \eta x) \times [\delta_\rho \mathcal{V}(x',\eta)\partial^\rho - \delta_\rho \mathcal{V}(x,\eta)\partial'^\rho]\delta^4(x'-x) \\
&= i\hbar \int d^4x\, \delta(\eta x - \tau)[(x_\nu - \eta_\nu \eta x)\delta_\mu - (x_\mu - \eta_\mu \eta x)\delta_\nu]\mathcal{V}(x,\eta) \\
&= i\hbar[\delta_\mu U_\nu(\eta,\tau) + \eta_\nu U_\mu(\eta,\tau) - \delta_\nu U_\mu(\eta,\tau) - \eta_\mu U_\nu(\eta,\tau)],
\end{aligned} \tag{5.11}$$

as desired.

In the absence of derivative coupling, local Lagrangian field theory provides the trivial special case in which $\mathcal{V} = \mathcal{L}_{\text{int}}$ and

$$\delta(\eta x' - \eta x)[\mathcal{L}_{\text{int}}(x'), \mathcal{L}_{\text{int}}(x)] = 0. \tag{5.12}$$

* Supported in part by the National Science Foundation Research Grant No. GP-8867.

[1] The role of the concept of hyperplane dependence in relativistic quantum theory is presented by G. N. Fleming, J. Math. Phys. **7**, 1959 (1966).

[2] Communicated to the author in many private conversations. The hyperplane independence of the scattering amplitude is also relevant to the problem of presenting a manifestly covariant rendition of the relativistic direct-interaction studies of: H. Osborn, Phys. Rev. **176**, 1514, 1523 (1968); F. Coester, Helv. Phys. Acta **38**, 7 (1965); and R. Fong and J. Sucher, J. Math. Phys. **5**, 456 (1964). The same remark applies to eventual quantized versions of the instantaneous interaction studies of: R. N. Hill, J. Math. Phys. **8**, 201, 1756 (1967); D. G. Currie and T. F. Jordan, Phys. Rev. **167**, 1178 (1968), and Ref. 4 therein.

[3] G. N. Fleming, J. Math. Phys. **9**, 193 (1968).

[4] The symbol δ_μ denotes differentiation with respect to η^μ. Since η is a unit 4-vector, the derivative satisfies $(\delta_\mu \eta_\nu) = g_{\mu\nu} - \eta_\mu \eta_\nu$ and $[\delta_\mu, \delta_\nu] = \eta_\mu \delta_\nu - \eta_\nu \delta_\mu$.

[5] These equations can be read off directly from Ref. 3 by setting $H' = V$, $N'_\mu = U_\mu$, and $K'_\mu = J'_\mu = 0$ in that paper.

[6] Matrix elements constructed from these vectors for scattering amplitudes are closely related to the covariant M functions of H. P. Stapp, Phys. Rev. **125**, 2139 (1962).

[7] S. C. McDonald, J. Math. Phys. **10**, 1234 (1969).

[8] This is also true for the hyperplane dependent rewrites of conventional theories such as: I. Goldberg and E. Marx, Nuovo Cimento **57B**, 485 (1968); E. A. Remler, Phys. Rev. **166**, 1710 (1968); F. Goto, Progr. Theoret. Phys. (Kyoto) **37**, 571 (1967); and J. G. Valatin, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd., **26**, No. 13 (1951).

[9] Notice that, unlike the similar relations derived for the energy density operator by P. A. M. Dirac [Rev. Mod. Phys. **34**, 592 (1962); Can. J. Math. **3**, 1 (1951)] and J. Schwinger [Phys. Rev. **130**, 406 (1963)] on the basis of Poincaré invariance, one does not here encounter any implication of microcausality.

Ray Theory of Diffraction by Open-Ended Waveguides. I. Field in Waveguides

S. W. LEE
Hughes Aircraft Company, Ground Systems Group, Fullerton, California 92634

(Received 20 November 1969; Revised Manuscript Received 9 March 1970)

This paper presents an extension of Keller's ray method to problems involving two or more parallel plates. The approach used here is to solve a canonical problem (two staggered plates) in a rigorous manner, and then extract its dominant asymptotic terms so that the result admits ray interpretation. It is found that the coupling effect between two plates may be accounted for by introducing two multiplicative factors $G_+(k \cos \theta_0)$ and $G_+(k \cos \theta)$ to the conventional half-plane diffraction coefficients, where (θ_0, θ) are the directions of incoming and outgoing rays at the edge. The function $G_+(\alpha)$ is the "plus part" of the transformed Green's function $G(\alpha) = 1 - \exp[-2b(\alpha^2 - k^2)^{\frac{1}{2}}]$ in the Wiener–Hopf technique, and b is the guide width. A table for $G_+(k \cos \theta)$ is provided so that the calculation of the modified field on the rays can be accomplished even with a slide rule. An application of the present ray method to the problem of radiation from an unflanged open-ended parallel-plate waveguide recovers the *exact* solution obtainable by the Wiener–Hopf technique. In Paper I of this series of papers, we present the main theory and discuss the field in waveguide, while in Paper II the field in free space, particularly that on the shadow boundary, will be examined.

1. INTRODUCTION

The edge-diffracted rays introduced by Keller[1] in 1957 together with the classical rays of geometrical optics have been used successfully in attacking a large number of boundary value problems in free space. Recently, Yee, Felsen, and Keller[2] extended the ray method to calculate the reflection coefficient from an open-ended parallel-plate waveguide [Fig. 1(a)] and obtained good numerical results for examples where the guide width (2b) is as small as one-third of a wavelength. In view of the fact that only a few waveguide discontinuity problems can be analytically solved, the extention of the ray method by Yee–Felsen–Keller (YFK) is indeed very significant.

In the YFK method, the solution to the waveguide problem in Fig. 1(a) is, as usual, "built up" from that of a *single* half-plane (i.e., Sommerfeld solution). The interaction between the two plates is then partially [the terms of $O(kb)^{-1}$ and higher are neglected] accounted for by considering the multiple reflections along the shadow boundary at $z = 0$. In this connection, one may raise the possibility of deriving a set of diffraction coefficients which takes care of the *coupling along the shadow boundary* automatically (without resorting to multiple reflections) and, therefore, can be used conveniently in diffraction problems involving open-ended waveguides. The present paper is an attempt to do this and to suggest a simple method of modifying the ray amplitude derived from a single half-plane so that it includes the coupling effect between two plates.

The approach used here is first to solve a canonical problem in a rigorous manner, arrange its solution in a form which admits ray interpretation, and finally generalize the results so that they may be applied to a variety of other problems involving parallel plates or wedges. The canonical problem considered is the diffraction of a plane wave by two parallel plates staggered a distance l (Fig. 2). By following a standard procedure, the problem can be formulated in terms of two coupled Wiener–Hopf equations. For the special case $l = 0$, these two Wiener–Hopf equations can be decoupled by considering separately the symmetrical and asymmetrical field components [Fig. 1(b) and (c)], and the resultant equations are exactly solvable.[3] However, if $l > 0$, the problem becomes much more complicated. In the present paper, we accomplish the decoupling of the two Wiener–Hopf equations by asymptotically expanding certain integrals in inverse powers of (ka). As in the conventional ray method, we retain only the terms of $O(ka)^{-\frac{1}{2}}$, and the results so obtained can be conveniently identified with rays of cylindrical waves emanating from the two edges.

The present work consists of two parts. In Paper I, we give the main results and examine the field in the waveguide. Owing to the existence of various shadow boundaries, the field outside the waveguide is quite involved and will be presented in Paper II, in the near future. There is no difficulty in obtaining analytical expressions for the field in the neighborhood of the shadow boundaries. However, extracting the dominant term and giving appropriate ray interpretations cannot be accomplished without considerable effort.

From Secs. 2–5 in Paper I, the solution of the canonical problem is presented for incident TM waves. The main results are then summarized in Sec. 6, which also includes generalizations to other polarizations and configurations. In Sec. 7, we give a few

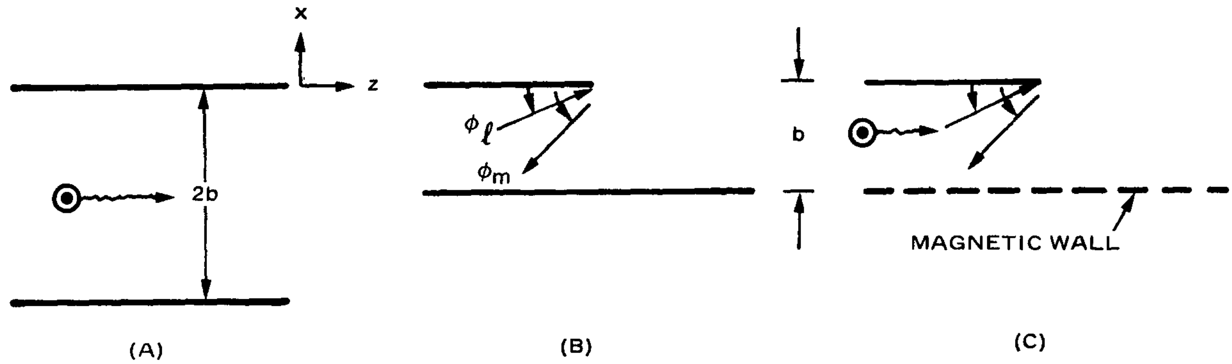


FIG. 1. Radiation from an open-ended parallel-plate waveguide: (a) geometry, (b) equivalent problem for an incident TM_{n0} or TE_{n0} with n even, (c) equivalent problem for an incident TM_{n0} or TE_{n0} with n odd.

examples to illustrate the applications of the ray method developed in previous sections to some typical problems. For those readers who are primarily interested in how to make use of the results in this paper, they may choose to bypass Secs. 2–5 and start with Sec. 6, which is self-contained.

2. FORMULATION

To develop a system of diffraction coefficients suitable for parallel-plate waveguide problems, one needs to consider all the four cases in Fig. 3 for different combinations of electric or magnetic walls, and polarizations. In Secs. 2–5, we consider only case (A), and give some detailed manipulations. Solutions to other cases can be found in a similar manner, and we present only their final results in Sec. 6.

The geometry of case (A) in Fig. 3 is enlarged in Fig. 2. The incident plane wave is given by

$$H_y^{(i)} = e^{-ik(\sin\theta_0 x + \cos\theta_0 y)}. \tag{2.1}$$

The angle θ_0 and all the other angles θ, θ', etc., to be introduced later, are defined to take values between π and $(-\pi)$. This is designed to avoid the ambiguity in the double-valued functions such as $\sin\frac{1}{2}\theta_0$.

The Wiener–Hopf formulation of the present problem varies slightly depending on whether $|\theta_0| < (\frac{1}{2}\pi)$ or $|\theta_0| > (\frac{1}{2}\pi)$. In applying the Wiener–Hopf technique, it is desirable to define the scattered field so that it attenuates as $|z| \to \infty$. For the definition used in (2.2), the scattered field in the shadow region does not satisfy this criterion if $|\theta_0| > (\frac{1}{2}\pi)$. In Secs. 2–5, we will further restrict ourselves to the case $|\theta_0| < (\frac{1}{2}\pi)$. Again, this restriction will be relaxed in the final results to be presented in Sec. 6. For $|\theta_0| < (\frac{1}{2}\pi)$, we define the scattered field H_y as

$$H_y(x, z) = H_y^{(\text{total})}(x, z) - H_y^{(i)}(x, z) \tag{2.2}$$

everywhere. The scattered field satisfies the wave equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\right)H_y(x, z) = 0, \tag{2.3}$$

where $k = k_1 + ik_2$, and $k_2 \to 0^+$ is the loss factor of the medium. It may be shown that $H_y = O[\exp(-k_2 z)]$

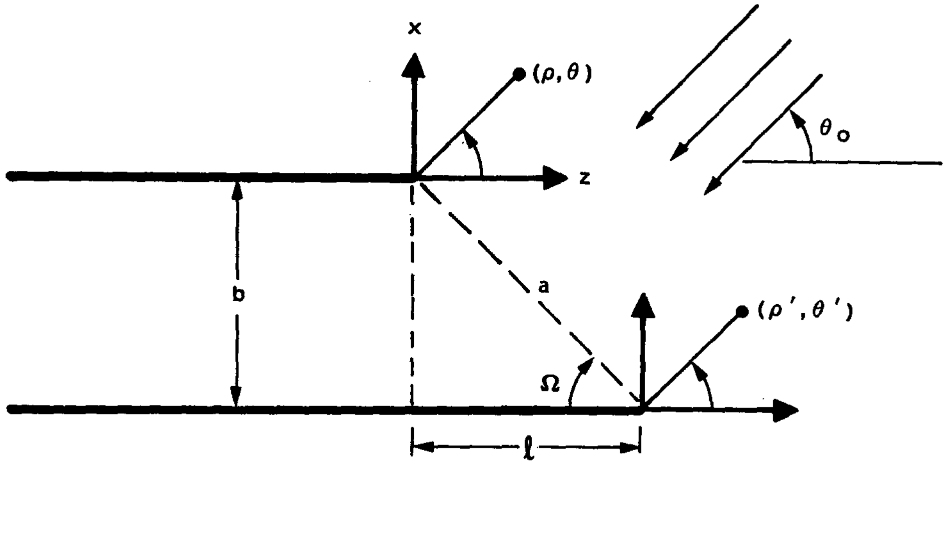


FIG. 2. Geometry of two electrically perfectly conducting parallel plates.

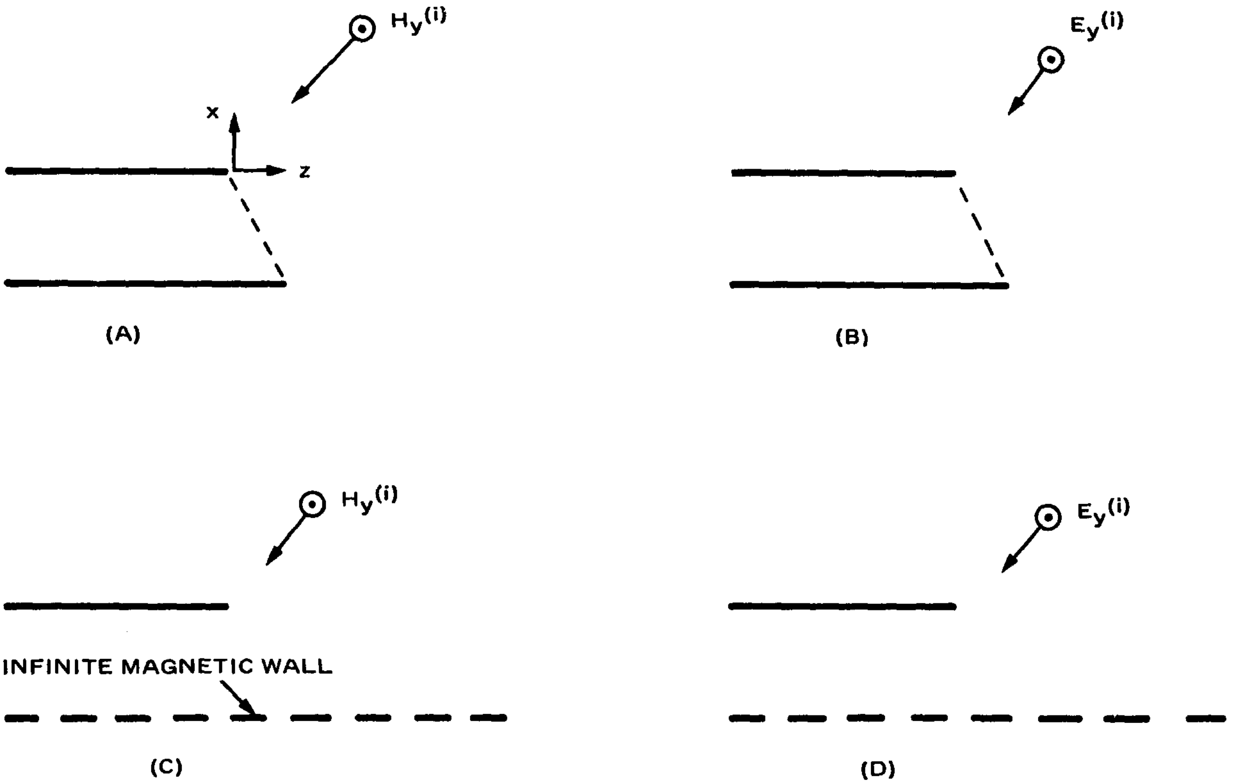


FIG. 3. Geometry of two parallel plates. The lower plate is either a semi-infinite electrically or an infinite magnetically perfectly conducting wall. The incident field is either TM or TE.

as $z \to +\infty$ and $H_y = O[\exp(k_2 \cos\theta_0 z)]$ as $z \to -\infty$. If we introduce the Fourier transform

$$\varphi(x, \alpha) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{\infty} H_y(x, z) e^{i\alpha z}\, dz, \quad \text{where} \quad \alpha = \sigma + i\tau, \tag{2.4}$$

then it follows from the asymptotic behavior of H_y that $\varphi(x, \alpha)$ is regular in the strip defined by

$$k_2 \cos\theta_0 > \tau > -k_2. \tag{2.5}$$

The transformed wave equation (2.3) reads

$$\left(\frac{\partial^2}{\partial x^2} - \gamma^2\right)\varphi(x, \alpha) = 0, \tag{2.6}$$

where $\gamma = (\alpha^2 - k^2)^{\frac{1}{2}}$. The branch cuts in the complex α plane are shown in Fig. 4 as wiggled lines, and the proper sheet of γ is chosen such that $\operatorname{Re}\gamma > 0$.

The solution of (2.6), satisfying the radiation condition, can be formally written as

$$\begin{aligned} \varphi(x, \alpha) &= A(\alpha)e^{-\gamma x}, & x > 0, \\ &= B(\alpha)e^{-\gamma x} + C(\alpha)e^{\gamma x}, & -b < x < 0, \\ &= D(\alpha)e^{\gamma x}, & x < -b, \end{aligned} \tag{2.7}$$

where A, B, C, and D are unknowns. Applying the boundary condition that $E_z(x, z)$ is continuous across $x = 0$ and $x = -b$, we obtain

$$A = B - C, \quad D = -Be^{2\gamma b} + C. \tag{2.8}$$

The next step is to match $H_y(x, z)$ [or $\varphi(x, \alpha)$] at $x = 0$ and $x = -b$.

First consider the matching at $x = 0$. Introduce the standard half-range transforms

$$\varphi_-(x, \alpha) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{0} H_y(x, z) e^{i\alpha z}\, dz, \tag{2.9a}$$

$$\varphi_+(x, \alpha) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{0}^{\infty} H_y(x, z) e^{i\alpha z}\, dz, \tag{2.9b}$$

where the subscripts "$-$," and "$+$" indicate that $\varphi_-(x, \alpha)$ is regular in the lower-half α plane defined by $\tau < k_2 \cos\theta_0$, and $\varphi_+(x, \alpha)$ is regular in the upper-half α plane defined by $\tau > -k_2$, respectively. In terms of these notations, the values of $\varphi(x, \alpha)$ at the two sides of $x = 0$ are

$$\varphi_-(0^+, \alpha) + \varphi_+(0^+, \alpha) = B(\alpha) - C(\alpha), \tag{2.10a}$$

$$\varphi_-(0^-, \alpha) + \varphi_+(0^-, \alpha) = B(\alpha) + C(\alpha), \tag{2.10b}$$

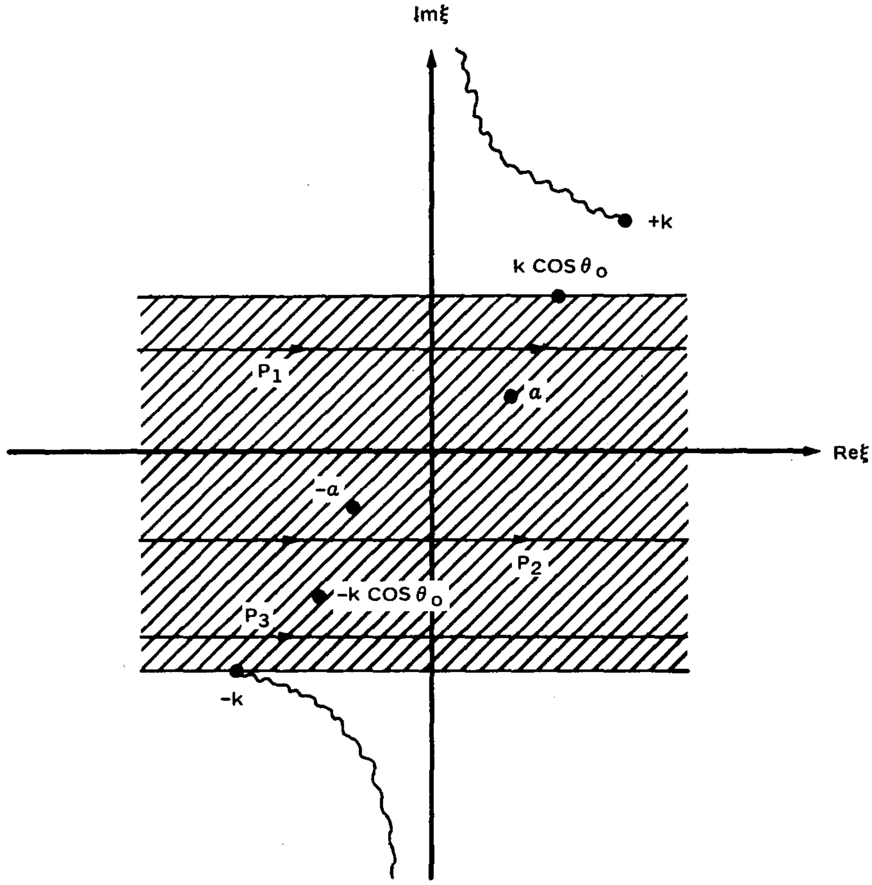


FIG. 4. Contours in complex ξ plane (for $|\theta_0| < \frac{1}{2}\pi$).

where we have made use of (2.8). Since $\varphi_+(0^+, \alpha) = \varphi_+(0^-, \alpha)$, the difference of (2.10a) and (2.10b) gives

$$C(\alpha) = -J_-(0, \alpha), \tag{2.11}$$

where $J_-(0, \alpha) = \frac{1}{2}[\varphi_-(0^+, \alpha) - \varphi_-(0^-, \alpha)]$ is proportional to the current on the plate at $x = 0$.

In order to match the field at $x = -b$, it is convenient to introduce the "shifted" half-range transforms

$$\tilde{\varphi}_-(x, \alpha) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{l} H_y(x, z) e^{i\alpha(z-l)}\, dz, \tag{2.12a}$$

$$\tilde{\varphi}_+(x, \alpha) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{l}^{\infty} H_y(x, z) e^{i\alpha(z-l)}\, dz. \tag{2.12b}$$

With these, we may derive in a similar manner

$$B(\alpha) = e^{i\alpha l} e^{-\gamma b} \tilde{J}_-(-b, \alpha), \tag{2.13}$$

where $\tilde{J}_-(-b, \alpha) = \frac{1}{2}[\tilde{\varphi}_-(-b^+, \alpha) - \tilde{\varphi}(-b^-, \alpha)]$ is proportional to the current at the plate at $x = -b$.

Finally, we apply the boundary condition at the two conducting plates. Written explicitly, the transformed $E_z(x, z)$ at $x = 0$ and $-b$ are

$$\varphi_-'(0, \alpha) + \varphi_+'(0, \alpha) = \gamma(-B + C), \tag{2.14a}$$

$$e^{i\alpha l}[\tilde{\varphi}_-'(-b, \alpha) + \tilde{\varphi}_+'(-b, \alpha)] = \gamma(-Be^{\gamma b} + Ce^{-\gamma b}), \tag{2.14b}$$

where the prime on, for example, $\varphi_-'(x, \alpha)$ means the derivative of $\varphi_-(x, \alpha)$ with respect to x. The condition that $E_z^{(\text{total})}(x, z)$ vanishes at $(x = 0, z < 0)$, and $(x = -b, z < l)$ implies

$$\varphi_-'(0, \alpha) = -\varphi'^{(i)}(0, \alpha) = \frac{k \sin \theta_0}{(2\pi)^{\frac{1}{2}}(\alpha - k \cos \theta_0)}, \tag{2.15a}$$

$$e^{i\alpha l}\tilde{\varphi}_-'(-b, \alpha) = -e^{i\alpha l}\tilde{\varphi}'^{(i)}(-b, \alpha) = \frac{k \sin \theta_0 \exp[-ika \cos(\theta_0 + \Omega)] e^{i\alpha l}}{\alpha - k \cos \theta_0}, \tag{2.15b}$$

where the angle Ω is shown in Fig. 2. Substituting (2.11), (2.13), and (2.15a) into (2.14a) gives

$$\frac{k \sin \theta_0}{(2\pi)^{\frac{1}{2}}(\alpha - k \cos \theta_0)} + \varphi_+'(0, \alpha) = -\gamma e^{i\alpha l} e^{-\gamma b} \tilde{J}_-(-b, \alpha) - \gamma J_-(0, \alpha),$$
$$-k_2 < \tau < k_2 \cos \theta_0. \tag{2.16}$$

Substituting (2.11), (2.13), and (2.15b) into (2.14b) gives

$$\frac{k\sin\theta_0 \exp[-ika\cos(\theta_0+\Omega)]e^{i\alpha l}}{(2\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}+e^{i\alpha l}\tilde{\varphi}'_+(-b,\alpha)$$
$$=-\gamma e^{i\alpha l}\tilde{J}_-(-b,\alpha)-\gamma e^{-\gamma b}J_-(0,\alpha),$$
$$-k_2<\tau<k_2\cos\theta_0. \quad (2.17)$$

Equations (2.16) and (2.17) are the desired Wiener–Hopf equations to be solved for the unknown currents $J_-(0,\alpha)$ and $\tilde{J}_-(-b,\alpha)$. Their solution is carried out in the next section.

3. SIMPLIFICATION OF WIENER-HOPF EQUATIONS

Let us concentrate first on the solution of (2.17). Rewrite it as

$$\frac{k\sin\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(2\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}\times\left(\frac{1}{(\alpha+k)^{\frac{1}{2}}}-\frac{1}{(k\cos\theta_0+k)^{\frac{1}{2}}}\right)+\frac{\tilde{\varphi}'_+(-b,\alpha)}{(\alpha+k)^{\frac{1}{2}}}+T_+(\alpha)$$
$$=-(\alpha-k)^{\frac{1}{2}}\tilde{J}_-(-b,\alpha)-\frac{k\sin\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(2\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)(k\cos\theta_0+k)^{\frac{1}{2}}}-[(\alpha+k)^{\frac{1}{2}}e^{-\gamma b}e^{-i\alpha l}J_-(0,\alpha)-T_+(\alpha)],$$
$$-k_2<\tau<k_2\cos\theta_0. \quad (3.1)$$

Here $T_+(\alpha)$ is the "+" part of

$$T(\alpha)=(\alpha+k)^{\frac{1}{2}}e^{-\gamma b}e^{-i\alpha l}J_-(0,\alpha) \quad (3.2)$$

and is given by[3]

$$T_+(x)=\frac{1}{2\pi i}\int_{P_2}\frac{(\xi-k)^{\frac{1}{2}}J_-(0,\xi)\exp[-(\xi^2-k^2)^{\frac{1}{2}}b]e^{-i\xi l}}{\xi-\alpha}\,d\xi. \quad (3.3)$$

The contour P_2 is indicated in Fig. 4. We will evaluate $T_+(\alpha)$ later. Returning to (3.1), we note that all the terms on the lhs are "+" functions, while all the terms on the rhs are "−" functions. With the help of the edge condition, it may be shown that both sides of (3.1) are equal to zero and the resultant equations are valid for all α. In particular, equating the rhs of (3.1) to zero gives

$$\tilde{J}_-(-b,\alpha)=\frac{T_+(\alpha)}{(\alpha-k)^{\frac{1}{2}}}-\left(\frac{k}{\pi}\right)^{\frac{1}{2}}\frac{\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(\alpha-k)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}-e^{-\gamma b}e^{-i\alpha l}J_-(0,\alpha). \quad (3.4)$$

This is the desired equation which expresses $\tilde{J}_-(-b,\alpha)$ in terms of $J_-(0,\alpha)$.

Next, substitution of (3.4) into the Wiener–Hopf equation (2.16) gives

$$\frac{k\sin\theta_0}{(2\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}+\varphi'_+(0,\alpha)$$
$$=\frac{k\sin\theta_0\exp[-ika\cos(\theta_0+\Omega)]e^{i\alpha l}e^{-\gamma b}(\alpha+k)^{\frac{1}{2}}}{2(k\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)\cos\frac{1}{2}\theta_0}-T_+(\alpha)e^{i\alpha l}e^{-\gamma b}(\alpha+k)^{\frac{1}{2}}-\gamma J_-(0,\alpha)G(\alpha),$$
$$-k_2<\tau<k_2\cos\theta_0, \quad (3.5)$$

where $G(\alpha)$ is given by

$$G(\alpha)=1-e^{-2\gamma b}. \quad (3.6a)$$

$G(\alpha)$ may be factored as

$$G(\alpha)=G_+(\alpha)G_-(\alpha)=G_+(\alpha)G_+(-\alpha), \quad (3.6b)$$

where

$$G_+(\alpha)=(2b)^{\frac{1}{2}}(k+\alpha)^{\frac{1}{2}}e^{-i(\frac{1}{4}\pi)}(\sin(kb)/kb)^{\frac{1}{2}}\times\exp\left[\frac{i\alpha b}{\pi}\left(1-C+\ln\frac{2\pi}{kb}+\tfrac{1}{2}i\pi\right)\right]\times\exp\left(\frac{ib\gamma}{\pi}\ln\frac{\alpha-\gamma}{k}\right)\prod_{n=1}^{\infty}\left(1+\frac{\alpha}{\gamma_n}\right)e^{i\alpha b/n\pi}. \quad (3.7)$$

In the above expression

$$\gamma_n=[(n\pi/b)^2-k^2]^{\frac{1}{2}}=(-i)[k^2-(n\pi/b)^2]^{\frac{1}{2}}$$
and $C=0.57721\cdots$.

Other forms of $G_+(\alpha)$ suitable for certain special situations are given in Appendix A. Dividing (3.5) by $(\alpha+k)^{\frac{1}{2}}G_+(\alpha)$ gives

$$\frac{k\sin\theta_0}{(2\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)(\alpha+k)^{\frac{1}{2}}G_+(\alpha)}+\frac{\varphi'_+(0,\alpha)}{(\alpha+k)^{\frac{1}{2}}G_+(\alpha)}$$
$$=\frac{k\sin\theta_0\exp[-ika\cos(\theta_0+\Omega)]e^{i\alpha l}e^{-\gamma b}}{2(k\pi)^{\frac{1}{2}}(\alpha-k\cos\theta_0)\cos\frac{1}{2}\theta_0G_+(\alpha)}-\frac{T_+(\alpha)e^{i\alpha l}e^{-\gamma b}}{G_+(\alpha)}-(\alpha-k)^{\frac{1}{2}}G_-(\alpha)J_-(0,\alpha),$$
$$-k_2<\tau<k_2\cos\theta_0. \quad (3.8)$$

Once again it may be shown the "+" and "−" parts of (3.8) are zero, respectively, and the resultant equations are valid for all α. In particular, equating the "−" part of (3.8) to zero gives

$$J_-(0,\alpha)=\frac{(-1)(k)^{\frac{1}{2}}\sin\frac{1}{2}\theta_0}{(\pi)^{\frac{1}{2}}G_+(k\cos\theta_0)}\times\frac{1}{(\alpha-k\cos\theta_0)(\alpha-k)^{\frac{1}{2}}G_-(\alpha)}+\frac{U_-(\alpha)-V_-(\alpha)}{(\alpha-k)^{\frac{1}{2}}G_-(\alpha)}. \quad (3.9)$$

Here $U_-(\alpha)$ is the "$-$" part of the first term on the rhs of (3.8) and is given by

$$U_-(\alpha) = \frac{-1}{2\pi i}\int_{P_1} \frac{k\sin\theta_0 \exp[-ika\cos(\theta_0+\Omega)]e^{i\xi l}\exp[-(\xi^2-k^2)^{\frac{1}{2}}b]}{2(k\pi)^{\frac{1}{2}}(\xi-k\cos\theta_0)\cos\frac{1}{2}\theta_0 G_+(\xi)(\xi-\alpha)}\,d\xi, \tag{3.10}$$

where P_1 is indicated in Fig. 4, and $V_-(\alpha)$ is the "$-$" part of the second term on the rhs of (3.8) and is given by

$$V_-(\alpha) = \frac{-1}{2\pi i}\int_{P_1} \frac{T_+(\xi)e^{i\xi l}\exp[-(\xi^2-k^2)^{\frac{1}{2}}b]}{G_+(\xi)(\xi-\alpha)}\,d\xi. \tag{3.11}$$

We also postpone the evaluation of $U_-(\alpha)$ and $V_-(\alpha)$ to the next section.

Summarizing the results obtained so far, we see that the formal solutions of $J_-(0,\alpha)$ and $\tilde{J}_-(-b,\alpha)$ are given in (3.9) and (3.4), respectively. Note that, since $V_-(\alpha)$ contains the unknown $J_-(0,\alpha)$ in $T_+(\alpha)$, we have not yet obtained the explicit solution of $J_-(0,\alpha)$. As will be shown later, the explicit result of $J_-(0,\alpha)$ [also of $\tilde{J}_-(-b,\alpha)$] is obtained in the present paper through the asymptotic evaluations of $T_+(\alpha)$, $U_-(\alpha)$, and $V_-(\alpha)$ for large $ka = k(b^2+l^2)^{\frac{1}{2}}$.

4. ASYMPTOTIC EVALUATION OF $U_-(\alpha)$, $V_-(\alpha)$, AND $T_+(\alpha)$

As will become clear later, the final solution of the present problem depends on the values of $U_-(\alpha)$, $V_-(\alpha)$, and $T_+(\alpha)$ for α in the range of $(-k,k)$. Therefore, it is convenient to write α as

$$\alpha = -k\cos w, \quad \text{where } |w| \le \pi. \tag{4.1}$$

Now first consider the evaluation of $U_-(\alpha = -k\cos w)$ in (3.10). A change of variable $\xi \to -\xi$ gives

$$U_-(-k\cos w) = \frac{(-1)(k)^{\frac{1}{2}}\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{2\pi i(\pi)^{\frac{1}{2}}} \times \int_{P_2}\frac{\exp[-(\xi^2-k^2)^{\frac{1}{2}}b]e^{-i\xi l}\,d\xi}{(\xi+k\cos\theta_0)(\xi-k\cos w)G_-(\xi)}. \tag{4.2}$$

In order to express $U_-(-k\cos w)$ in terms of Fresnel integrals at a later stage, we shift the contour P_2 in (4.2) across the pole at $\xi = -k\cos\theta_0$ to P_3 (Fig. 4). This gives

$$U_-(-k\cos w) = \frac{(-1)\sin\frac{1}{2}\theta_0\exp[ikb(\sin\theta_0+\sin|\theta_0|)]}{(k\pi)^{\frac{1}{2}}(\cos w+\cos\theta_0)G_+(k\cos\theta_0)} + \tilde{U}_-(-k\cos w), \tag{4.3a}$$

where

$$\tilde{U}_-(-k\cos w) = \frac{(-1)(k)^{\frac{1}{2}}\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{2\pi i(\pi)^{\frac{1}{2}}} \times \int_{P_3}\frac{\exp[-(\xi^2-k^2)^{\frac{1}{2}}b-i\xi l]}{(\xi+k\cos\theta_0)(\xi-k\cos w)G_-(\xi)}\,d\xi. \tag{4.3b}$$

The integral in (4.3b) can not be evaluated explicitly. In the present paper we evaluate $\tilde{U}_-(-k\cos w)$ asymptotically by dropping terms higher than $(ka)^{-\frac{1}{2}}$. As a preparatory step for using the saddle-point method, the poles at $\xi = -k\cos\theta_0$ and $k\cos w$ in (4.3b) should first be isolated. To this end, rewrite (4.3b) as

$$\begin{aligned}\tilde{U}_-(-k\cos w) = {}& \frac{(-1)(k)^{\frac{1}{2}}\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{2\pi i(\pi)^{\frac{1}{2}}}\\ &\times\Bigg[\int_{P_3}\left(\frac{(\xi-k)^{\frac{1}{2}}}{G_+(-\xi)}-\frac{(-i)(k+k\cos\theta_0)^{\frac{1}{2}}}{G_+(k\cos\theta_0)}\right)[(\xi+k\cos\theta_0)(\xi-k)^{\frac{1}{2}}]^{-1}e^{-(\xi^2-k^2)^{\frac{1}{2}}b-i\xi l}\,d\xi\\ &-\int_{P_3}\left(\frac{(\xi-k)^{\frac{1}{2}}}{G_+(-\xi)}-\frac{(-i)(k-k\cos w)^{\frac{1}{2}}}{G_+(-k\cos w)}\right)[(\xi-k\cos w)(\xi-k)^{\frac{1}{2}}]^{-1}e^{-(\xi^2-k^2)^{\frac{1}{2}}b-i\xi l}\,d\xi\\ &+\frac{(-i)(k+k\cos\theta_0)^{\frac{1}{2}}}{G_+(-k\cos\theta_0)}\int_{P_3}\frac{\exp[-(\xi^2-k^2)^{\frac{1}{2}}b-i\xi l]}{(\xi+k\cos\theta_0)(\xi-k)^{\frac{1}{2}}}\,d\xi\\ &-\frac{(-i)(k-k\cos w)^{\frac{1}{2}}}{G_+(-k\cos w)}\int_{P_3}\frac{\exp[-(\xi^2-k^2)^{\frac{1}{2}}b-i\xi l]}{(\xi-k\cos w)(\xi-k)^{\frac{1}{2}}}\,d\xi\Bigg].\end{aligned} \tag{4.4}$$

The integrands in the first two terms in (4.4) are smooth functions in the neighborhood of

$$\xi = -k\cos\theta_0$$

and $\xi = k\cos w$, respectively, and can be asymptotically evaluated by the saddle-point method in a standard manner. The third and fourth integrals may be identified with the Sommerfeld half-plane solutions.

Omitting the details, we give here only the final result:

$$\begin{aligned}
&\tilde{U}_{-}(-k\cos w)\\
&= \frac{(-1)2i\sin\frac{1}{2}\Omega\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(2\pi)^{\frac{1}{2}}(\cos w+\cos\theta_0)}\\
&\quad\times\frac{\exp[i(ka-\frac{1}{4}\pi)]}{(2\pi ka)^{\frac{1}{2}}}\Bigg[\frac{1}{k(\cos\theta_0-\cos\Omega)}\\
&\quad\times\left(\frac{(k+k\cos\Omega)^{\frac{1}{2}}}{G_+(k\cos\Omega)}-\frac{(k+k\cos\theta_0)^{\frac{1}{2}}}{G_+(k\cos\theta_0)}\right)\\
&\quad+\frac{1}{k(\cos w+\cos\Omega)}\\
&\quad\times\left(\frac{(k+k\cos\Omega)^{\frac{1}{2}}}{G_+(k\cos\Omega)}-\frac{(k-k\cos w)^{\frac{1}{2}}}{G_+(-k\cos w)}\right)\Bigg]\\
&\quad+\frac{\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(k\pi)^{\frac{1}{2}}(\cos w+\cos\theta_0)}\\
&\quad\times\left(\frac{Q[a,\Omega\mid\pi-|\theta_0|]}{G_+(k\cos\theta_0)}-\frac{\operatorname{sgn} w\,Q[a,\Omega\mid w]}{G_+(-k\cos w)}\right).
\end{aligned}\tag{4.5}$$

The notation used in (4.5) is explained as

$$\begin{aligned}
\operatorname{sgn} w &= +1, \quad \text{if } w>0,\\
&= -1, \quad \text{if } w<0,
\end{aligned}\tag{4.6}$$

and $Q[\rho,\theta\mid\theta_0]$ is the Sommerfeld half-plane solution given by

$$\begin{aligned}
&Q[\rho,\theta\mid\theta_0]\\
&= -e^{-ik\rho\cos(\theta-\theta_0)}\\
&\quad+\frac{e^{-\frac{1}{4}i\pi}}{(\pi)^{\frac{1}{2}}}\{e^{-ik\rho\cos(\theta+\theta_0)}F[+(2k\rho)^{\frac{1}{2}}\cos\tfrac{1}{2}(\theta+\theta_0)]\\
&\quad+e^{-ik\rho\cos(\theta-\theta_0)}F[-(2k\rho)^{\frac{1}{2}}\cos\tfrac{1}{2}(\theta-\theta_0)]\},
\end{aligned}\tag{4.7}$$

where $F(x)$ is the (complex) Fresnel integral

$$F(x)=\int_x^\infty e^{it^2}\,dt.\tag{4.8}$$

Physically, $Q[\rho,\theta\mid\theta_0]$ may be identified as the scattered (not total) field observed at (ρ,θ) by a half-plane (the half-plane is located at $0<\rho<\infty$ and $\theta=\pi$) due to an incident plane wave propagating in the direction θ_0 [see (2.1)]. For large $k\rho$, we may use the asymptotic formula for the Fresnel integral and obtain the leading terms of $Q[\rho,\theta\mid\theta_0]$. Dropping the terms higher than $(k\rho)^{-\frac{1}{2}}$, we obtain the result

$$\begin{aligned}
Q[\rho,\theta\mid\theta_0] &= \operatorname{sgn}(\theta\theta_0)e^{-ik\rho\cos(|\theta|+|\theta_0|)}\\
&\quad+Q^{(d)}[\rho,\theta\mid\theta_0], \quad \pi<|\theta|+|\theta_0|,\\
&= Q^{(d)}[\rho,\theta\mid\theta_0], \quad \text{elsewhere},
\end{aligned}\tag{4.9a}$$

$$Q^{(d)}[\rho,\theta\mid\theta_0]=\frac{e^{i(k\rho-\frac{1}{4}\pi)}}{(2\pi k\rho)^{\frac{1}{2}}}\frac{2i\sin\frac{1}{2}\theta_0\sin\frac{1}{2}\theta}{\cos\theta_0+\cos\theta}, \quad \text{if } |\theta|+|\theta_0|\neq\pi,\tag{4.9b}$$

where $Q^{(d)}$ is the well-known field on the rays diffracted at the edge of a single half-plane. For $|\theta|+|\theta_0|\approx\pi$, i.e., in the neighborhood of shadow boundaries, the value of $Q[\rho,\theta\mid\theta_0]$ has to be evaluated from the exact formula given in (4.7). In particular, for the scattered field exactly on the shadow boundaries, one has

$$\begin{aligned}
&Q[\rho,\theta\mid\theta_0]\\
&= \tfrac{1}{2}e^{ik\rho}-\frac{e^{-\frac{1}{4}i\pi}}{(\pi)^{\frac{1}{2}}}e^{ik\rho\cos2\theta_0}F[(2k\rho)^{\frac{1}{2}}\sin|\theta_0|],\\
&\qquad\text{if } \theta=\operatorname{sgn}\theta_0[\pi-|\theta_0|],\\
&= (-1)\left\{\tfrac{1}{2}e^{ik\rho}-\frac{e^{-\frac{1}{4}i\pi}}{(\pi)^{\frac{1}{2}}}e^{ik\rho\cos2\theta_0}F[(2k\rho)^{\frac{1}{2}}\sin|\theta_0|]\right\},\\
&\qquad\text{if } \theta=-\operatorname{sgn}\theta_0[\pi-|\theta_0|].
\end{aligned}\tag{4.10}$$

A simplified version of $\tilde{U}_{-}(\alpha)$ can be obtained by making use of (4.9) in (4.5). The result is

$$\begin{aligned}
&\tilde{U}_{-}(-k\cos w)\\
&= \frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\frac{\exp[-ika\cos(\theta_0+\Omega)]}{(\pi k)^{\frac{1}{2}}G_+(k\cos\Omega)}\\
&\quad\times\frac{-2i\sin\frac{1}{2}\Omega\cos\frac{1}{2}\Omega\sin\frac{1}{2}\theta_0}{(\cos w+\cos\Omega)(\cos\theta_0-\cos\Omega)}\\
&\quad+\frac{\exp[ikb(\sin\theta_0+\sin|\theta_0|)]}{(k\pi)^{\frac{1}{2}}G_+(k\cos\theta_0)}\\
&\quad\times\frac{\sin\frac{1}{2}\theta_0}{\cos\theta_0+\cos w}C(0<|\theta_0|<\Omega)\\
&\quad-\frac{\exp[-ika\cos(\theta_0+\Omega)-ika\cos(\Omega+|w|)]}{(k\pi)^{\frac{1}{2}}G_+(-k\cos w)}\\
&\quad\times\frac{\sin\frac{1}{2}\theta_0}{\cos\theta_0+\cos w}C(0<\pi-|w|<\Omega),
\end{aligned}\tag{4.11}$$

where $C(\theta_1<\theta<\theta_2)$ is the usual Heaviside function for an angular section and is defined by

$$\begin{aligned}
C(\theta_1<\theta<\theta_2) &= 1, \quad \text{if } \theta_1<\theta<\theta_2,\\
&= 0, \quad \text{elsewhere}.
\end{aligned}\tag{4.12}$$

In the derivation of (4.11), we have made use of the following approximations:

$$\theta'\approx\theta,\tag{4.13a}$$

$$\begin{aligned}
\rho' &\approx \rho, &&\text{in amplitude factors},\\
&\approx \rho-a\cos(\Omega+\theta), &&\text{in phase factors},
\end{aligned}\tag{4.13b}$$

which is valid for the case $k\rho\gg ka\gg1$. Apparently, the simplified result in (4.11) is not valid whenever the value of $\tilde{U}_{-}(-k\cos w)$ blows up (corresponding to various shadow boundaries). Such a situation occurs, for example, when $|w|\approx\pi-\Omega$, $|\theta_0|=\Omega$, etc. For those cases, the expression of $\tilde{U}_{-}(-k\cos w)$ given in (4.5) must be used.

In a similar manner, we can evaluate V_- ($\alpha = -k\cos w$) and T_+ ($\alpha = -k\cos w$) from (3.11) and (3.3) by the saddle-point method. The final results, if we drop the terms higher than $(ka)^{-\frac{1}{2}}$, are

$$V_-(-k\cos w) = \frac{i(2k)^{\frac{1}{2}}\sin\frac{1}{2}\Omega}{k(\cos w + \cos\Omega)}\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\left(\frac{T_+(k\cos\Omega)(k+k\cos\Omega)^{\frac{1}{2}}}{G_+(k\cos\Omega)} - \frac{T_+(-k\cos w)(k-k\cos w)^{\frac{1}{2}}}{G_+(-k\cos w)}\right) + \frac{(\operatorname{sgn} w)T_+(-k\cos w)}{G_+(-k\cos w)}Q[a,\Omega\,|\,w], \tag{4.14}$$

$$\begin{aligned}T_+(-k\cos w) = {} & \left(\frac{k}{\pi}\right)^{\frac{1}{2}}\frac{\sin\frac{1}{2}\theta_0 G(k\cos\theta_0)}{G_+(k\cos\theta_0)}C(\theta_0>0)\\
&\times\left\{\frac{-1}{k(\cos w+\cos\theta_0)}\left(\frac{(\operatorname{sgn}\theta_0)Q[a,\Omega\,|\,\theta_0]}{G_-(k\cos\theta_0)} - \frac{Q[a,\Omega\,|\,\pi-|w|]}{G_-(-k\cos w)}\right) - \frac{i\sin\frac{1}{2}\Omega}{k(\cos w+\cos\theta_0)}\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\right.\\
&\times\left[\left(\frac{2\cos\frac{1}{2}\Omega}{G_+(k\cos\Omega)} - \frac{2\sin\frac{1}{2}\theta_0}{G_-(k\cos\theta_0)}\right)\Big/(\cos\Omega+\cos\theta_0)\right.\\
&\left.\left.-\left(\frac{2k\cos\frac{1}{2}\Omega}{G_+(k\cos\Omega)} - \frac{(2k)^{\frac{1}{2}}(k+k\cos w)^{\frac{1}{2}}}{G_-(-k\cos w)}\right)\Big/k(\cos\Omega-\cos w)\right]\right\}\\
&-\frac{(2k)^{\frac{1}{2}}\sin\frac{1}{2}\Omega}{k(\cos\Omega-\cos w)}\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\left((-i)(k+k\cos\Omega)^{\frac{1}{2}}\frac{[\tilde U_-(-k\cos\Omega)-V_-(-k\cos\Omega)]}{G_+(k\cos\Omega)}\right.\\
&\left.+i(k+k\cos w)^{\frac{1}{2}}\frac{[\tilde U_-(-k\cos w)-V_-(-k\cos w)]}{G_+(k\cos w)}\right)\\
&+\left(\frac{\tilde U_-(-k\cos w)-V_-(-k\cos w)}{G_+(k\cos w)}\right)Q[a,\Omega\,|\,\pi-|w|].\end{aligned}\tag{4.15}$$

It should be noted that (4.14) and (4.15) are two coupled linear algebraic equations for $V_+(-k\cos w)$ and $T_+(-k\cos w)$, and may be solved without difficulty. [Note that $T_+(k\cos\Omega)$, etc., are not considered as unknowns. They may be obtained by setting $w = (\pi-\Omega)$ into the solution of $T_+(-k\cos w)$ once the latter is known.] However, the result is too lengthy to be useful. In the following analysis, we will compute $V_-(-k\cos w)$ from (4.14) by using

$$\tilde T_+(-k\cos w) = T_+(-k\cos w)\big|_{V_-(-k\cos w)\to 0} \tag{4.16}$$

instead of $T_+(-k\cos w)$ itself. If this approximation and the formula (4.9) are used in (4.14) and (4.15), the simplified versions of $V_-(-k\cos w)$ and $T_+(-k\cos w)$ may be obtained. The results are

$$\begin{aligned}V_-(-k\cos w) &= \frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\frac{i\sin\Omega\tilde T_+(k\cos\Omega)}{G_+(k\cos\Omega)[\cos w+\cos\Omega]}\\
&\quad+\frac{\tilde T_+(-k\cos w)\exp[-ika\cos(\Omega+w)]}{G_+(-k\cos w)}\\
&\quad\times C(0<\pi-|w|<\Omega),\end{aligned}\tag{4.17}$$

$$\begin{aligned}T_+(-k\cos w) &= \frac{(-1)\sin\frac{1}{2}\theta_0 G_-(k\cos\theta_0)}{(k\pi)^{\frac{1}{2}}}C(\theta_0>0)\\
&\quad\times\left[\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\frac{2i\sin(-\frac{1}{2}\Omega)\cos\frac{1}{2}\Omega}{G_+(k\cos\Omega)}\right.\\
&\quad\times\frac{1}{(\cos\Omega+\cos\theta_0)(\cos w-\cos\Omega)}\\
&\quad+\frac{1}{\cos\theta_0+\cos w}\\
&\quad\times\left(\frac{\exp[-ika\cos(\theta_0+\Omega)]}{G_-(k\cos\theta_0)}C(0<\pi-\theta_0<\Omega)\right.\\
&\quad\left.\left.-\frac{\exp[ika\cos(\Omega-|w|)]}{G_-(-k\cos w)}C(0<|w|<\Omega)\right)\right]\\
&\quad+\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\frac{i\sin\Omega}{\cos\Omega-\cos w}\\
&\quad\times\left(\frac{\tilde U_-(-k\cos\Omega)-V_-(-k\cos\Omega)}{G_+(k\cos\Omega)}\right)\\
&\quad+\left(\frac{\tilde U_-(-k\cos w)-V_-(-k\cos w)}{G_-(-k\cos w)}\right)\\
&\quad\times\exp[ika\cos(\Omega-|w|)]C(0<|w|<\Omega).\end{aligned}\tag{4.18}$$

Once again, the simplified results in (4.17) and (4.18) are not valid whenever the values of $V_-(-k\cos w)$ and $T_+(-k\cos w)$ blow up. In those situations, we have to use the expressions in (4.14) and (4.15).

5. FIELD IN WAVEGUIDE

In the previous section, we have obtained the explicit asymptotic results for $U_-(\alpha)$, $V_-(\alpha)$, and $T_+(\alpha)$.

Then (3.4) and (3.9) give the solutions of the induced currents on the two plates. Substituting (3.4) and (3.9) into (2.8), (2.11), and (2.13), we may determine the coefficients $A(\alpha)$, $B(\alpha)$, $C(\alpha)$, and $D(\alpha)$ and the complete field solution. The results are given as follows:

$$\varphi(x,\alpha) = \left[\left(\frac{k}{\pi}\right)^{\frac{1}{2}} \frac{(-1)\sin\frac{1}{2}\theta_0 G_+(\alpha)G_-(k\cos\theta_0)}{(\alpha-k)^{\frac{1}{2}}(\alpha-k\cos\theta_0)} C\,(\theta_0>0) + \frac{[\tilde{U}_-(\alpha)-V_-(\alpha)]G_+(\alpha)}{(\alpha-k)^{\frac{1}{2}}} + \frac{T_+(\alpha)e^{i\alpha l}e^{-\gamma b}}{(\alpha-k)^{\frac{1}{2}}} - \left(\frac{k}{\pi}\right)^{\frac{1}{2}} \times \frac{\sin\frac{1}{2}\theta_0 \exp[-ika\cos(\Omega+\theta_0)]e^{i\alpha l}e^{-\gamma b}}{(\alpha-k)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}\right]e^{-\gamma x}, \quad \text{for } x>b, \tag{5.1}$$

$$\varphi(x,\alpha) = \left[\frac{T_+(\alpha)}{(\alpha-k)^{\frac{1}{2}}} - \left(\frac{k}{\pi}\right)^{\frac{1}{2}} \times \frac{\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(\alpha-k)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}\right] \times e^{i\alpha l}e^{-\gamma(x+b)} - 2J_-(0,\alpha)e^{-\gamma b}\cosh\gamma(x+b), \quad \text{for } 0>x>(-b), \tag{5.2}$$

$$\varphi(x,\alpha) = \left[\left(\frac{k}{\pi}\right)^{\frac{1}{2}} \frac{\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(\alpha-k)^{\frac{1}{2}}(\alpha-k\cos\theta_0)} - \frac{T_+(\alpha)}{(\alpha-k)^{\frac{1}{2}}}\right]e^{i\alpha l}e^{\gamma(x+b)}, \quad \text{for } x<(-b). \tag{5.3}$$

To obtain the explicit solution in the space domain, i.e., $H_y(x,z)$, the inverse Fourier transform of $\varphi(x,\alpha)$ needs to be taken. For the field in various shadow boundaries outside the waveguide, the asymptotic evaluation of the inverse transform in order to obtain simple expressions suitable for ray interpretation is quite involved and will be dealt with in Paper II. At the present, we will consider only the field in the waveguide which, for $0>x>-b$, is given by

$$H_y(x,z) = \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty+i\tau}^{\infty+i\tau}\left[\frac{T_+(\alpha)}{(\alpha-k)^{\frac{1}{2}}} - \left(\frac{k}{\pi}\right)^{\frac{1}{2}} \times \frac{\sin\frac{1}{2}\theta_0\exp[-ika\cos(\theta_0+\Omega)]}{(\alpha-k)^{\frac{1}{2}}(\alpha-k\cos\theta_0)}\right] \times e^{-\gamma(x+b)-i\alpha(z-l)}\,d\alpha - \frac{1}{(2\pi)^{\frac{1}{2}}}\int_{-\infty+i\tau}^{\infty+i\tau} 2J_-(0,\alpha)e^{-\gamma b}\cosh\gamma(x+b)e^{-i\alpha z}\,d\alpha, \quad -k_2<\tau<k_2\cos\theta_0. \tag{5.4}$$

For $z<0$, we may deform the integral path to the upper-half α plane to enclose the following singularities: (i) the branch cut at $\alpha=k$; (ii) the pole at $\alpha=k\cos\theta_0$; and (iii) the infinite poles at $\alpha=k$ and $\alpha=\{i\gamma_m\}$ in the factor $[(\alpha-k)^{\frac{1}{2}}G_-(\alpha)]$. {Recalling (3.7), the factor $[(\alpha-k)^{\frac{1}{2}}G_-(\alpha)]$ has a simple zero, not a branch point, at $\alpha=k$.} It may be shown that, for the integral along the branch cut starting at $\alpha=k$, the contributions from the first and second integrals in (5.4) cancel each other and therefore the result is zero, as may be expected from physical intuition. Thus, the field in the waveguide region is all contributed from discrete poles in the transform domain, which corresponds to normal modes of the waveguide in the spatial domain.

The evaluation of (5.4) due to the pole contribution is quite straightforward. The result is

$$H_y(x,z) = H_y^{(\mathrm{inc})} + H_y^{(\mathrm{mode})}. \tag{5.5}$$

The first term is due to the pole at $\alpha=k\cos\theta_0$ and is given by

$$H_y^{(\mathrm{inc})} = -e^{-ik\rho\cos(\theta-\theta_0)}, \tag{5.6}$$

which cancels exactly the incident field as expected. The second term in (5.5) is due to the poles at $\alpha=k$ and $\alpha=\{i\gamma_m\}$, and may be expressed as

$$H_y^{(\mathrm{mode})} = \sum_{m=0}^{\infty} 2c_m\cos\frac{m\pi}{b}x e^{\gamma_m z}, \tag{5.7}$$

where $(2c_m)$ is the mode coefficient of the TM_{m0} mode and is given by

$$c_m = -i(2\pi)^{\frac{1}{2}}\,\mathrm{Res}\,J_-(0,i\gamma_m) = -i(2\pi)^{\frac{1}{2}}(i\gamma_m+k)^{\frac{1}{2}}G_+(i\gamma_m)\left[\frac{d}{d\alpha}\gamma G(\alpha)\right]^{-1}_{\alpha=i\gamma_m} \times \left(\frac{(-1)(k)^{\frac{1}{2}}\sin\frac{1}{2}\theta_0 G_-(k\cos\theta_0)}{(\pi)^{\frac{1}{2}}(i\gamma_m-k\cos\theta_0)}C\,(\theta_0>0) + \tilde{U}_-(i\gamma_m) - V_-(i\gamma_m)\right). \tag{5.8}$$

It should be remarked that even though all the analysis so far was based on the fact that $|\theta_0|<\frac{1}{2}\pi$, the result in (5.8) as a matter of fact is valid for all $|\theta_0|<\pi$. Usually, we are interested in those $\{c_m\}$ corresponding to propagating modes. For propagating modes, (5.8) can be simplified, and admits ray interpretations. This will be detailed below.

First, introduce φ_m such that

$$\sin\varphi_m = \left(\frac{m\pi}{b}\right)\Big/k, \quad 0\le\varphi_m<\tfrac{1}{2}\pi. \tag{5.9}$$

The angle φ_m may be identified with the direction of propagation of the two superimposed plane waves associated with TM_{m0} mode. Using the simplified version of $\tilde{U}(-k\cos w)$ and $V_-(-k\cos w)$ (with

$w = \pi - \varphi_m$) given in (4.11) and (4.17), we can simplify the expression for c_m in (7.6) to yield the following results:

$$c_m = \sum_{n=1}^{4} c_m^{(n)}. \tag{5.10}$$

The first term is

$$\begin{aligned} c_m^{(1)} = \Bigg(& \frac{2i \sin \frac{1}{2}\theta_0 \sin \frac{1}{2}(\varphi_m - \pi)}{\cos \theta_0 + \cos (\varphi_m - \pi)} \\ & \times \left(\frac{G_-[k \cos (\varphi_m - \pi)]}{G_+(k \cos \theta_0)} \right. \\ & \times [C(-\Omega < \theta_0) - e^{i2kb \sin \theta_0} C(\Omega < \theta_0)] \Bigg) \\ & \times \left(\left\{ \frac{d}{d\alpha} [\gamma G(\alpha)] \right\}^{-1} \right)_{\alpha = k \cos \varphi_m}. \end{aligned} \tag{5.11}$$

Some interpretations of (5.11) are in order:

(i) The factor in the first term is the ray amplitude due to the diffraction at the upper edge [Fig. 5(a)], as if the lower plate had not existed. [Compare it with (4.9b), and note that the factor $\exp i(k\rho - \frac{1}{4}\pi)/(2\pi k\rho)^{\frac{1}{2}}$ is regarded as the spreading factor of the cylindrical wave.] The scattered ray is in the direction of the plane wave associated with TM_{m0} mode.

(ii) The second factor is the modification introduced by the lower plate. For $\theta_0 < -\Omega$, the factor is zero, as the ray is blocked by the lower plate. For $|\theta_0| < \Omega$, the factor becomes

$$\frac{G_-[k \cos (\varphi_m - \pi)]}{G_+(k \cos \theta_0)}, \tag{5.12}$$

which may be interpreted as the contribution of the coupling along the shadow boundary at $z = 0$ (Fig. 2). As $b \to \infty$, while retaining a small loss, $G(\alpha)$ and thus $G_+(\alpha)$ and $G_-(\alpha)$ approach one. Hence, the factor in (5.12) becomes unity as expected. In the region $\Omega < \theta_0 < \frac{1}{2}\pi$, this factor becomes

$$\frac{G_-[k \cos (\varphi_m - \pi)]}{G_+(k \cos \theta_0)} (1 - e^{i2kb \sin \theta_0}). \tag{5.13}$$

The second term in (5.13) is due to the specular reflection at the lower plate before the ray strikes at the upper edge. Finally, in the region $\frac{1}{2}\pi < \theta_0 < \pi$, this becomes

$$G_-(k \cos \theta_0) G_-(k \cos (\varphi_m - \pi)), \tag{5.14}$$

which is again due to the coupling between the two plates. [In arriving at (5.14), we made use of the factor $G_-(\alpha) = G(\alpha)/G_+(\alpha)$.]

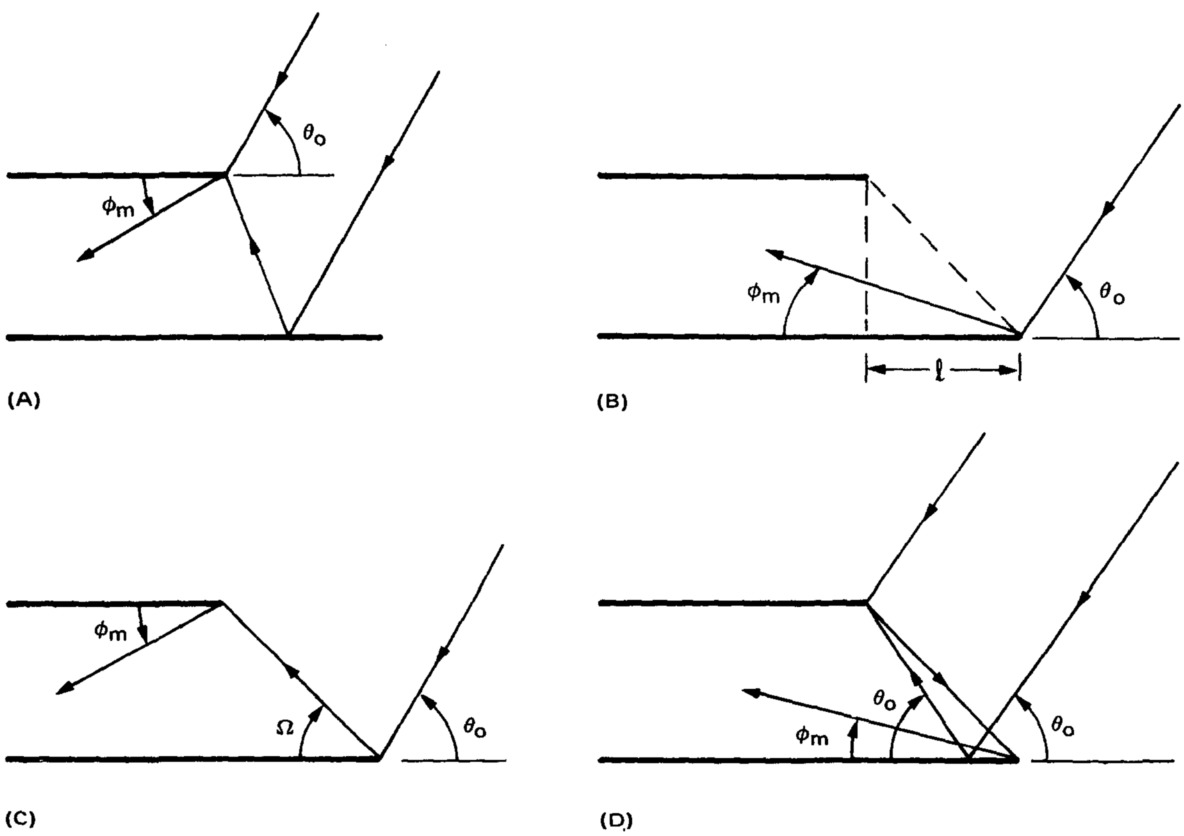


FIG. 5. Ray contributions to the modal coefficients as given in (5.10).

(iii) The third bracket in (5.11) may be interpreted as the *conversion factor from ray to mode*. Explicitly, it is given by

$$\left(\left\{\frac{d}{d\alpha}[\gamma G(\alpha)]\right\}^{-1}\right)_{\alpha=k\cos\varphi_m} = (2\epsilon_m kb\cos\varphi_m)^{-1}, \quad (5.15)$$

where $\epsilon_m = 2$, if $m = 0$, and $\epsilon_m = 1$, if $m \neq 0$. This factor was previously derived by Yee, Felsen, and Keller[2] through image consideration and by Deschamps[4] from the energy point of view.

The second term in (5.10) is given by

$$c_m^{(2)} = \left(e^{-ika\cos(\theta_0+\Omega)}\frac{2i\sin\frac{1}{2}\theta_0\sin\frac{1}{2}(\pi-\varphi_m)}{\cos\theta_0+\cos(\pi-\varphi_m)}\right) \times (e^{ika\cos(\Omega-\varphi_m)})(2\epsilon_m kb\cos\varphi_m)^{-1} \times C[-\pi<\theta_0<(\pi-\Omega)]C(0<\varphi_m<\Omega). \quad (5.16)$$

This is due to the ray diffracted at the lower edge [Fig. 5(b)]. Since there is no shadow boundary formed due to the presence of the upper plate, the ray amplitude is not modified. This particular ray can get into the waveguide only if $0 < \varphi_m < \Omega$, which explains the factor $C(0 < \varphi_m < \Omega)$ in (5.16). The additional phase factor

$$e^{ika\cos(\Omega-\varphi_m)} = (-1)^m e^{i(k\cos\varphi_m)l} \quad (5.17)$$

may be associated with the additional path length from the lower edge to the waveguide aperture. From $(x = -b, z = l)$ to $(x = -b, z = 0)$, the distance is l, and the propagation constant is $k\cos\varphi_m$. This results in the phase delay $\exp(ikl\cos\varphi_m)$. To account for the factor $(-1)^m$ in (5.17), we rewrite (5.7) as

$$H_y^{(\mathrm{mode})} = \sum_{m=0}^{\infty} c_m(e^{i(m\pi/b)x} + e^{-i(m\pi/b)x})e^{-i[k-(m\pi/b)^2]^{\frac{1}{2}}z}. \quad (5.18)$$

Observe that, at $(x = -b, z = 0)$, the ray traveling in the direction $(\pi - \varphi_m)$ has an amplitude $(-1)^m c_m$. Hence, the $(-1)^m$ in (5.17) is a normalization factor.

The third and fourth terms in (5.10) are given by

$$c_m^{(3)} = \left(e^{-ika\cos(\theta_0+\Omega)}\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\frac{2i\sin\frac{1}{2}\theta_0\sin\frac{1}{2}(\pi-\Omega)}{\cos\theta_0+\cos(\pi-\Omega)}\right) \times \left(\frac{2i\sin(-\frac{1}{2}\Omega)\sin\frac{1}{2}(\varphi_m-\pi)}{\cos\Omega+\cos(\varphi_m-\pi)} \times \frac{G_-[k\cos(\varphi_m-\pi)]}{G_+(k\cos\Omega)}\right)(2\epsilon_m kb\cos\varphi_m)^{-1} \times C(-\pi<\theta_0<\pi-\Omega), \quad (5.19)$$

$$c_m^{(4)} = \left\{\frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}}\frac{2i\sin\frac{1}{2}\theta_0\sin(-\frac{1}{2}\Omega)}{\cos\theta_0+\cos\Omega} \times \frac{1}{G_+(k\cos\theta_0)G_+(k\cos\Omega)} \times [C(-\Omega<\theta_0) - e^{i2kb\sin\theta_0}(\Omega<\theta_0)]\right\} \times \left(\frac{2i\sin\frac{1}{2}(\pi-\Omega)\sin\frac{1}{2}(\varphi_m-\pi)}{\cos(\pi-\Omega)+\cos(\varphi_m-\pi)} \times \frac{(-1)^m e^{ikl\cos\varphi_m}}{2\epsilon_m kb\cos\varphi_m}C(0<\varphi_m<\Omega)\right), \quad (5.20)$$

which are the contributions from the rays bouncing between the edges once before emanating from their respective edges [Fig. 5(c) and (d)].

Before concluding this section, let us consider a special case $|\theta_0| = \varphi_q$ for a particular q. This occurs when the incident wave is in the same direction as one of the plane waves associated with TM_{q0} mode in the waveguide. Under this condition, $c_q^{(1)}$ and $c_q^{(2)}$ blow up individually. This is because of the fact that we have used the simplified versions of $\tilde{U}_-(-k\cos w)$ and $V_-(-k\cos w)$ in (5.8), which are not valid whenever their values approach infinity. [Mathematically, this corresponds to *double* pole at $\xi = -k\cos\varphi_q$ for $\tilde{U}_-(k\cos\varphi_q)$ as given in (4.3b).] Instead, in computing (5.8), we should use $\bar{U}_-(-k\cos w)$ in (4.5) and $V_-(-k\cos w)$ in (4.14). For this particular case, however, we can obtain an equivalent result by simply combining $c_q^{(1)}$ and $c_q^{(2)}$ as a *single* term. For $\Omega < \theta_0 < \frac{1}{2}\pi$, we have

$$c_q^{(1)} + c_q^{(2)} = [2i\sin\tfrac{1}{2}\varphi_q\sin\tfrac{1}{2}(\varphi_q-\pi)]\left(\frac{1}{2\epsilon_q kb\cos\varphi_q}\right)\left(\frac{1-\exp(i2kb\sin\theta_0)}{\cos\theta_0-\cos\varphi_q}\right)_{\theta_0\to\varphi_q} = \frac{1}{\epsilon_q}, \quad \text{if} \quad \theta_0 = \varphi_q. \quad (5.21)$$

For $|\theta_0| < \Omega$, we have

$$c_q^{(1)} + c_q^{(2)} = \frac{\pm\sin\varphi_q}{i\epsilon_q 2kb\cos\varphi_q} \times \left(\frac{[G_+(k\cos\varphi_m)/G_+(k\cos\theta_0)] - \exp[+ika\cos(\Omega-\varphi_m) - ika\cos(\theta_0+\Omega)]}{\cos\theta_0-\cos\varphi_q}\right)_{\theta_0\to(\pm\varphi_q)}$$
$$= (i\epsilon_q 2kb\cos\varphi_q)^{-1}\left(\frac{\pm k\sin\varphi_q G_+'(k\cos\varphi_q)}{G_+(k\cos\varphi_q)} + ika\sin(\Omega\pm\varphi_q)\exp[ikb(\sin\varphi_m)(1+\operatorname{sgn}\theta_0)]\right), \quad \text{if} \quad \theta_0 = \pm\varphi_q, \quad (5.22)$$

where $G'_+(\alpha)$ is the derivative of $G_+(\alpha)$ with respect to its argument. Making use of the relation $G(\alpha) = G_+(\alpha)G_+(-\alpha)$, we easily show that

$$G'_+(\alpha) = \frac{(2\alpha b/\gamma)e^{-2\gamma b}}{G_+(-\alpha) - G_+(\alpha)}. \quad (5.23)$$

Thus, $G'_+(\alpha)$ can be expressed in terms of $G_+(\alpha)$ [a short table for $G_+(\alpha)$ will be given later].

6. SUMMARY OF MAIN RESULTS AND GENERALIZATION

A study of the results in the previous section, as well as the field outside the waveguide (which is to be presented in Paper II), we may state the following two main conclusions relating to the diffraction of an incident TM plane wave given in (2.1) by the parallel-plate structure shown in Fig. 2:

Modified Ray Amplitude. For the special case $b \to \infty$, the field on the ray diffracted at the edge of the *upper plate* is the well-known expression

$$H_y = \left(\frac{2i \sin \frac{1}{2}\theta_0 \sin \frac{1}{2}\theta}{\cos \theta_0 + \cos \theta}\right)\frac{e^{i(k\rho - \frac{1}{4}\pi)}}{(2\pi k\rho)^{\frac{1}{2}}}, \quad \text{if } b \to \infty, \quad (6.1)$$

where θ_0 is the direction of the incident wave plane, and (ρ, θ) is the observation point. Both θ_0 and θ are counted such that $|\theta_0|, |\theta| \le \pi$. For finite b, this field should be modified due to the coupling between the two plates with the results

$$H_y = \left(\frac{2i[f(\theta_0) \sin \frac{1}{2}\theta_0][g(\theta) \sin \frac{1}{2}\theta]}{\cos \theta_0 + \cos \theta}\right)\frac{e^{i(k\rho - \frac{1}{4}\pi)}}{(2\pi k\rho)^{\frac{1}{2}}}, \quad \text{for upper edge.} \quad (6.2)$$

The factor $f(\theta_0)$ is the modification associated with the incoming ray

$$f(\theta_0) = G_-(k \cos \theta_0), \quad \tfrac{1}{2}\pi < \theta_0 < \pi,$$
$$= \frac{1}{G_+(k \cos \theta_0)}, \quad -\Omega < \theta_0 < \tfrac{1}{2}\pi, \quad (6.3)$$

where $G_+(\alpha) = G_-(-\alpha)$ is given in (3.7). The factor $g(\theta)$ in (6.2) is the modification associated with the outgoing ray

$$g(\theta) = G_-(k \cos \theta), \quad |\theta| > \Omega \quad \text{and} \quad \theta \neq \Omega,$$
$$= \frac{1}{G_+(k \cos \theta)}, \quad |\theta| < \Omega \quad \text{and} \quad \theta \neq \Omega. \quad (6.4)$$

{Alternatively, the function $g(\theta)$ may be made identical to $f(\theta_0)$ if we do not count the ray which is specularly reflected from the lower plate before striking at the upper edge [cf. Eq. (5.13)].}

For $\theta \approx \Omega$, the result is quite involved and will be presented in Paper II. (In computing field in the waveguide, the case $\theta \approx \Omega$ does not arise.) The modification for the field on the diffracted ray introduced in (6.2) may be regarded as the consequence of multiple reflections and diffractions along the shadow boundary at $z = 0$. Thus, for diffraction at the *lower edge*, no such shadow boundary exists, and the ray amplitude is computed *as if the upper plate were absent.*

Ray-to-Mode Conversion. Let us express the field in the waveguide as

$$H_y(x, z) = \sum_m 2c_m \cos \frac{m\pi}{a} x e^{\pm\gamma_m z}. \quad (6.5)$$

For propagating modes, the contribution to c_m due to each ray is equal to the amplitude [the first factor in (6.2)] multiplied by the conversion factor

$$\left(\left\{\frac{d}{d\alpha}[\gamma G(\alpha)]\right\}^{-1}\right)_{\alpha = k \cos \varphi_m} = \frac{1}{2\epsilon_m kb \cos \varphi_m}, \quad (6.6)$$

where $\cos \varphi_m = [1 - (m\pi/kb)]^{\frac{1}{2}}$, $\epsilon_m = 2$ if $m = 0$, and $\epsilon_m = 1$ if $m \neq 0$. In calculating rays from the lower edge, there is an additional multiplying factor

$$(-1)^m e^{ikl \cos \varphi_m}. \quad (6.7)$$

The exponential factor is due to the phase delay from $(x = -b, z = l)$ to $(x = -b, z = 0)$. The presence of $(-1)^m$ follows from the fact that the ray traveling in the direction $(\pi - \varphi_m)$ as given in (6.5) has an amplitude $(-1)^m c_m$ at $(x = -b, z = 0)$.

With the above rules, it is now a simple matter to write down the expressions for some diffraction problems of parallel-plate waveguides. Several illustrating examples will be given in the next section. For numerical computations, we need the values of $G_+(kx)$ for $|x| < 1$ and a given kb. The numerical data presented in Table I is prepared for this purpose. For clarity, we plot the value of $G_+(kx)$ for certain kb in Fig. 6. Note that $|G_+(kx)|$ is far from being unity for negative x but approaches to unity as $x \to 1$. [The value of $G_+(kx)$ for negative x will be used for computing the field outside the waveguide.] With the help of Table I, the calculation of the various rays to obtain the solution of a diffraction problem can be accomplished even with a slide rule!

So far we have discussed only the TM case. A similar analysis has been carried out for the TE case. We will omit the details and simply summarize the final results as below [Fig. 3(b)]:

Modified Ray Amplitude. For finite b, the diffraction ray at the *upper edge* is given by

$$E_y = \left(\frac{-2i[f(\theta_0) \cos \frac{1}{2}\theta_0][g(\theta) \cos \frac{1}{2}\theta]}{\cos \theta_0 + \cos \theta}\right)\frac{e^{i(k\rho - \frac{1}{4}\pi)}}{(2\pi k\rho)^{\frac{1}{2}}}, \quad \text{for upper edge,} \quad (6.8)$$

TABLE I. Value of $G_+(kx)$.

Values of $G_+(kx)$ for $kb = 1.0$[a]

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(1.30, −16.4)
−0.95	(0.50, −67.6)	0.05	(1.30, −15.3)
−0.90	(0.69, −60.2)	0.10	(1.30, −14.3)
−0.85	(0.82, −54.7)	0.15	(1.29, −13.4)
−0.80	(0.92, −50.1)	0.20	(1.29, −12.5)
−0.75	(0.99, −46.2)	0.25	(1.29, −11.7)
−0.70	(1.05, −42.8)	0.30	(1.28, −10.9)
−0.65	(1.10, −39.7)	0.35	(1.28, −10.2)
−0.60	(1.14, −36.9)	0.40	(1.27, −9.6)
−0.55	(1.18, −34.4)	0.45	(1.27, −8.9)
−0.50	(1.20, −32.0)	0.50	(1.27, −8.3)
−0.45	(1.23, −29.9)	0.55	(1.26, −7.8)
−0.40	(1.24, −27.9)	0.60	(1.25, −7.3)
−0.35	(1.26, −26.1)	0.65	(1.25, −6.8)
−0.30	(1.27, −24.4)	0.70	(1.24, −6.3)
−0.25	(1.28, −22.8)	0.75	(1.24, −5.9)
−0.20	(1.29, −21.3)	0.80	(1.23, −5.5)
−0.15	(1.29, −20.0)	0.85	(1.23, −5.1)
−0.10	(1.29, −18.7)	0.90	(1.22, −4.8)
−0.05	(1.30, −17.5)	0.95	(1.22, −4.5)
		1.00	(1.21, −4.2)

Zeros of $G_+(kx)$: None

Values of $G_+(kx)$ for $kb = 3.0$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.53, 40.9)
−0.95	(2.11, −35.2)	0.05	(0.51, 34.6)
−0.90	(2.56, −14.1)	0.10	(0.51, 28.5)
−0.85	(2.70, 1.3)	0.15	(0.51, 23.0)
−0.80	(2.67, 13.7)	0.20	(0.52, 18.2)
−0.75	(2.56, 23.9)	0.25	(0.53, 14.2)
−0.70	(2.40, 32.5)	0.30	(0.55, 11.0)
−0.65	(2.22, 39.9)	0.35	(0.57, 8.4)
−0.60	(2.02, 46.1)	0.40	(0.59, 6.3)
−0.55	(1.82, 51.3)	0.45	(0.61, 4.6)
−0.50	(1.64, 55.6)	0.50	(0.63, 3.3)
−0.45	(1.46, 58.9)	0.55	(0.65, 2.2)
−0.40	(1.29, 61.3)	0.60	(0.67, 1.4)
−0.35	(1.14, 62.7)	0.65	(0.68, 0.7)
−0.30	(1.00, 63.0)	0.70	(0.70, 0.2)
−0.25	(0.88, 62.2)	0.75	(0.71, −0.2)
−0.20	(0.77, 60.2)	0.80	(0.73, −0.5)
−0.15	(0.69, 57.0)	0.85	(0.74, −0.8)
−0.10	(0.62, 52.5)	0.90	(0.75, −1.0)
−0.05	(0.57, 47.1)	0.95	(0.76, −1.1)
		1.00	(0.77, −1.3)

Zeros of $G_+(kx)$: None

Values of $G_+(kx)$ for $kb = 2.0$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(1.35, 12.3)
−0.95	(1.12, −63.1)	0.05	(1.32, 12.6)
−0.90	(1.46, −49.3)	0.10	(1.29, 12.9)
−0.85	(1.64, −39.2)	0.15	(1.26, 13.0)
−0.80	(1.75, −31.1)	0.20	(1.23, 13.0)
−0.75	(1.81, −24.4)	0.25	(1.21, 12.9)
−0.70	(1.84, −18.7)	0.30	(1.19, 12.8)
−0.65	(1.84, −13.8)	0.35	(1.17, 12.6)
−0.60	(1.82, −9.5)	0.40	(1.15, 12.4)
−0.55	(1.79, −5.8)	0.45	(1.14, 12.1)
−0.50	(1.76, −2.6)	0.50	(1.12, 11.8)
−0.45	(1.72, 0.2)	0.55	(1.11, 11.5)
−0.40	(1.68, 2.6)	0.60	(1.10, 11.2)
−0.35	(1.63, 4.7)	0.65	(1.09, 10.9)
−0.30	(1.59, 6.5)	0.70	(1.08, 10.5)
−0.25	(1.54, 8.0)	0.75	(1.07, 10.2)
−0.20	(1.50, 9.3)	0.80	(1.06, 9.9)
−0.15	(1.46, 10.3)	0.85	(1.06, 9.5)
−0.10	(1.42, 11.2)	0.90	(1.05, 9.2)
−0.05	(1.38, 11.8)	0.95	(1.04, 8.9)
		1.00	(1.04, 8.6)

Zeros of $G_+(kx)$: None

Values of $G_+(kx)$ for $kb = 3.1416$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.00, 135.0)
−0.95	(1.96, 162.7)	0.05	(0.08, 137.7)
−0.90	(2.35, −174.3)	0.10	(0.14, 140.3)
−0.85	(2.44, −157.2)	0.15	(0.20, 142.6)
−0.80	(2.39, −143.2)	0.20	(0.25, 144.8)
−0.75	(2.25, −131.3)	0.25	(0.30, 146.8)
−0.70	(2.08, −120.8)	0.30	(0.34, 148.7)
−0.65	(1.88, −111.6)	0.35	(0.38, 150.4)
−0.60	(1.68, −103.3)	0.40	(0.41, 152.0)
−0.55	(1.48, −95.8)	0.45	(0.44, 153.5)
−0.50	(1.28, −88.9)	0.50	(0.47, 154.8)
−0.45	(1.10, −82.7)	0.55	(0.49, 156.1)
−0.40	(0.93, −77.0)	0.60	(0.51, 157.3)
−0.35	(0.77, −71.8)	0.65	(0.53, 158.4)
−0.30	(0.63, −67.0)	0.70	(0.55, 159.4)
−0.25	(0.49, −62.5)	0.75	(0.57, 160.3)
−0.20	(0.37, −58.4)	0.80	(0.58, 161.2)
−0.15	(0.26, −54.7)	0.85	(0.60, 162.0)
−0.10	(0.16, −51.2)	0.90	(0.61, 162.8)
−0.05	(0.08, −48.0)	0.95	(0.62, 163.5)
		1.00	(0.63, 164.1)

Zeros of $G_+(kx)$: $x = -0.003$,

TABLE I (*continued*).

Values of $G_+(kx)$ for $kb = 3.2$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.34, 136.7)
−0.95	(1.93, 162.7)	0.05	(0.41, 139.4)
−0.90	(2.29, −173.9)	0.10	(0.47, 141.9)
−0.85	(2.33, −156.6)	0.15	(0.52, 144.2)
−0.80	(2.23, −142.4)	0.20	(0.57, 146.4)
−0.75	(2.06, −130.2)	0.25	(0.61, 148.4)
−0.70	(1.85, −119.7)	0.30	(0.64, 150.2)
−0.65	(1.62, −110.3)	0.35	(0.68, 151.9)
−0.60	(1.39, −101.9)	0.40	(0.70, 153.4)
−0.55	(1.17, −94.3)	0.45	(0.73, 154.9)
−0.50	(0.96, −87.4)	0.50	(0.75, 156.2)
−0.45	(0.77, −81.2)	0.55	(0.77, 157.4)
−0.40	(0.59, −75.4)	0.60	(0.79, 158.6)
−0.35	(0.42, −70.1)	0.65	(0.80, 159.6)
−0.30	(0.28, −65.3)	0.70	(0.82, 160.6)
−0.25	(0.14, −60.8)	0.75	(0.83, 161.5)
−0.20	(0.02, −56.8)	0.80	(0.84, 162.4)
−0.15	(0.09, 127.0)	0.85	(0.85, 163.1)
−0.10	(0.18, 130.5)	0.90	(0.86, 163.9)
−0.05	(0.27, 133.7)	0.95	(0.87, 164.5)
		1.00	(0.87, 165.2)

Zeros of $G_+(kx)$: $x = -0.190$,

Values of $G_+(kx)$ for $kb = 5.0$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(1.38, −171.8)
−0.95	(1.92, 174.0)	0.05	(1.34, −171.0)
−0.90	(1.57, −150.8)	0.10	(1.30, −170.5)
−0.85	(0.93, −125.0)	0.15	(1.26, −170.2)
−0.80	(0.27, −104.3)	0.20	(1.22, −170.2)
−0.75	(0.31, 93.0)	0.25	(1.19, −170.3)
−0.70	(0.78, 107.8)	0.30	(1.17, −170.4)
−0.65	(1.14, 120.5)	0.35	(1.15, −170.7)
−0.60	(1.40, 131.7)	0.40	(1.13, −171.0)
−0.55	(1.58, 141.4)	0.45	(1.11, −171.4)
−0.50	(1.68, 149.9)	0.50	(1.10, −171.8)
−0.45	(1.74, 157·2)	0.55	(1.09, −172.1)
−0.40	(1.75, 163.6)	0.60	(1.08, −172.5)
−0.35	(1.74, 169.1)	0.65	(1.07, −172.8)
−0.30	(1.71, 173.7)	0.70	(1.06, −173.2)
−0.25	(1.66, 177.6)	0.75	(1.06, −173.5)
−0.20	(1.61, −179.1)	0.80	(1.05, −173.8)
−0.15	(1.55, −176.5)	0.85	(1.05, −174.1)
−0.10	(1.49, −174.5)	0.90	(1.04, −174.3)
−0.05	(1.44, −172.9)	0.95	(1.04, −174.6)
		1.00	(1.04, −174.8)

Zeros of $G_+(kx)$: $x = -0.778$,

Values of $G_+(kx)$ for $kb = 4.0$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(1.23, 159.6)
−0.95	(1.69, 164.7)	0.05	(1.23, 161.9)
−0.90	(1.75, −166.7)	0.10	(1.24, 163.9)
−0.85	(1.52, −145.6)	0.15	(1.23, 165.6)
−0.80	(1.19, −128.6)	0.20	(1.23, 167.2)
−0.75	(0.83, −114.2)	0.25	(1.22, 168.5)
−0.70	(0.49, −101.7)	0.30	(1.21, 169.8)
−0.65	(0.18, −90.8)	0.35	(1.21, 170.8)
−0.60	(0.10, 98.9)	0.40	(1.20, 171.7)
−0.55	(0.34, 107.5)	0.45	(1.19, 172.6)
−0.50	(0.54, 115.2)	0.50	(1.18, 173.3)
−0.45	(0.70, 122.1)	0.55	(1.17, 173.9)
−0.40	(0.84, 128.3)	0.60	(1.17, 174.5)
−0.35	(0.94, 133.9)	0.65	(1.16, 174.9)
−0.30	(1.03, 138.9)	0.70	(1.15, 175.4)
−0.25	(1.09, 143.4)	0.75	(1.14, 175.7)
−0.20	(1.14, 147.4)	0.80	(1.14, 176.1)
−0.15	(1.18, 151.0)	0.85	(1.13, 176.4)
−0.10	(1.20, 154.2)	0.90	(1.12, 176.6)
−0.05	(1.22, 157.0)	0.95	(1.12, 176.9)
		1.00	(1.11, 177.1)

Zeros of $G_+(kx)$: $x = -0.619$,

Values of $G_+(kx)$ for $kb = 6.0$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.75, −143.1)
−0.95	(2.22, −164.6)	0.05	(0.71, −148.1)
−0.90	(1.18, −122.3)	0.10	(0.69, −153.1)
−0.85	(0.05, 88.7)	0.15	(0.68, −157.7)
−0.80	(1.06, 113.6)	0.20	(0.68, −161.7)
−0.75	(1.78, 134.4)	0.25	(0.69, −165.0)
−0.70	(2.23, 152.2)	0.30	(0.71, −167.7)
−0.65	(2.46, 167.4)	0.35	(0.72, −169.8)
−0.60	(2.51, −179.4)	0.40	(0.74, −171.6)
−0.55	(2.45, −168.0)	0.45	(0.75, −172.9)
−0.50	(2.31, −158.3)	0.50	(0.77, −174.0)
−0.45	(2.12, −150.1)	0.55	(0.78, −174.9)
−0.40	(1.92, −143.4)	0.60	(0.79, −175.6)
−0.35	(1.71, −138.1)	0.65	(0.80, −176.2)
−0.30	(1.50, −134.4)	0.70	(0.81, −176.7)
−0.25	(1.32, −132.2)	0.75	(0.82, −177.0)
−0.20	(1.15, −131.5)	0.80	(0.83, −177.4)
−0.15	(1.01, −132.4)	0.85	(0.84, −177.6)
−0.10	(0.90, −134.8)	0.90	(0.85, −177.8)
−0.05	(0.81, −138.5)	0.95	(0.85, −178.0)
		1.00	(0.86, −178.2)

Zeros of $G_+(kx)$: $x = -0.852$,

TABLE I (*continued*).

Values of $G_+(kx)$ for $kb = 6.2832$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.01, 135.0)
−0.95	(1.97, −145.5)	0.05	(0.10, 138.9)
−0.90	(0.85, −100.4)	0.10	(0.19, 142.4)
−0.85	(0.37, 113.0)	0.15	(0.26, 145.5)
−0.80	(1.31, 140.1)	0.20	(0.33, 148.3)
−0.75	(1.93, 163.0)	0.25	(0.38, 150.8)
−0.70	(2.26, −177.2)	0.30	(0.43, 153.0)
−0.65	(2.38, −159.8)	0.35	(0.47, 155.0)
−0.60	(2.33, −144.3)	0.40	(0.50, 156.8)
−0.55	(2.18, −130.5)	0.45	(0.53, 158.4)
−0.50	(1.97, −118.1)	0.50	(0.56, 159.8)
−0.45	(1.72, −106.9)	0.55	(0.58, 161.1)
−0.40	(1.47, −96.8)	0.60	(0.60, 162.3)
−0.35	(1.22, −87.8)	0.65	(0.61, 163.3)
−0.30	(0.98, −79.6)	0.70	(0.63, 164.3)
−0.25	(0.77, −72.2)	0.75	(0.64, 165.1)
−0.20	(0.57, −65.6)	0.80	(0.66, 165.9)
−0.15	(0.40, −59.6)	0.85	(0.67, 166.6)
−0.10	(0.24, −54.2)	0.90	(0.68, 167.3)
−0.05	(0.11, −49.3)	0.95	(0.68, 167.9)
		1.00	(0.69, 168.4)

Zeros of $G_+(kx)$: $x = -0.866, -0.003,$

Values of $G_+(kx)$ for $kb = 7.0$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(1.15, 155.5)
−0.95	(1.50, −141.2)	0.05	(1.16, 158.7)
−0.90	(0.16, −91.4)	0.10	(1.17, 161.4)
−0.85	(0.95, 125.4)	0.15	(1.18, 163.7)
−0.80	(1.58, 155.0)	0.20	(1.17, 165.7)
−0.75	(1.80, 180.0)	0.25	(1.17, 167.4)
−0.70	(1.72, −158.5)	0.30	(1.16, 168.9)
−0.65	(1.47, −139.7)	0.35	(1.16, 170.1)
−0.60	(1.12, −123.2)	0.40	(1.15, 171.1)
−0.55	(0.75, −108.5)	0.45	(1.14, 172.0)
−0.50	(0.39, −95.5)	0.50	(1.13, 172.8)
−0.45	(0.06, −83.9)	0.55	(1.13, 173.5)
−0.40	(0.23, 106.4)	0.60	(1.12, 174.0)
−0.35	(0.47, 115.6)	0.65	(1.11, 174.5)
−0.30	(0.66, 123.7)	0.70	(1.11, 174.9)
−0.25	(0.81, 130.9)	0.75	(1.10, 175.3)
−0.20	(0.93, 137.2)	0.80	(1.10 175.6)
−0.15	(1.01, 142.8)	0.85	(1.09, 175.9)
−0.10	(1.07, 147.7)	0.90	(1.09, 176.2)
−0.05	(1.12, 151.9)	0.95	(1.08, 176.4)
		1.00	(1.08, 176.6)

Zeros of $G_+(kx)$: $x = -0.894, -0.441,$

Values of $G_+(kx)$ for $kb = 6.4$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.48, 138.3)
−0.95	(1.91, −144.9)	0.05	(0.56, 142.1)
−0.90	(0.73, −99.1)	0.10	(0.63, 145.5)
−0.85	(0.48, 114.9)	0.15	(0.69, 148.5)
−0.80	(1.37, 142.3)	0.20	(0.73, 151.2)
−0.75	(1.90, 165.6)	0.25	(0.77, 153.6)
−0.70	(2.14, −174.3)	0.30	(0.80, 155.7)
−0.65	(2.15, −156.7)	0.35	(0.83, 157.6)
−0.60	(2.02, −141.0)	0.40	(0.85, 159.3)
−0.55	(1.79, −127.0)	0.45	(0.87, 160.8)
−0.50	(1.52, −114.5)	0.50	(0.88, 162.1)
−0.45	(1.23, −103.3)	0.55	(0.90, 163.3)
−0.40	(0.95, −93.2)	0.60	(0.91, 164.4)
−0.35	(0.68, −84.1)	0.65	(0.92, 165.3)
−0.30	(0.44, −75.9)	0.70	(0.92, 166.2)
−0.25	(0.22, −68.5)	0.75	(0.93, 167.0)
−0.20	(0.03, −61.9)	0.80	(0.93, 167.7)
−0.15	(0.13, 124.0)	0.85	(0.94, 168.3)
−0.10	(0.27, 129.3)	0.90	(0.94, 168.9)
−0.05	(0.38, 134.1)	0.95	(0.95, 169.4)
		1.00	(0.95, 169.9)

Zeros of $G_+(kx)$: $x = -0.871, -0.190,$

Values of $G_+(kx)$ for $kb = 9.4248$

x	$G_+(kx)$	x	$G_+(kx)$
−1.00	(0.00, 0.0)	0.00	(0.01, −45.0)
−0.95	(0.40, 88.6)	0.05	(0.13, −40.3)
−0.90	(1.69, −24.1)	0.10	(0.23, −36.1)
−0.85	(2.02, 25.6)	0.15	(0.31, −32.5)
−0.80	(1.24, 65.7)	0.20	(0.38, −29.4)
−0.75	(0.11, 99.6)	0.25	(0.43, −26.6)
−0.70	(0.95, −51.2)	0.30	(0.48, −24.2)
−0.65	(1.72, −25.6)	0.35	(0.52, −22.2)
−0.60	(2.17, −3.0)	0.40	(0.55, −20.3)
−0.55	(2.35, 17.1)	0.45	(0.58, −18.7)
−0.50	(2.32, 35.0)	0.50	(0.60, −17.3)
−0.45	(2.14, 51.0)	0.55	(0.62, −16.1)
−0.40	(1.88, 65.3)	0.60	(0.64, −15.0)
−0.35	(1.59, 78.0)	0.65	(0.65, −14.1)
−0.30	(1.29, 89.4)	0.70	(0.67, −13.2)
−0.25	(1.00, 99.5)	0.75	(0.68, −12.4)
−0.20	(0.74, 108.5)	0.80	(0.69, −11.7)
−0.15	(0.51, 116.4)	0.85	(0.70, −11.1)
−0.10	(0.31, 123.4)	0.90	(0.70, −10.5)
−0.05	(0.14, 129.6)	0.95	(0.71, −10.0)
		1.00	(0.72, −9.5)

Zeros of $G_+(kx)$: $x = -0.943, -0.745, -0.003,$

[a] Tabulated in the form of (magnitude, phase in degrees).

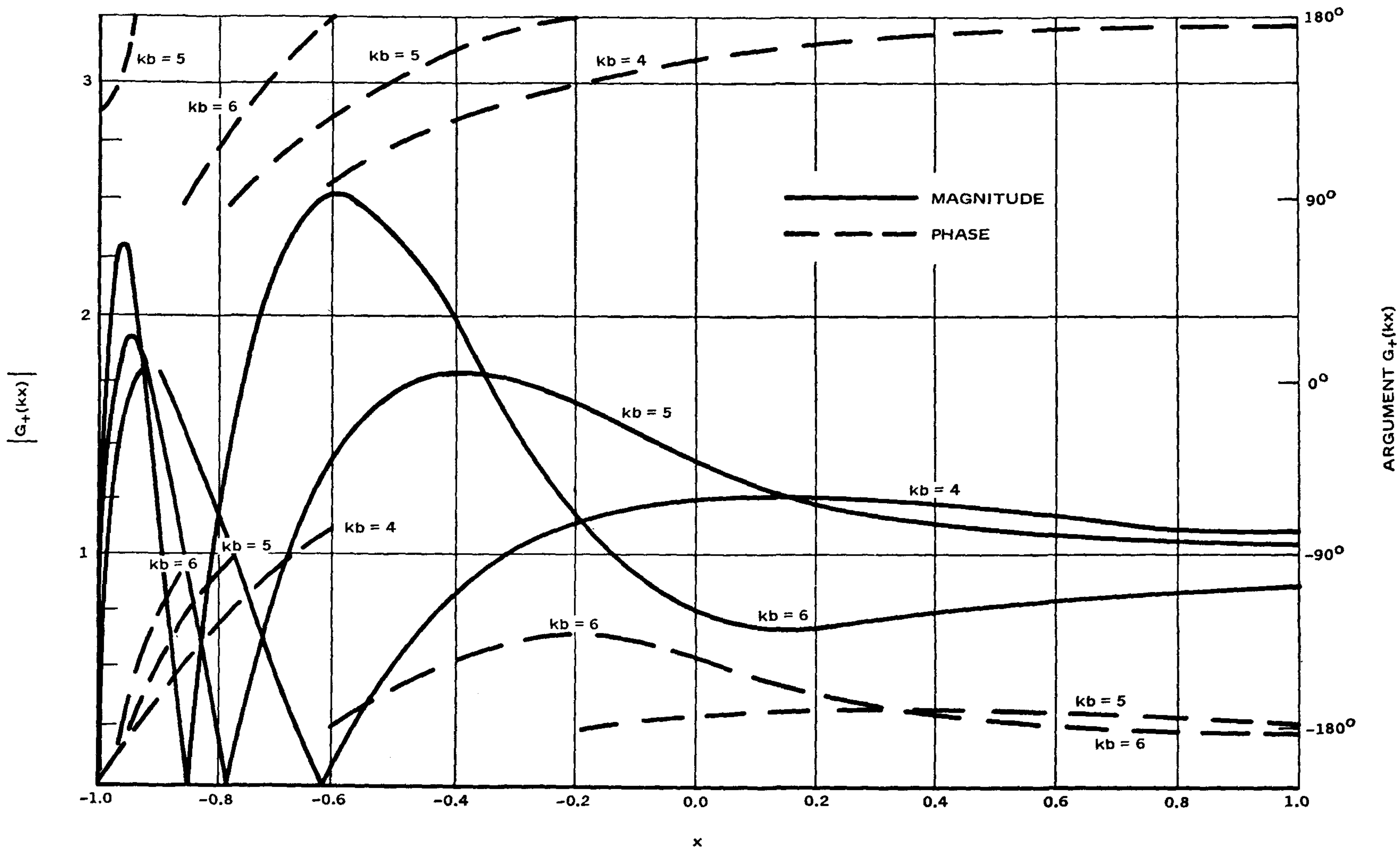


FIG. 6. Values of $G_+(kx)$ computed from (3.7).

where $f(\theta_0)$ and $g(\theta)$ are again given in (6.3) and (6.4). For the diffraction ray at the lower edge, no modification from the corresponding half-plane solution is needed.

Ray-to-Mode Conversion. Let us express the field in the waveguide as

$$E_y = \sum_m 2ic_m\left(\pm\sin\frac{m\pi}{a}x\right)e^{\pm\gamma_m z}. \tag{6.9}$$

Note that,for the wave traveling in the $-z$ direction, we have added -1 in the parenthesis. This is designed to normalize the amplitude of the ray traveling in the direction $(\varphi_m - \pi)$ (so that it becomes unity at $x = z = 0$). With the above definition for the mode, the ray-to-mode conversion factor is again given by (6.6).

Finally, consider a special case $\Omega = \frac{1}{2}\pi$ (Fig. 2). The results given in (5.11) through (5.20) are not valid for this special case, since they were derived from the simplified versions of $\tilde{U}_-(-k\cos w)$, $V_-(-k\cos w)$, and $T_+(-k\cos w)$. Fortunately for this special case, we may reduce our problem into symmetrical and asymmetrical parts by introducing *infinitely* large electric or magnetic ground planes. Hence, we need the solutions to the two configurations shown in Fig. 3(c) and (d). Again, we will omit the details, and give only the final results:

Modified Ray Amplitude. The field on a diffracted ray is modified in the same manner as (6.2) for the TM case and (6.8) for the TE case, except that, instead of (3.6) and (3.7), we have

$$\tilde{G}(\alpha) = 1 + e^{-2\gamma b}, \tag{6.10}$$

$$\tilde{G}_+(\alpha) = \tilde{G}_-(-\alpha) = (2\cos kb)^{\frac{1}{2}} \times \exp\left\{\frac{i\alpha b}{\pi}\left[1 - C + \ln\left(\frac{\pi}{2kb}\right) + i\tfrac{1}{2}\pi\right]\right\} \times \exp\left(\frac{ib\gamma}{\pi}\ln\frac{\alpha-\gamma}{k}\right) \times \prod_{n=1}^{\infty}\left(1 + \frac{\alpha}{i\gamma_{n-\frac{1}{2}}}\right)e^{i\alpha b/(n-\frac{1}{2})\pi}, \tag{6.11}$$

where $\gamma_{n-\frac{1}{2}} = \{[(n-\frac{1}{2})\pi/b]^2 - k^2\}^{\frac{1}{2}}$.

Ray-to-Mode Conversion. Let us express the fields in the waveguide as

$$\text{TM:}\quad H_y = \sum_m 2c_m\cos\frac{(2m-1)x}{2b}\exp(\pm\gamma_{m-\frac{1}{2}}z), \tag{6.12}$$

$$\text{TE:}\quad E_y = \sum_m 2ic_m\left(\pm\sin\frac{(2m-1)x}{2b}\right)\exp(\pm\gamma_{m-\frac{1}{2}}z). \tag{6.13}$$

Then the conversion factor is again in the form of (6.6) with $\tilde{G}(\alpha)$ given in (6.10), i.e.,

$$\left(\left\{\frac{d}{d\alpha}[\gamma\tilde{G}(\alpha)]\right\}^{-1}\right)_{\alpha=k\cos\varphi_{m-\frac{1}{2}}} = \frac{1}{2kb\cos\varphi_{m-\frac{1}{2}}}, \tag{6.14}$$

where $\sin\varphi_{m-\frac{1}{2}} = (m-\frac{1}{2})/kb$, and $0 < \varphi_{m-\frac{1}{2}} < \frac{1}{2}\pi$.

7. APPLICATION TO PROBLEMS

As a first example, consider the problem of radiation from an open-ended waveguide, as shown in Fig. 1(a). Let the incident wave from the left be either TM or TE mode:

(i) Sym. TM:

$$H_y^{(i)} = 2\cos\frac{l\pi}{b}x\exp(-\gamma_l z); \tag{7.1a}$$

(ii) Asym. TM:

$$H_y^{(i)} = 2\cos\frac{(l-\frac{1}{2})\pi}{b}x\exp(-\gamma_{l-\frac{1}{2}}z); \tag{7.1b}$$

(iii) Sym. TE:

$$E_y^{(i)} = 2i\sin\frac{(l-\frac{1}{2})\pi}{b}x\exp(-\gamma_{l-\frac{1}{2}}z); \tag{7.1c}$$

(iv) Asym. TE:

$$E_y^{(i)} = 2i\sin\frac{l\pi}{b}x\exp(-\gamma_l z). \tag{7.1d}$$

Then the configuration in Fig. 1(a) can be reduced to that in Fig. 1(b) for cases (i) and (iv) and to that in Fig. 1(c) for cases (ii) and (iii). Let us denote the reflected fields as:

(i) Sym. TM:

$$H_y = \sum_m 2c_m\cos\frac{m\pi}{b}x\exp(\gamma_m z); \tag{7.2a}$$

(ii) Asym. TM:

$$H_y = \sum_m 2c_m\cos\frac{(m-\frac{1}{2})\pi}{b}x\exp(\gamma_{m-\frac{1}{2}}z); \tag{7.2b}$$

(iii) Sym. TE:

$$E_y = \sum_m 2ic_m\left(-\sin\frac{(m-\frac{1}{2})\pi}{b}x\right)\exp(\gamma_{m-\frac{1}{2}}z); \tag{7.2c}$$

(iv) Asym. TE:

$$E_y = \sum_m 2ic_m\left(-\sin\frac{m\pi}{b}x\right)\exp(\gamma_m z). \tag{7.2d}$$

There is only a single ray diffracted at the edge.

According to the rules given in Sec. 6, the results can be written down immediately:

(i) Sym. TM:

$$c_m = \frac{2i\{\sin \frac{1}{2}(\varphi_l - \pi)G_{-}[k \cos (\varphi_l - \pi)]\}\{\sin \frac{1}{2}(\varphi_m - \pi)G_{-}[k \cos (\varphi_m - \pi)]\}}{2kb\epsilon_m \cos \varphi_m[\cos (\varphi_l - \pi) + \cos (\varphi_m - \pi)]}; \tag{7.3a}$$

(ii) Asym. TM:

$$c_m = \frac{2i\{\sin \frac{1}{2}(\varphi_{l-\frac{1}{2}} - \pi)\tilde{G}_{-}[k \cos (\varphi_{l-\frac{1}{2}} - \pi)]\}\{\sin \frac{1}{2}(\varphi_{m-\frac{1}{2}} - \pi)\tilde{G}_{-}[k \cos (\varphi_{m-\frac{1}{2}} - \pi)]\}}{2kb \cos \varphi_{m-\frac{1}{2}}[\cos (\varphi_{l-\frac{1}{2}} - \pi) + \cos (\varphi_{m-\frac{1}{2}} - \pi)]}; \tag{7.3b}$$

(iii) Sy. TE:

$$c_m = \frac{-2i\{\cos \frac{1}{2}(\varphi_{l-\frac{1}{2}} - \pi)\tilde{G}_{-}[k \cos (\varphi_{l-\frac{1}{2}} - \pi)]\}\{\cos \frac{1}{2}(\varphi_{m-\frac{1}{2}} - \pi)\tilde{G}_{-}[k \cos (\varphi_{m-\frac{1}{2}} - \pi)]\}}{2kb \cos \varphi_{m-\frac{1}{2}}[\cos (\varphi_{l-\frac{1}{2}} - \pi) + \cos (\varphi_{m-\frac{1}{2}} - \pi)]}; \tag{7.3c}$$

(iv) Asym. TE:

$$c_m = \frac{-2i\{\cos \frac{1}{2}(\varphi_l - \pi)G_{-}[k \cos (\varphi_l - \pi)]\}\{\cos \frac{1}{2}(\varphi_m - \pi)G_{-}[k \cos (\varphi_m - \pi)]\}}{2kb \cos \varphi_m[\cos (\varphi_l - \pi) + \cos (\varphi_m - \pi)]}. \tag{7.3d}$$

The above results may be compared with the exact solution (by Wiener–Hopf technique) by Vajnshtejn[5] and Noble.[3] They are *identical*. This is not surprising in view of the fact that $ka \to \infty$. An interesting feature of (7.3) is that, even though the ray method is developed in the present paper for propagating modes, (7.3) is valid for evanescent mode (with φ_m imaginary) as well. [This is because of the fact, in evaluating $\tilde{U}_{-}(-k \cos w)$ in Sec. 4, we assume w to be real.] With the help of Table I, the computation of (7.3) is quite simple.

As a second example, let us consider the problem of radiation from a flanged waveguide having an internal wedge angle β, with an incident TM_{10} mode given in (7.1a). For the special case $b \to \infty$, the ray amplitude is given by the well-known result[2]:

$$H_y = \left\{\frac{\pi}{i(2\pi - \beta)} \sin \frac{\pi^2}{(2\pi - \beta)} \times \left[\left(\cos \frac{\pi(\theta - \theta_0)}{(2\pi - \beta)} - \cos \frac{\pi^2}{(2\pi - \beta)}\right)^{-1} + \left(\cos \frac{\pi(2\pi + \theta + \theta_0)}{(2\pi - \beta)} - \cos \frac{\pi^2}{(2\pi - \beta)}\right)^{-1}\right]\right\} \times \frac{e^{i(k\rho - \frac{1}{4}\pi)}}{(2\pi k\rho)^{\frac{1}{2}}}, \quad b \to \infty, \tag{7.4}$$

which reduces to (6.1) as the wedge angle $\beta \to 0$. For finite b, the ray method gives the solution for the reflection coefficients:

$$c_m = \left\{\frac{\pi}{i(2\pi - \beta)} \sin \frac{\pi^2}{(2\pi - \beta)} \times \left[\left(\cos \frac{\pi(\varphi_l - \varphi_m)}{(2\pi - \beta)} - \cos \frac{\pi^2}{(2\pi - \beta)}\right)^{-1} + \left(\cos \frac{\pi(\varphi_l + \varphi_m)}{(2\pi - \beta)} - \cos \frac{\pi^2}{(2\pi - \beta)}\right)^{-1}\right]\right\} \times \frac{G_{-}[k \cos (\varphi_l - \pi)]G_{-}[k \cos (\varphi_m - \pi)]}{2\epsilon_m kb \cos \varphi_m}. \tag{7.5}$$

For the special cases $\beta = 0$ and $\frac{1}{2}\pi$, the self-reflection coefficient of TEM mode becomes

$$c_0 = G_{+}^2(k)/i4kb \qquad \text{(unflanged waveguide, } \beta = 0\text{)}, \tag{7.6}$$

$$c_0 = G_{+}^2(k)/i3\sqrt{3}\, kb \qquad \text{(flanged waveguide, } \beta = \pi/2\text{)}, \tag{7.7}$$

which show that the reflection coefficient for a flanged waveguide is simply 0.77 times (or 1.15 dB down) of that for an unflanged waveguide [recall that (7.6) is exact]. Note that there is no known exact solution to the flanged waveguide. In the past years, several approximate methods have been suggested. The analyses of these methods are generally quite involved, and the results are by no means as simple as the one given in (7.7). By making use of the table for $G_{+}(\alpha)$, (7.7) can be easily calculated, and an example of such calculations is given in Figs. 7 and 8. For comparison, we have also plotted the result obtained by YFK method.[2] Note that the agreement between YFK method and our method is excellent except when b is very close to a multiple of half wavelength (at the onset of a new propagating mode). When $b = n\lambda/2$, YFK method gives a smooth curve, while (7.7) yields a dip with discontinuous derivative. [From (A1), it is easy to show that the slope of $|G_{+}(\alpha)|$ approaches infinity as $b \to (n\lambda/2)$ from the right-hand side.] It is known from problems with exact solutions [such as (7.6)] that a discontinuity in derivative should be expected at the on-set of a new propagating mode.

In the final example, we compute the diffraction of an incident TM wave given in (2.1) by an array of three waveguides (Fig. 9). Let us concentrate on the field in the middle waveguides and express it in the form of (7.2a). A straightforward ray tracing gives the following results:

$$c_m = \sum_{n=1}^{8} c_m^{(n)}. \tag{7.8}$$

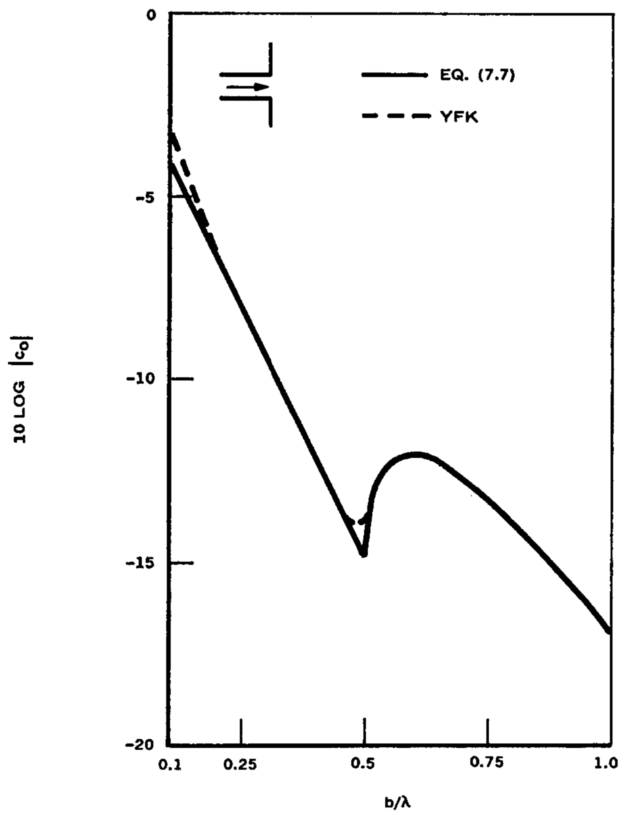


FIG. 7. Magnitude of the reflection coefficient of TEM mode from a flanged waveguide.

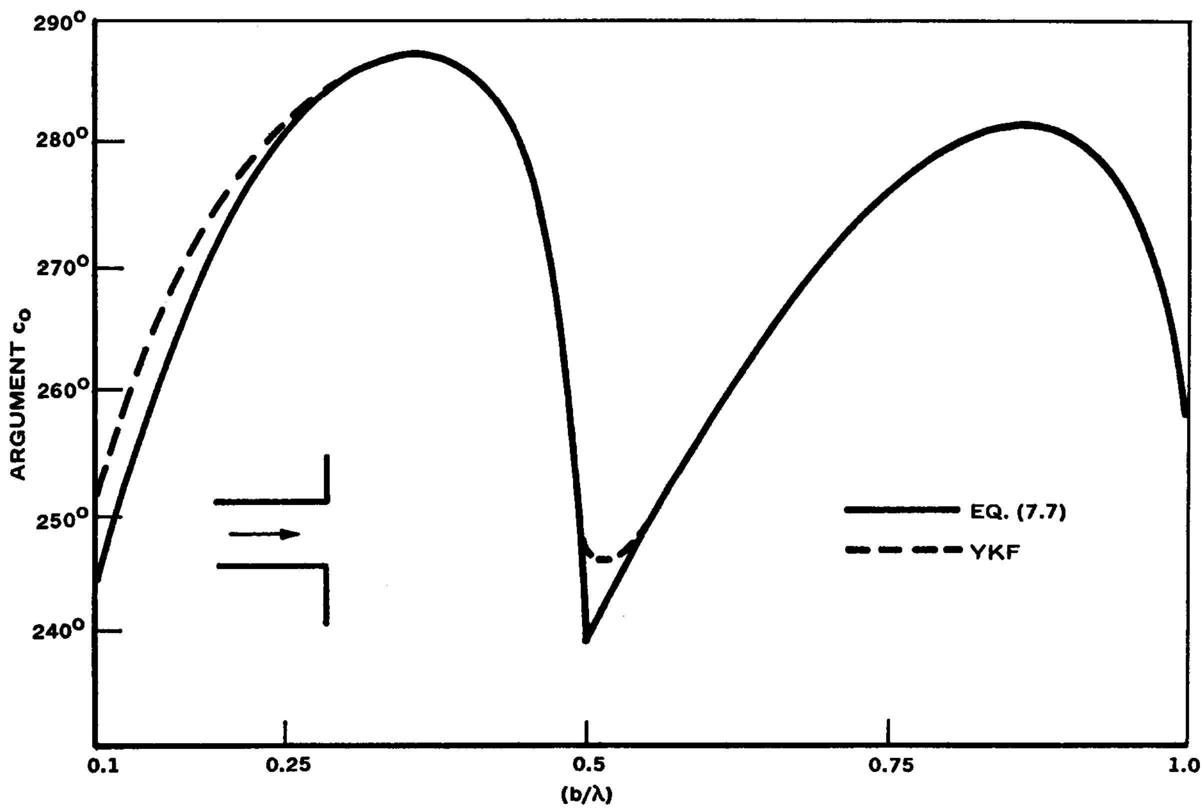


FIG. 8. Phase of the reflection coefficient of TEM mode from a flanged waveguide.

The eight terms in (7.8) are due to eight different rays which are sketched in Fig. 9. From the ray paths and the rules in Sec. 6, it is quite straightforward to write down their explicit forms, except perhaps $c_m^{(6)}$ and $c_m^{(8)}$. In the case of ray 6, the outgoing ray at edge $x = 0$, $z = 0$ falls exactly on the shadow boundary of the incoming ray. Consequently, the simple formula in (6.2) fails to give a finite value. A remedy for this situation is to use (4.10), and the result becomes

$$\begin{aligned} c_m^{(6)} = \Bigg(& e^{ika\cos(\theta_0+\Omega)} \frac{e^{i(ka-\frac{1}{4}\pi)}}{(2\pi ka)^{\frac{1}{2}}} \frac{2i\sin\frac{1}{2}\theta_0 \sin(-\frac{1}{2}\Omega)}{\cos\theta_0 + \cos(-\Omega)} \\ & \times \frac{c(-\Omega < \theta_0) - \exp(i2kb\sin\theta_0) C(\Omega < \theta_0)}{G_+(k\cos\theta_0) G_+[k\cos(-\Omega)]} \Bigg) \\ & \times \Bigg(\tfrac{1}{2} e^{ika} + \frac{e^{-\frac{1}{4}i\pi}}{(\pi)^{\frac{1}{2}}} e^{ika\cos 2(\pi-\Omega)} \\ & \times F[(2ka)^{\frac{1}{2}} \sin(\pi-\Omega)] \Bigg) \left(\frac{G_-[k\cos(\pi-\Omega)]}{G_+[k\cos(-\Omega)]} \right) \\ & \times \Bigg(\frac{2i\sin\frac{1}{2}(\pi-\Omega)\sin\frac{1}{2}(\varphi_m-\pi)}{\cos(\pi-\Omega)+\cos(\varphi_m-\pi)} \\ & \times G_-[k\cos(\pi-\Omega)] G_-[k\cos(\varphi_m-\pi)] \\ & \times \frac{(-1)^m \exp(ikl\cos\varphi_m)}{2\epsilon_m kb\cos\varphi_m} C(0 < \varphi_m < \Omega) \Bigg). \end{aligned} \quad (7.9)$$

The same modification can be used to determine $c_m^{(8)}$.

From the above examples, we note that the ray method for calculating the diffraction or radiation problems involving parallel-plate waveguides is just as straightforward as their counterparts in free space.

ACKNOWLEDGMENTS

The author appreciates the encouragement from Professor G. A. Deschamps, Professor L. B. Felsen, and Professor W. R. Jones.

APPENDIX: OTHER FORMS OF $G_+(\alpha)$

In addition to the expression of $G_+(\alpha)$ given in (3.7), we will give three formulas suitable for certain special situations:

(*i*) *Magnitude of* $G_+(\alpha)$:

$$|G_+(k\cos\theta)| = 2\cos\tfrac{1}{2}\theta e^{-\frac{1}{2}kb\cos\theta} \left(\frac{\sin(kb\sin\theta)}{\sin\theta} \right)^{\frac{1}{2}} \times \prod_{n=1}^{N} \left(\frac{\cos\theta[1-(n\pi/kb)^2]^{\frac{1}{2}}}{[\sin^2\theta - (n\pi/kb)^2]^{\frac{1}{2}}} \right), \quad (A1)$$

where N is the number of propagating modes, i.e., $\{\gamma_n\}$ are real for $n > N$.

(*ii*) *Integral expression of* $G_+(\alpha)$ *for large kb:* For large kb, the infinite product expression given in (3.7)

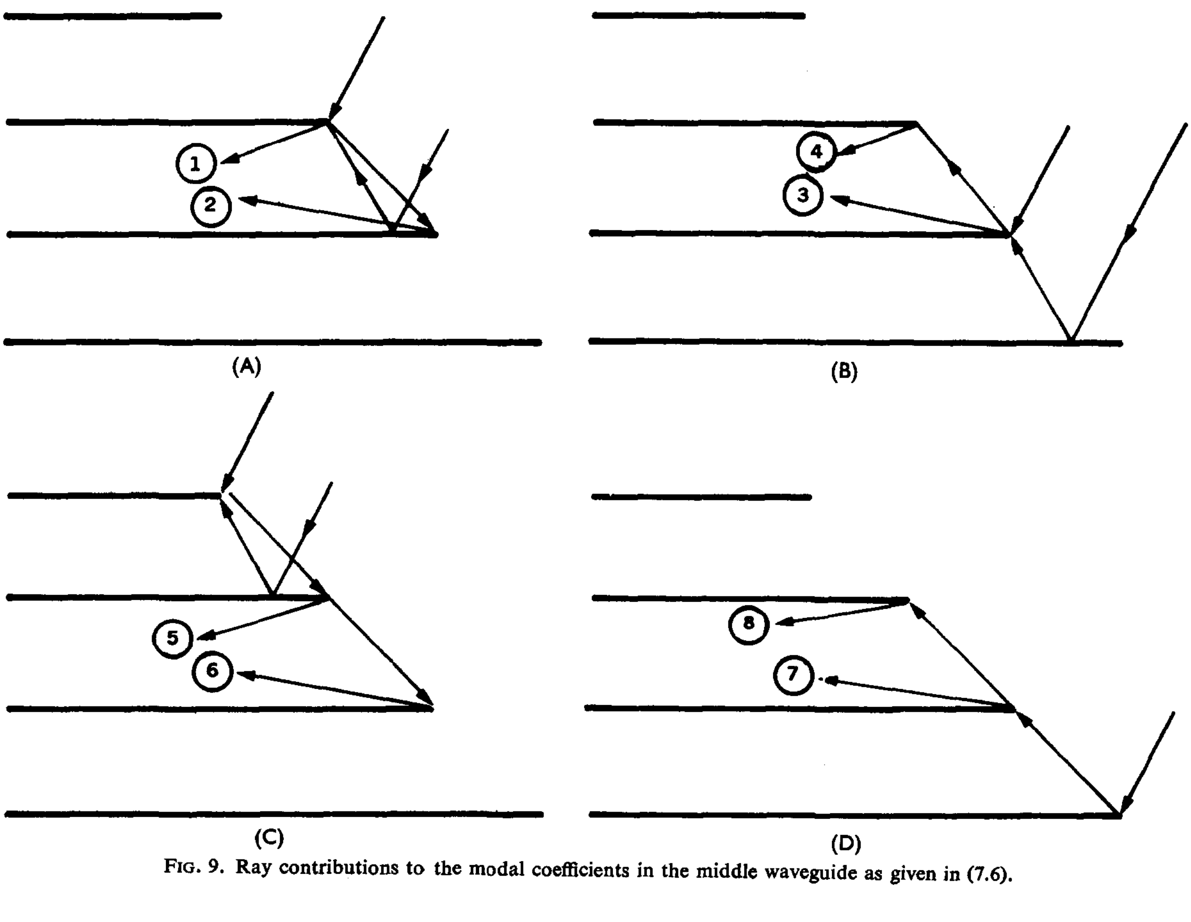


FIG. 9. Ray contributions to the modal coefficients in the middle waveguide as given in (7.6).

converges slowly. By making use of a new factorization formula,[6] it may be shown that

$$G_+(\alpha) = (2 \sin kb)^{\frac{1}{2}} e^{i(\frac{1}{2}kb - \frac{1}{4}\pi)} \times \exp\left\{\frac{1}{2\pi i}\int_P \left[\ln\left(1 + \frac{2\alpha b}{[s(s - i4kb)]^{\frac{1}{2}}}\right)\right] \times \frac{e^{i2kb}}{e^s - e^{i2kb}}\, ds\right\}, \quad \text{(A2)}$$

where the path P goes from $-\infty$ to $0+$ below the real axis of s, circles around the origin, and then goes from $0+$ to $+\infty$ above the real axis. Note that the integrand in the above integral decays exponentially as $\exp(-s)$ and therefore converges quite rapidly. For very large kb, the above integral can be evaluated asymptotically with the result

$$G_+(\alpha) \sim e^{-\frac{1}{4}i\pi}(2\pi\delta)^{\frac{1}{2}}\left(1 + \frac{\alpha b}{(2\pi\delta kb)^{\frac{1}{2}}}\right) \times \exp\left(-0.824\frac{(1-i)\alpha b}{(2kb)^{\frac{1}{2}}}\right),$$

$$kb \gg 1,\ |\alpha| \ll k, \quad \text{and} \quad |\delta| < 0.25, \quad \text{(A3)}$$

where δ is defined from the relation

$$kb = \pi(q + \delta), \quad q = \text{an integer}. \quad \text{(A4)}$$

The formula in (A3) was also given by Vajnshtejn in a paper discussing the laser resonnator.[7] Recently, Bowman gave an interesting expansion of $G_+(\alpha)$ for large ka, and showed its relation to the ray method used by YFK.[8]

[1] J. B. Keller, J. Appl. Phys. **28**, 426 (1957); also, "A Geometric Theory of Diffraction," *Calculus of Variation and its Applications, Symposium of Applied Mathematics* (McGraw-Hill, New York, 1958), Vol. 8, pp. 27–52.

[2] H. Y. Yee, L. B. Felsen, and J. B. Keller, SIAM J. Appl. Math. **16**, 269 (1968).

[3] B. Noble, *Method Based on Wiener–Hopf Technique* (Macmillan, New York, 1958).

[4] G. A. Deschamps, private communication.

[5] L. A. Vajnshtejn, "Theory of Diffraction and Method of Factorization," Izd. Soviet Radio, Moscow, 1966 (English translation by P. Beckmann, The Golden Press, Boulder, Colorado, 1969).

[6] R. Mittra and S. W. Lee, *Analytical Methods in the Theory of Guided Wave* (Macmillan, New York, 1970).

[7] L. A. Vajnshtejn, Zh. Eksp. Teor. Fiz. **44**, 1050 (1963) [Sov. Phys. JETP **17**, 709 (1963)].

[8] J. J. Bowman, IEEE Trans. Antennas Propagation **18**, 131 (1970).

Propagation of Electromagnetic Waves over a Smooth Multisection Curved Earth—An Exact Theory*†

JAMES R. WAIT
Theoretical Studies Groups, ESSA Research Laboratories, Boulder, Colorado 80302

(Received 24 March 1970)

In this paper, we present a self-contained exposition of ground-wave propagation over a smooth inhomogeneous earth model. A factorization procedure is used to obtain an exact solution for a 2-dimensional cylindrical model which may have any number of homogeneous sections. It is demonstrated that, under most conditions in practice, the solution for the mode conversion coefficients can be simplified to a form which was derived earlier by an application of the compensation theorem. Formulas are given which permit the refinement of earlier approximate calculations if such accuracy is required.

1. INTRODUCTION

The theory of ground-wave propagation has been developed mainly for a homogeneous earth. While the results so developed are quite adequate for many purposes, there is a notable difficulty when one is confronted with predicting the field strength for propagation over a path which is partly land and partly sea. While some useful methods have been devised to deal with mixed paths, the theory is not in a particularly sound state. Also, experimental data taken under controlled conditions are scarce. For these reasons, it seems desirable to examine the theory with a view to providing confidence in some of the existing computational methods. Also, hopefully, we point the way to more accurate methods which may be needed in future engineering applications such as precise radio navigation.

We choose a 2-dimensional cylindrical model of the earth. This can be justified for a spherical-earth model where one can show that the curvature transverse to the propagation path does not play a significant role.[1] Our task here is to compute the field for vertically polarized waves circulating in an azimuthal direction around an earth which is composed of any number of homogeneous segments. We will consider initially a path of two segments in order to obtain certain transmission and reflection coefficients. Then we indicate, by considering a three-section path, how the theory may be generalized. Unlike most previous developments, we allow for the existence of waves which are reflected at discontinuities such as coast lines.

2. CONCISE TREATMENT FOR THE HOMOGENEOUS PATH

We, first of all, consider a fully homogeneous path as indicated in Fig. 1. Here we have chosen cylindrical coordinates (r, θ, z) with the surface of the earth at $r = a$. The source of vertically polarized waves is a z-directed uniform magnetic line source of strength K volts. For a time factor $\exp(i\omega t)$, the primary magnetic field of the line source has only a z component H_z^p which, at a distance R, is given by

$$H_z^p = KH_0^{(2)}(kR) \sim K[(2i)/(\pi kR)]^{\frac{1}{2}} \exp(-ikR), \quad (1)$$

where $k = 2\pi/(\text{wavelength})$ and where $H_0^{(2)}$ is a Hankel function with its indicated asymptotic approximation. We are now interested in the expression H_z^p in a contour integral form which is suitable for matching boundary conditions at $r = a$. The latter are exemplified by the statement that the resultant tangential fields E_θ and H_z are related by the specified surface impedance Z. Explicitly,

$$E_\theta = -Z_1 H_z \quad \text{at} \quad r = a. \quad (2)$$

This, in turn, is equivalent to $\partial H_z/\partial x = i\Delta_1 H_z$, where $x = kr$ is set equal to ka. Here, $\Delta_1 = Z_1/\eta_0$, $\eta_0 = 120\pi$.

Using a standard addition theorem for Bessel functions, we can write

$$H_z^p = K \sum_{p=-\infty}^{+\infty} H_p^{(2)}(kr_0)J_p(kr)e^{-ip\theta} \quad \text{for} \quad r < r_0 \quad (3)$$

or in the equivalent integral form

$$H_z^p = K\int_{-\infty}^{+\infty} J_\nu(kr)H_\nu^{(2)}(kr_0)e^{-i\nu\theta} \sum_{q=-\infty}^{\infty} e^{-2\pi iq\nu}\, d\nu \quad \text{for} \quad r < r_0, \quad (4)$$

where p and q, in the above, are integers. In (3) and (4) above, we merely exchange r and r_0 if the results are to apply to the condition $r > r_0$.

We now construct a suitable integral form for H_z which behaves like H_z^p as $R \to 0$ and, at the same time, it satisfies the boundary condition at $r = a$. Without

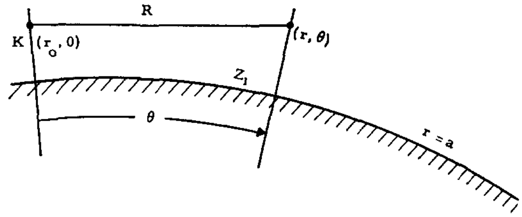


FIG. 1. Homogeneous earth model.

difficulty, we find, for $r < r_0$, that[1]

$$H_z = \tfrac{1}{2}K\int_{-\infty}^{+\infty} H_\nu(kr_0)[H_\nu^{(1)}(kr) + {}_1R_\nu H_\nu^{(2)}(kr)] \times e^{-i\nu\theta}\sum_{q=-\infty}^{\infty} e^{-i2\pi q\nu}\, d\nu, \quad (5)$$

where

$${}_1R_\nu = -\frac{H_\nu^{(1)\prime}(ka) - i\Delta_1 H_\nu^{(1)}(ka)}{H_\nu^{(2)\prime}(ka) - i\Delta_1 H_\nu^{(2)}(ka)}, \quad (6)$$

and where

$$H_\nu^{(1),(2)\prime}(ka) = [\partial H_\nu^{(1),(2)}(x)/\partial x]_{x=ka}. \quad (7)$$

While (5) is an exact solution for the homogeneous cylindrical surface, it is not useful for calculation. A more convenient form is obtained by deforming the contour about the singularities of the integrand. For the situation specified, these are poles at $\nu = \pm\nu_m$, where the ν_m, $m = 1, 2, 3, \cdots$, in the fourth quadrant, are solutions of

$$1/({}_1R_\nu) = 0. \quad (8)$$

Thus, if $\theta > 0$, we close the contour in (5) by an infinite semicircle in the lower half of the ν plane for terms containing $q = 0, 1, 2, \cdots$ and close the contour in the upper half-plane for terms containing $q = -1, -2, -3, \cdots$. By Jordan's lemma, the contributions from the semicircles can be shown to be vanishingly small. Thus, by Cauchy's theorem,

$$H_z = -\pi iK \sum_{m=1,2,3,\cdots} H_{\nu_m}^{(2)}(kr_0)H_{\nu_m}^{(2)}(kr)\left[\frac{\partial}{\partial\nu}\left(\frac{1}{{}_1R_\nu}\right)\right]_{\nu=\nu_m}^{-1} \times \left\{[e^{-i\nu_m\theta} + e^{-i\nu_m(2\pi-\theta)}]\sum_{q=0,1,2,\cdots} e^{-2\pi iq\nu_m}\right\}. \quad (9)$$

As indicated in (9), the summation includes all roots ν_m which are in the fourth quadrant of the complex ν plane. Clearly, H_z can be interpreted as the sum of the "creeping wave" modes of order m which each propagate around the cylinder any number of times in both directions. For lossy surfaces (i.e., Re $\Delta_1 > 0$) and for an electrically large cylinder (i.e, $ka \gg 1$), the curly bracket term in (9) can be replaced by its dominant term $\exp(-i\nu_m\theta)$ for θ less than, but near, π. In what follows, we will assume this is the case.

Thus, for future reference, we write

$$H_z = \sum_m A_m H_{\nu_m}^{(2)}(kr)\exp(-i\nu_m\theta), \quad (10)$$

where

$$A_m = -\pi iK[H_{\nu_m}^{(2)}(kr_0)]\Big/\left[\frac{\partial}{\partial\nu}\left(\frac{1}{{}_1R_\nu}\right)\right]_{\nu=\nu_m}. \quad (11)$$

A more explicit and useful form of the coefficient A_m is obtained by employing the Wronskian relation

$$H_\nu^{(2)}(x)H_\nu^{(1)\prime}(x) - H_\nu^{(1)}(x)H_\nu^{(2)\prime}(x) = 4i/(\pi x) \quad (12)$$

and the modal condition

$$H_\nu^{(2)\prime}(ka) - i\Delta_1 H_\nu^{(2)}(ka) = 0 \quad (13)$$

to show that

$$H_{\nu_m}^{(1)\prime}(ka) - i\Delta_1 H_{\nu_m}^{(1)}(ka) = \frac{4i}{\pi ka}[H_{\nu_m}^{(2)}(ka)]^{-1}. \quad (14)$$

Thus, (11) is equivalent to

$$A_m = -\frac{4K}{ka}\frac{H_{\nu_m}^{(2)}(kr_0)}{H_{\nu_m}^{(2)}(ka)} \times \left(\frac{\partial}{\partial\nu}[H_\nu^{(2)\prime}(ka) - i\Delta_1 H_\nu^{(2)}(ka)]\right)_{\nu=\nu_m}^{-1}. \quad (15)$$

In the special case where the observer is on the surface of the cylinder, we see that (10) has the explicit form

$$H_z]_{r=a} = -\frac{4K}{ka}\sum_m \frac{H_{\nu_m}^{(2)}(kr_0)\exp(-i\nu\theta_m)}{\{\partial[H_\nu^{(2)\prime}(ka) - i\Delta_1 H_\nu^{(2)}(ka)]/\partial\nu\}_{\nu=\nu_m}}, \quad (16)$$

where again we have utilized (12). The contour integral representation equivalent to (16) is clearly given by

$$H_z]_{r=a} = \frac{2K}{i\pi ka}\int_{-\infty}^{+\infty}\frac{H_\nu^{(2)}(kr_0)e^{-i\nu\theta}}{H_\nu^{(2)\prime}(ka) - i\Delta_1 H_\nu^{(2)}(ka)}\, d\nu. \quad (17)$$

3. FORMULATION FOR A TWO-SECTION PATH

We now consider the extension to a two-section mixed path. We shall use the dual integral equation formulation used so successfully by Clemmow[2] and Thompson.[3] For the present type of problem, it seems to be more convenient than the more conventional Green's function formulation,[4,5] which eventually requires solving an integral equation. The latter is equivalent to the dual integral equation mentioned above and, in both cases, the Wiener–Hopf method[6] of factorization can be brought to bear.

The situation is illustrated in Fig. 2. Now, the surface impedance of the lower boundary is sectionally uniform. For $\theta < \theta_1$, the surface impedance is Z_1 but, for $\theta > \theta_1$, the surface impedance is Z_2.

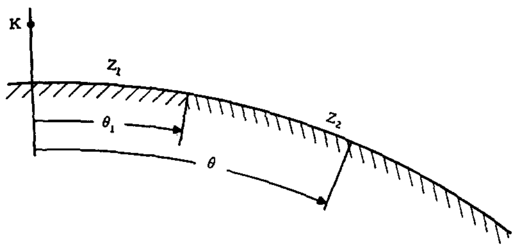


FIG. 2. Two-section mixed path.

For the present situation we assume that the resultant magnetic field H'_z can be written

$$H'_z = H_z + \delta H_z,$$

where H_z is the field for a fully uniform situation (i.e. $Z_2 = Z_1$), while δH_z is the modification resulting from the change of surface impedance from Z_1 to Z_2 for $\theta > \theta_1$. The assigned task is to find an expression for δH_z.

Now, both H'_z and H_z satisfy the wave equation in the region $r > a$. Thus, δH_z will also satisfy the same wave equation, and this leads us to construct the following integral representation:

$$\delta H_z = \frac{2K}{i\pi ka}\int_{-\infty}^{+\infty} \frac{\beta(\nu)H_\nu^{(2)}(kr)}{H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)}}\, e^{-i\nu(\theta-\theta_1)}\, d\nu. \quad (18)$$

Here the function $\beta(\nu)$ is yet to be determined, and the other factors in (18) are introduced for later convenience. Also, for simplicity in (18) and in what follows, we use the convention for Hankel functions of argument ka indicated by

$$H_\nu^{(j)}(ka) = H_\nu^{(j)}, \quad j = 1, 2,$$

and

$$\left(\frac{\partial H_\nu^{(j)}(x)}{\partial x}\right)_{x=ka} = H_\nu^{(j)\prime}, \quad j = 1, 2.$$

We note that the presence of the Hankel function $H_\nu^{(2)}(kr)$ in the integrand of (18) has the appropriate radial dependence for the free-space region $r > a$.

The impedance boundary conditions at $r = a$ are now written

$$\frac{\partial H'_z}{\partial x} = i\Delta_1 H'_z \quad \text{for} \quad \theta < \theta_1$$
$$= i\Delta_2 H'_z \quad \text{for} \quad \theta > \theta_1, \quad \text{at} \quad x = ka. \quad (19)$$

But we know that

$$\frac{\partial H_z}{\partial x} = i\Delta_1 H_z, \quad \text{for all } \theta_1 \text{ at } \quad x = ka.$$

Therefore, a compact statement of the boundary condition at $r = a$ is

$$\left[\left(\frac{\partial}{\partial x} - i\Delta_1\right)\delta H_z\right]_{x=ka} = 0 \quad \text{for} \quad \theta < \theta_1, \quad (20)$$

$$\left[\left(\frac{\partial}{\partial x} - i\Delta_2\right)\delta H_z\right]_{x=ka} = i(\Delta_2 - \Delta_1)[H_z]_{x=ka} \quad \text{for} \quad \theta > \theta_1. \quad (21)$$

Using the integral representations (17) and (18) and the boundary conditions (20) and (21), we find that

$$\int_{-\infty}^{+\infty} \beta(\nu)e^{-i\nu(\theta-\theta_1)}\, d\nu = 0 \quad \text{for} \quad \theta < \theta_1 \quad (22)$$

and

$$\int_{-\infty}^{+\infty} F(\nu)\beta(\nu)e^{-i\nu(\theta-\theta_1)}\, d\nu = i(\Delta_2 - \Delta_1)\int_{-\infty}^{+\infty} G(\hat{\nu})e^{-i\hat{\nu}\theta}\, d\hat{\nu} \quad \text{for} \quad \theta > \theta_1, \quad (23)$$

where

$$F(\nu) = \frac{H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)}}{H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)}} \quad (24)$$

and

$$G(\nu) = \frac{H_\nu^{(2)}(kr_0)}{H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)}}. \quad (25)$$

To reduce the dual integral equations to a simpler form, we use the integral representation

$$\beta(\nu) = \int_{-\infty}^{+\infty} \beta(\nu, \hat{\nu})\, d\hat{\nu} \quad (26)$$

in (22) and assume that the order of integration may be inverted. Then, the pair (21) and (22) are equivalent to

$$\int_{-\infty}^{+\infty} \beta(\nu, \hat{\nu})e^{-i\nu(\theta-\theta_1)}\, d\nu = 0 \quad \text{for} \quad \theta < \theta_1 \quad (27)$$

and

$$\int_{-\infty}^{+\infty} F(\nu)\beta(\nu, \hat{\nu})e^{-i\nu(\theta-\theta_1)}\, d\nu = i(\Delta_2 - \Delta_1)G(\hat{\nu})e^{-i\hat{\nu}\theta} \quad \text{for} \quad \theta > \theta_1. \quad (28)$$

We now follow the Wiener–Hopf procedure[6] and introduce the factorization

$$1/F(\nu) = M^+(\nu)N^-(\nu). \quad (29)$$

Here $M^+(\nu)$ is regular except for poles in the finite part of the complex ν plane and has neither zeros or poles in the *upper* half-plane. Similarly, $N^-(\nu)$ is also regular except for poles in the finite part of the complex ν plane, but it has neither zeros or poles in the lower half-plane.

With the split indicated by (29), we are now in a position to write the formal solution of (27) and (28) as

$$\beta(\nu, \hat{\nu}) = -\frac{(\Delta_2 - \Delta_1)}{2\pi}\, e^{-i\hat{\nu}\theta_1}\, \frac{M^+(\nu)N^-(\hat{\nu})}{\nu - \hat{\nu}}\, G(\hat{\nu}) \quad (30)$$

under the important provision that the contours of integration in (27) and (28) be indented *above* the pole at $\nu = \hat{\nu}$. Thus, when the contour in (27) is closed by an infinite semicircle in the upper half-plane, there are no singularities of the integral enclosed by the contour, and the value of the integral is zero as it should be. Similarly, if the contour in (28) is closed by an infinite semicircle in the lower half-plane, the residue at the enclosed pole at $\nu = \hat{\nu}$ yields the right-hand side of (28).

In carrying out the confirmation of the solution posed by (30), we have invoked Jordan's lemma which requires that $M^+(\nu)$ and $N^-(\nu)$ are bounded at infinity in the upper and lower half-planes, respectively.

Using (26) and (30), we obtain the solution

$$\beta(\nu) = -\frac{\Delta_2 - \Delta_1}{2\pi} M^+(\nu) \int_{-\infty}^{+\infty} \frac{N^-(\hat{\nu})}{\nu - \hat{\nu}} G(\hat{\nu}) e^{-i\hat{\nu}\theta_1} \, d\hat{\nu}, \tag{31}$$

where the integration contour is allowed to pass *below* the pole at $\nu = \hat{\nu}$. Since θ_1 is essentially positive, a residue series representation for (31) is obtained by closing the integration contour in (31) in the lower half-plane. Thus, we find that

$$\beta(\nu) = i(\Delta_2 - \Delta_1) M^+(\nu) \times \sum_m \frac{N^-(\nu_m)}{(\nu - \nu_m)} \frac{H_{\nu_m}^{(2)}(kr_0) e^{-i\nu_m\theta_1}}{[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu=\nu_m}}. \tag{32}$$

The summation here extends over the roots ν_m, $m = 1, 2, 3, \cdots$, of

$$H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)} = 0, \tag{33}$$

which is equivalent to (8).

Using (32) and (18), we now write down the explicit form for the formal solution of the problem:

$$\delta H_z = \frac{2K(\Delta_2 - \Delta_1)}{\pi ka} \sum_m \int_{-\infty}^{+\infty} N^-(\nu_m) M^+(\nu) \times \frac{H_{\nu_m}^{(2)}(kr_0) H_\nu^{(2)}(kr) e^{-i\nu(\theta-\theta_1)} e^{-i\nu_m\theta_1}}{(\nu - \nu_m)[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu_m} (H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})} d\nu. \tag{34}$$

The integration contour in the ν plane here is along the real axis which has no singularities on it. To obtain numerical results for θ near θ_1, it appears that (34) must be integrated numerically. If, on the other hand, kr is sufficiently large, a saddle-point method could be applied to effect the ν integration. The latter approach was discussed by Clemmow[2] and Thompson.[3] Instead, we consider only the residue series evaluation of (34) in the two cases where $\theta < \theta_1$ and $\theta > \theta_1$. These representations are uniformly valid everywhere, although their convergence is poor when θ is near θ_1. An alternate method for handling this region is discussed later.

4. TREATMENT OF REFLECTED WAVES

In the case of the reflected waves (i.e., $\theta - \theta_1 < 0$), we close the integration contour in (34) in the upper half-plane. The enclosed singularities are poles which occur at $\nu = -\nu_p$, where the $\nu_p, p = 1, 2, 3, \ldots$, are solutions of (33). This statement follows from the identity

$$H_{-\nu}^{(2)}(x) = e^{-i\nu\pi} H_\nu^{(2)}(x). \tag{35}$$

To evaluate the residues of the poles at $\nu = -\nu_p$, we note that

$$\begin{aligned}\left[\frac{\partial}{\partial\nu}(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})\right]_{\nu=-\nu_p} &= \left[\frac{\partial}{\partial\nu}[e^{i\nu\pi}(H_{-\nu}^{(2)\prime} - i\Delta_1 H_{-\nu}^{(2)})]\right]_{-\nu=\nu_p} \\ &= e^{-i\nu_p\pi}\left[\frac{\partial}{\partial\nu}(H_{-\nu}^{(2)\prime} - i\Delta_1 H_{-\nu}^{(2)})\right]_{-\nu=\nu_p} \\ &= -e^{-i\nu_p\pi}\left[\frac{\partial}{\partial\nu}(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})\right]_{\nu=\nu_p},\end{aligned} \tag{36}$$

where we have utilized the condition

$$H_{-\nu}^{(2)\prime} - i\Delta_1 H_{-\nu_p}^{(2)} = 0. \tag{37}$$

Then without difficulty we find, for $\theta < \theta_1$, that

$$\delta H_z = -\frac{4K}{ka} \sum_m \sum_p \frac{H_{\nu_m}^{(2)}(kr_0) H_{\nu_p}^{(2)}(kr)}{H_{\nu_m}^{(2)}(ka)} \times R_{p,m} \frac{e^{i\nu_p(\theta-\theta_1)} e^{-i\nu_m\theta_1}}{[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu_m}}, \tag{38}$$

where

$$R_{p,m} = -i(\Delta_2 - \Delta_1) \frac{N^-(\nu_m) N^-(\nu_p)}{\nu_p + \nu_m} \times \frac{H_{\nu_m}^{(2)}(ka)}{[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu_p}}. \tag{39}$$

Here we have used the property that $M^+(-\nu_p) = N^-(\nu_p)$. We may also write (38) in the form

$$\delta H_z = \sum_m \sum_p A_m R_{p,m} H_{\nu_p}^{(2)}(kr) e^{i\nu_p(\theta-\theta_1)} e^{-i\nu_m\theta_1}, \tag{40}$$

where A_m is given by (15). The interpretation of $R_{p,m}$ as a reflection factor follows on comparing (40) with (10).

5. TREATMENT OF THE TRANSMITTED WAVES

We consider the residue series evaluation of (34) for the transmitted waves (i.e., $\theta - \theta_1 > 0$). This requires closing the contour in the lower half-plane. From the definition (29), we see that

$$\frac{M^+(\nu)}{H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)}} = \frac{1}{(H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)})N^-(\nu)}. \quad (41)$$

This tells us that relevant poles occur at $\nu = \mu_n$, $n = 1, 2, 3, \cdots$, which are solutions of

$$H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)} = 0$$

in the lower half-plane. In addition, there is a set of poles at $\nu = \nu_m$. Thus, (34) is equivalent to

$$\delta H_z = -2\pi i\Bigg(\frac{2K(\Delta_2 - \Delta_1)}{\pi ka}\sum_m\sum_n \frac{N^-(\nu_m)}{N^-(\mu_n)}\frac{H_{\nu_m}^{(2)}(kr_0)H_{\mu_n}^{(2)}(kr)}{(\mu_n - \nu_m)[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu_m}[\partial(H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)})/\partial\nu]_{\mu_n}}$$
$$+ \frac{2K(\Delta_2 - \Delta_1)}{\pi ka}\sum_m \frac{H_{\nu_m}^{(2)}(kr_0)H_{\nu_m}^{(2)}(kr)}{[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu_m}(H_{\nu_m}^{(2)\prime} - i\Delta_2 H_{\nu_m}^{(2)})}\Bigg). \quad (42)$$

In the denominator of the second series above, we observe that

$$H_{\nu_m}^{(2)\prime} - i\Delta_2 H_{\nu_m}^{(2)} = -i(\Delta_2 - \Delta_1)H_{\nu_m}^{(2)} \quad (43)$$

by virtue of the relation

$$H_{\nu_m}^{(2)\prime} - i\Delta_1 H_{\nu_m}^{(2)} = 0. \quad (44)$$

As a consequence, this series is identically equal to $-H_z$, which is confirmed by noting the form of (10). Thus, on using the identity $\delta H_z = H_z' - H_z$, we obtain from (42) the following expression for the total field for the region $\theta > \theta_1$:

$$H_z' = -\frac{4K}{ka}\sum_m\sum_n \frac{H_{\nu_m}^{(2)}(kr_0)}{H_{\nu_m}^{(2)}(ka)} \times \frac{H_{\mu_n}^{(2)}(kr)T_{n,m}}{[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu_m}} e^{-i\mu_n(\theta-\theta_1)}e^{-i\nu_m\theta_1}, \quad (45)$$

where

$$T_{n,m} = \frac{i(\Delta_2 - \Delta_1)}{\mu_n - \nu_m}\frac{N^-(\nu_m)}{N^-(\mu_n)}\frac{H_{\nu_m}^{(2)}(ka)}{[\partial(H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)})/\partial\nu]_{\nu=\mu_n}}. \quad (46)$$

We note that (45), for the transmitted modes, is expressible in the form

$$H_z' = \sum_m\sum_n A_m T_{n,m} H_{\mu_n}^{(2)}(kr)e^{-i\mu_n(\theta-\theta_1)}e^{-i\nu_m\theta_1}, \quad (47)$$

which is analogous to (40) for the reflected modes.

6. EXTENSION TO A THREE-SECTION PATH

The series representations given by (40) and (47) constitute the exact solutions for the reflected and transmitted modes at the surface impedance discontinuity at $\theta = \theta_1$. The form of these expressions is such that the results may be generalized to a multisection path. First of all, we consider a three-section path, and then we generalize the results to any number of homogeneous sections.

A three-section path is illustrated in Fig. 3. Here the surface impedance is Z_1 for $\theta < \theta_1$, Z_2 for $\theta_1 < \theta < \theta_2$, and Z_3 for $\theta > \theta_2$. The solution can be constructed by suitably superimposing the two-section solutions discussed in the preceding chapter. The idea is that a mode of order m incident at the discontinuity $\theta = \theta_1$ is reflected as a sum of p modes back to the source according to the reflection factor $^{\leftarrow}R_{p,m}^{(1)}$, which is identically equal to (39). The superimposed arrow pointed toward the left indicates the direction of the reflected modes, and the superscript (1) refers to the discontinuity at $\theta = \theta_1$. In addition, n modes are transmitted beyond the discontinuity at $\theta = \theta_1$, according to the transmission factor $^{\rightarrow}T_{n,m}^{(1)}$. Each of these n modes now propagate toward the discontinuity at $\theta = \theta_2$, where they are converted into l modes and transmitted into the final region. Reflection also takes place at $\theta = \theta_2$, and these reflected modes of order q are in turn reflected at $\theta = \theta_1$, and so on.

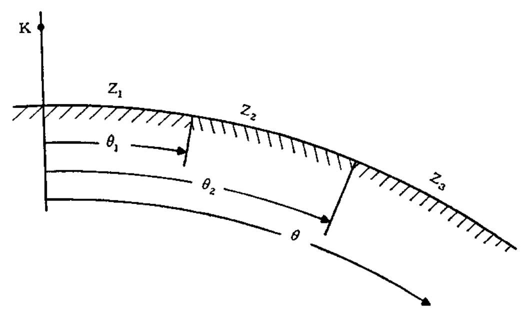


FIG. 3. Three-section mixed path.

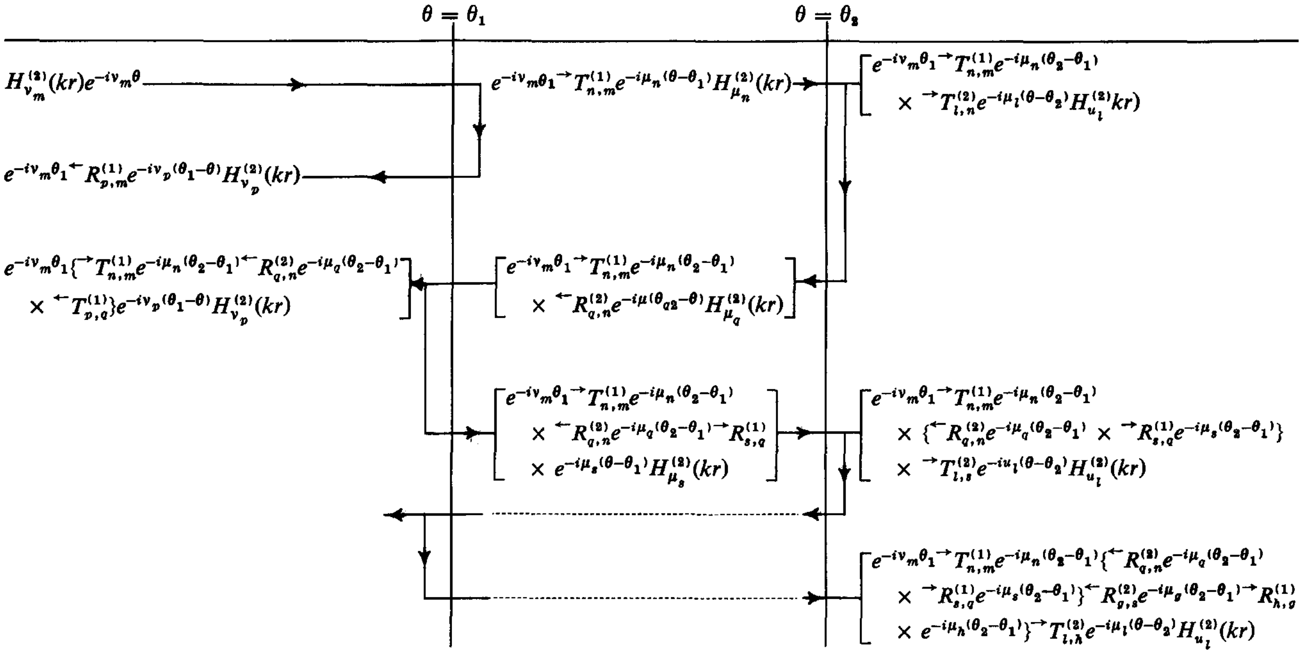


FIG. 4. Flow chart to illustrate modal interaction of the two impedance discontinuities.

The physics of the reflection–conversion and transmission–conversion processes are illustrated in Fig. 4. In this flow chart, we see the mode structure of the field in each of the three regions for a single mode of order m incident from the left The total field in each region is obtained by summing over all converted modes of order n, l, q, p, s, g, h, etc. The arrows over the reflection and transmission (conversion) factors indicate the direction of flow and the superscripts (1) and (2) refer to the discontinuition a $\theta = \theta_1$ and $\theta = \theta_2$, respectively. For example, ${}^{\leftarrow}T^{(1)}_{p,q}$ is given by

$$ {}^{\leftarrow}T^{(1)}_{p,q} = i\frac{(\Delta_1 - \Delta_2)}{\nu_p - \mu_q}\frac{N^{-}(\mu_q)}{N^{-}(\nu_p)}\frac{H^{(2)}_{\mu_q}(ka)}{[\partial(H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})/\partial\nu]_{\nu=\nu_p}} \tag{48} $$

in analogy to (46), which is ${}^{\rightarrow}T^{(1)}_{n,m}$. In a similar fashion, we find that

$$ {}^{\rightarrow}T^{(2)}_{l,n} = i\frac{(\Delta_3 - \Delta_1)}{u_l - \mu_n}\frac{N^{-}(\mu_n)}{N^{-}(u_l)} \times \frac{H^{(2)}_{\mu_n}(ka)}{[\partial(H_\nu^{(2)\prime} - i\Delta_3 H_\nu^{(2)})/\partial\nu]_{\nu=u_l}}, \tag{49} $$

where $\Delta_3 = Z_3/\eta_0$.

In the flow chart depicted in Fig. 4, we show only the low-order interactions between the two discontinuities. To express the results in a comprehensive fashion, we can use a matrix description. For example, if the incident mode from the left, in the region $\theta < \theta_1$, has the form

$$ A_m H^{(2)}_{\nu_m}(kr)e^{-i\nu_m\theta}, \tag{50} $$

then the modes transmitted into region $\theta > \theta_2$ have the form

$$ \sum_l A_m T_{l,m} H^{(2)}_{u_l}(kr)e^{-iu_l(\theta-\theta_2)}, \tag{51} $$

where $T_{l,m}$ is the element in the lth column and the nth row of a matrix $[T]$. This matrix is to be obtained from the operation

$$ [T] = [{}^{\rightarrow}T^{(2)}]([1] + [{}^{\rightarrow}R^{(1)}][{}^{\leftarrow}R^{(2)}] + [{}^{\rightarrow}R^{(1)}][{}^{\leftarrow}R^{(2)}] \times [{}^{\rightarrow}R^{(1)}][{}^{\leftarrow}R^{(2)}] + \cdots)[D][{}^{\rightarrow}T^{(1)}], \tag{52} $$

where [1] is the unity (diagonal) matrix. Here, the general reflection (conversion) factors are

$$ [R^{(j)}] = [D]\begin{bmatrix} R^{(j)}_{1,1} & R^{(j)}_{1,2} & R^{(j)}_{1,3} & \cdots \\ R^{(j)}_{2,1} & R^{(j)}_{2,2} & \cdots & \cdots \\ R^{(j)}_{3,1} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \tag{53a} $$

where

$$ [D] = \begin{bmatrix} e^{-i\mu_1(\theta_2-\theta_1)} & \cdots & 0 & \cdots \\ 0 & e^{-i\mu_2(\theta_2-\theta_1)} & \cdots & \cdots \\ 0 & 0 & e^{-i\mu_3(\theta_2-\theta_1)} & \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{53b} $$

is the product of a square matrix and a diagonal matrix (both of infinite size). Also, the general transmission (conversion) factors are

$$[T^{(j)}] = \begin{bmatrix} T^{(j)}_{1,1} & T^{(j)}_{1,2} & T^{(j)}_{1,3} & \cdots \\ T^{(j)}_{2,1} & T^{(j)}_{2,2} & \cdots\cdots & \\ T^{(j)}_{3,1} & \cdots\cdots & & \\ \cdots & & & \end{bmatrix}. \quad (54)$$

Arrows are to be added to the elements in the square matrices in (53) and (54) as appropriate. Also, $j = 1$ or 2 designates the discontinuity at $\theta = \theta_1$ or $\theta = \theta_2$, respectively, for the three-section path being considered.

An equivalent form of (52) is readily seen to be

$$[T] = [{}^{\rightarrow}T^{(2)}]([1] - [{}^{\rightarrow}R^{(1)}][{}^{\leftarrow}R^{(2)}])^{-1}[D][{}^{\rightarrow}T^{(1)}], \quad (55)$$

which involves a matrix inversion. Formally, this is an exact result for the three-section path but, in practice, only a finite number of modes is considered in individual cases. Thus, the matrices will be truncated so that the final results will be approximate; yet the error can be made arbitrarily small if sufficient computing effort is expended.

7. MULTISECTION PATH

We now come to the final generalization to a multisection path with J homogeneous sections. The surface impedance is now Z_1 for $\theta < \theta_1$, Z_2 for $\theta_1 < \theta < \theta_2$, Z_3 for $\theta_2 < \theta < \theta_3$, and so on, until it is Z_{J-1} for $\theta_{J-2} < \theta < \theta_{J-1}$ and, finally, Z_J for $\theta > \theta_{J-1}$. Then, again, if the incident field from the left is a mode of order m of the form

$$A_m H^{(2)}_{\nu_m}(kr)e^{-i\nu_m\theta}, \quad (56)$$

the transmitted modes of order l in the region $\theta > \theta_{J-1}$ are of the form

$$\sum_l A_m T_{l,m} H^{(2)}_{u_l}(u_l r)e^{-iu_l(\theta-\theta_{J-1})}. \quad (57)$$

Here $T_{l,m}$ is the element in the lth row and the mth column of the matrix $[T]$ given by the product

$$[T] = \prod_{j=J,J-1,\cdots,4,3} [{}^{\rightarrow}T^{(j-1)}]([1] + [{}^{\rightarrow}R^{(j-2)}][{}^{\leftarrow}R^{(j-1)}]) \times [D^{(j-2)}][{}^{\rightarrow}T^{(1)}], \quad (58)$$

which is to be used if $J \geq 3$. Here $D^{(j-2)}$ is the appropriate transmission matrix of the form given by (53b). The presence of the inverse matrix here and in (55) results from the interaction of the modes between adjacent surface impedance discontinuities. In the case where the reflection (conversion) coefficients can be neglected, the inverse matrix can be replaced by [1]. In this case, we have the simple product formula

$$[T] = \prod_{j=J,J-1,\cdots,4,3} [{}^{\rightarrow}T^{(j-1)}][D^{(j-2)}]{}^{\rightarrow}T^{(1)}, \quad (59)$$

which holds for $J \geq 3$. Clearly, (59) accounts for forward mode conversions, but it neglects the mode reconversion which can occur if $J \geq 3$.

8. REDUCTION OF THE SOLUTION

The residue series solutions discussed in the preceding sections are in terms of cylindrical wavefunctions of complex order and large real argument. We now wish to reduce the results to a more useful form for computational work. For the time being, we restrict our discussion to the two-section mixed path with surface impedance Z_1 and Z_2 on either side of a discontinuity at $\theta = \theta_1$.

Using (10), (40), and (47), we can write the following expressions for the resultant field H_z' on either side of the impedance discontinuity:

$$H_z' = \sum_m a_m e^{-i\nu_m\theta}G^{(1)}_m(kr) + \sum_m a_m e^{-i\nu_m\theta_1}\sum_p r_{p,m}e^{-i\nu_p(\theta_1-\theta)}G^{(1)}_p(kr), \quad (60)$$

for $\theta < \theta_1$, and

$$H_z' = \sum_m a_m e^{-i\nu_m\theta_1}\sum S_{n,m}e^{-i\mu_n(\theta-\theta_1)}G^{(2)}_n(kr), \quad (61)$$

for $\theta > \theta_1$, where we have introduced the following new notation:

$$a_m = H^{(2)}_{\nu_m}A_m = -\frac{4K}{ka}\frac{H^{(2)}_{\nu_m}(kr_0)}{[\partial(H_\nu' - i\Delta_1 H_\nu)/\partial\nu]_{\nu=\nu_m}}, \quad (62)$$

$$r_{p,m} = \frac{H^{(2)}_{\nu_p}}{H^{(2)}_{\nu_m}}R_{p,m} = -i(\Delta_2 - \Delta_1)\frac{N^-(\nu_m)N^-(\nu_p)}{(\nu_p + \nu_m)} \times \frac{H^{(2)}_{\nu_p}}{[\partial(H^{(2)\prime}_\nu - i\Delta_1 H^{(2)}_\nu)/\partial\nu]_{\nu=\nu_p}}, \quad (63)$$

$$S_{n,m} = \frac{H^{(2)}_{\mu_n}}{H^{(2)}_{\nu_m}}T_{n,m} = i\frac{(\Delta_1 - \Delta_2)}{(\nu_m - \mu_n)}\frac{N^-(\nu_m)}{N^-(\mu_n)} \times \frac{H^{(2)}_{\mu_n}}{[\partial(H^{(2)\prime}_\nu - i\Delta_2 H^{(2)}_\nu)/\partial\nu]_{\nu=\mu_n}}, \quad (64)$$

$$G^{(1)}_m(kr) = H^{(2)}_{\nu_m}(kr)/H^{(2)}_{\nu_m}, \quad (65)$$

and

$$G^{(2)}_n(kr) = H^{(2)}_{\mu_n}(kr)/H^{(2)}_{\mu_n}. \quad (66)$$

We now introduce the Airy approximations for the Hankel functions.[7] For present purposes, these take the form

$$H^{(1)}_\nu(x) \simeq -(i/\pi^{\frac{1}{2}})(2/x)^{\frac{1}{3}}w_2(\tau), \quad (67)$$

$$H^{(2)}_\nu(x) \simeq (i/\pi^{\frac{1}{2}})(2/x)^{\frac{1}{3}}w_1(\tau), \quad (68)$$

where $\tau = (\nu - x)(2/x)^{\frac{1}{3}}$ and $w_1(\tau)$ and $w_2(\tau)$ are Airy functions. These representations are valid provided $x \gg 1$ and $|\nu - x| \ll x^{\frac{2}{3}}$, which means that the low-order modes of lowest attenuation are adequately described. Also, to within the same approximation, the derivatives of the Hankel functions can be represented by the derivatives of the Airy functions. Thus,

$$H_\nu^{(1)\prime}(x) \simeq (i/\pi^{\frac{1}{2}})(2/x)^{\frac{2}{3}} w_2'(\tau), \tag{69}$$

$$H_\nu^{(2)\prime}(x) \simeq -(i/\pi^{\frac{1}{2}})(2/x)^{\frac{2}{3}} w_1'(\tau). \tag{70}$$

It is useful to note that the Airy functions used here are simply related to the more conventional forms. Thus,

$$w_1(\tau) = \pi^{\frac{1}{2}}[\mathrm{Bi}\,(\tau) - i\,\mathrm{Ai}\,(\tau)] \tag{71}$$

and

$$w_2(\tau) = \pi^{\frac{1}{2}}[\mathrm{Bi}\,(\tau) + i\,\mathrm{Ai}\,(\tau)], \tag{72}$$

where $\mathrm{Ai}\,(\tau)$ and $\mathrm{Bi}\,(\tau)$ have been tabulated extensively.[8]

We now readily find that the desired roots ν_m and μ_n are obtained from

$$\nu_m \cong ka + (ka/2)^{\frac{1}{3}} t_m^{(1)} \tag{73}$$

and

$$\mu_n \cong ka + (ka/2)^{\frac{1}{3}} t_n^{(2)}, \tag{74}$$

where $t_m^{(1)}$ and $t_n^{(2)}$ are roots of

$$w_1'(t) - q_1 w_1(t) = 0 \tag{75}$$

and

$$w_1'(t) - q_2 w_1(t) = 0, \tag{76}$$

respectively. Here q_1 and q_2 are normalized impedance parameters given by

$$q_1 = -i(ka/2)^{\frac{1}{3}}\Delta_1, \qquad \Delta_1 = Z_1/\eta_0,$$

and

$$q_2 = -i(ka/2)^{\frac{1}{3}}\Delta_2, \qquad \Delta_2 = Z_2/\eta_0.$$

We note that (75) and (76) are the Airy function approximations for the exact forms

$$H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)} = 0, \quad \nu = \nu_m, \tag{77}$$

and

$$H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)} = 0, \quad \nu = \mu_n, \tag{78}$$

respectively.

Using the Airy representations, we easily find that the reflection (conversion) coefficient and the transmission (conversion) coefficient as defined by (63) and (64) are now approximated by

$$r_{p,m} \cong -\frac{(q_2 - q_1)N^-(\nu_m)N^-(\nu_p)}{[4(ka/2)^{\frac{2}{3}} + t_p^{(1)} + t_m^{(1)}](t_p^{(1)} - q_1^2)} \tag{79}$$

and

$$S_{n,m} \cong \frac{q_1 - q_2}{t_m^{(1)} - t_n^{(2)}} \frac{N^-(\nu_m)}{N^-(\mu_n)} \frac{1}{(t_n^{(2)} - q_2^2)}. \tag{80}$$

In obtaining these forms, we have used the fact that $w_1(t)$ and $w_2(t)$ satisfy the Stokes differential equation

$$\frac{d^2}{dt^2} w_{1,2}(t) - t w_{1,2}(t) = 0 \tag{81}$$

for any argument t, while (75) and (76) tell us that

$$w_1'(t_p^{(1)}) = q_1 w_1(t_p^{(1)}) \tag{82}$$

and

$$w_1'(t_n^{(2)}) = q_2 w_1(t_n^{(2)}). \tag{83}$$

9. THE FACTORIZATION

We now return to the problem of factorizing $F(\nu)$, which is defined by (24). We write this in the equivalent form

$$F(\nu) = f_2(\nu)/f_1(\nu), \tag{84}$$

where

$$f_1(\nu) = (H_\nu^{(2)\prime} - i\Delta_1 H_\nu^{(2)})e^{-i\nu\pi/2}$$

and

$$f_2(\nu) = (H_\nu^{(2)\prime} - i\Delta_2 H_\nu^{(2)})e^{-i\nu\pi/2}.$$

Using the identity $H_{-\nu}^{(2)}(x) = e^{-i\nu\pi}H_\nu^{(2)}(x)$, we readily deduce that both $f_1(\nu)$ and $f_2(\nu)$ are even functions of ν. Also, $f_1(\nu)$ and $f_2(\nu)$ have no singularities in the finite ν plane, but they do have zeros at $\nu = \nu_s$ and $\nu = \mu_s$, where the subscript s indicates the order of the root. We are now permitted to use the infinite product theorem to obtain

$$f_1(\nu) = f_1(0) \prod_s \left(1 - \frac{\nu^2}{\nu_s^2}\right) \tag{85}$$

and

$$f_2(\nu) = f_2(0) \prod_s \left(1 - \frac{\nu^2}{\mu_s^2}\right). \tag{86}$$

Then, (84) may be written

$$F(\nu) = F(0) \prod_s \left(1 - \frac{\nu}{\mu_s}\right)\left(1 + \frac{\nu}{\mu_s}\right) \Big/ \left(1 - \frac{\nu}{\nu_s}\right)\left(1 + \frac{\nu}{\nu_s}\right)$$

$$= \frac{1}{M^+(\nu)N^-(\nu)}. \tag{87}$$

Noting that $\nu = \nu_s$ and $\nu = \mu_s$ occur in the lower half ν plane, we see by inspection that

$$M^+(\nu) = [F(0)]^{-\frac{1}{2}} \prod_s \left(1 - \frac{\nu}{\nu_s}\right) \Big/ \left(1 - \frac{\nu}{\mu_s}\right) \tag{88}$$

is regular and free of zeros in the upper half-plane. With similar reasoning, we find that

$$N^-(\nu) = [F(0)]^{-\frac{1}{2}} \prod_s \left(1 + \frac{\nu}{\nu_s}\right) \Big/ \left(1 + \frac{\nu}{\mu_s}\right). \tag{89}$$

We defer a detailed study here of these infinite products. However, we can obtain two significant pieces of information. First of all, the quantity

$N^{-}(\nu_m)N^{-}(\nu_p)$ appearing in (79) is of the order of $[F(0)]^{-1}$, where

$$F(0) = \frac{H_0^{(2)\prime} - i\Delta_2 H_0^{(2)}}{H_0^{(2)\prime} - i\Delta_1 H_0^{(2)}} \simeq \frac{1+\Delta_2}{1+\Delta_1}, \quad (90)$$

which is itself of the order of unity. The second quantity of interest is the ratio $N^{-}(\nu_m)/N^{-}(\mu_n)$ which occurs in (46) and (64) for the transmission coefficients. From (89), it is evident that

$$\frac{N^{-}(\nu_m)}{N^{-}(\mu_n)} = \prod_s \left(1+\frac{\mu_n}{\mu_s}\right)\left(1+\frac{\nu_m}{\nu_s}\right) \Big/ \left(1+\frac{\mu_n}{\nu_s}\right)\left(1+\frac{\nu_m}{\mu_s}\right). \quad (91)$$

For many cases of practical interest, such as ground wave propagation over mixed paths, the infinite product given by (91) is quite close to unity. This is a consequence of the fact that the ratio ν_s/μ_s itself does not differ much from unity if $|\Delta_1|$ and $|\Delta_2| \ll 1$. In fact, in general, we can show that

$$\lim_{s\to\infty} (\nu_s/\mu_s) = 1.$$

The consequence of the above discussion is that the transmission (conversion) coefficient $S_{n,m}$ can be well approximated by the remarkably simple relationship

$$S_{n,m} \cong \frac{q_1 - q_2}{t_m^{(1)} - t_n^{(2)}} \frac{1}{t_n^{(2)} - q_2^2}. \quad (92)$$

Apparently, there is no such simple expression for the reflection factor $r_{p,m}$, although an order-of-magnitude estimate suggests that

$$\frac{r_{p,m}}{S_{n,m}} \sim \frac{t_m^{(1)} - t_n^{(2)}}{t_m^{(1)} + t_p^{(1)}} \frac{t_n^{(2)} - q_2^2}{t_p^{(1)} - q_2^2} \frac{1}{4(ka/2)^{\frac{2}{3}}}, \quad (93)$$

which, at least for the lower-order modes, has a magnitude very small compared with unity (typically of the order of 10^{-4}). Thus, the neglect of the reflected mode at the discontinuity has some justification in fact.

10. FINAL REMARKS

As a final closure in this discussion, we take the simplified form $S_{n,m}$ and use it to obtain an explicit, albeit approximate, expression for the resultant magnetic field H_z' for $\theta > \theta_1$ above a two-section path. In this case, we introduce an "attenuation function" W' which is defined as follows:

$$H_z' = H_0 W',$$

where

$$H_0 = 2KH_0^{(2)}(ka\theta) \simeq 2K\left(\frac{2i}{\pi ka\theta}\right)^{\frac{1}{2}} e^{-ika\theta} \quad (94)$$

is a reference field which is numerically equal to the field of the magnetic line source of strength K located on the surface of a flat perfect conductor at a *linear* distance $ka\theta$. Having the field normalized in this fashion means that the "attenuation function" W' reduces to unity if the earth were flat and perfectly conducting.

In order to facilitate the discussion, we also introduce some of the dimensionless parameters which are now common in ground-wave propagation theory. These are $x_1 = (ka/2)^{\frac{1}{3}}\theta_1$, $x = (ka/2)^{\frac{1}{3}}\theta$, $y_0 = [2/(ka)]^{\frac{1}{3}}k(r_0 - a)$, and $y = [2/(ka)]^{\frac{1}{3}}k(r - a)$. Then, without difficulty, we find that

$$W' \simeq W \simeq (\pi x)^{\frac{1}{2}} e^{-i\pi/4} \times \sum_{m=1,2,3,\cdots} \frac{1}{t_m^{(1)} - q_1^2} e^{-ixt_m^{(1)}} G_m^{(1)}(y_0) G_m^{(1)}(y), \quad \text{for } x < x_1, \quad (95)$$

and

$$W' \simeq (\pi x)^{\frac{1}{2}} e^{-i\pi/4} \sum_{m=1,2,3,\cdots} \frac{1}{t_m^{(1)} - q_1^2} e^{-ix_1 t_m^{(1)}} G_m^{(1)}(y_0) \times \sum_{n=1,2,3,\cdots} S_{n,m} e^{-i(x-x_1)t_n^{(2)}} G_n^{(2)}(y), \quad \text{for } x > x_1, \quad (96)$$

where

$$S_{n,m} = (q_1 - q_2)(t_m^{(1)} - t_n^{(2)})^{-1}(t_n^{(2)} - q_2^2)^{-1}. \quad (97)$$

It is of considerable interest to observe that (96) has precisely the same form as the expression derived from an application of the compensation theorem applied to an equivalent problem.[7] In fact, (96), for $y = y_0 = 0$, is equivalent to the integral formulas

$$W' = W(x, q_1) + \left(\frac{x}{\pi i}\right)^{\frac{1}{2}}(q_2 - q_1) \times \int_0^{x-x_1} \frac{W(x-\hat{x}, q_1)W(\hat{x}, q_2)}{[\hat{x}(x-\hat{x})]^{\frac{1}{2}}} d\hat{x} \quad (98)$$

$$= W(x, q_2) + \left(\frac{x}{\pi i}\right)^{\frac{1}{2}}(q_1 - q_2) \times \int_0^{x} \frac{W(x-\hat{x}, q_2)W(\hat{x}, q_1)}{[\hat{x}(x-\hat{x})]^{\frac{1}{2}}} d\hat{x}, \quad (99)$$

where

$$W(x, q_j) = \left(\frac{\pi x}{i}\right)^{\frac{1}{2}} \sum_{m=1,2,3,\cdots} \frac{1}{t_m^{(j)} - q_j^2} e^{-ixt_m^{(j)}}, \quad j = 1, 2. \quad (100)$$

That the forms given by (98) and (99) lead back to (96) can be readily verified by simple integration, provided we recognize that

$$(q_2 - q_1) \sum_{n=1,2,3,\cdots} \frac{1}{(t_n^{(2)} - t_m^{(1)})(t_n^{(2)} - q_2^2)} = 1. \quad (101)$$

This, in turn, is a consequence of the identity

$$\frac{w_1(t)}{w_1'(t) - q_2 w_1(t)} = \sum_n \frac{1}{(t_n^{(2)} - q_2^2)(t - t_n^{(2)})}, \quad (102)$$

which can be verified by integrating both sides over a contour in the complex t plane which encloses the poles at $t = t_n^{(2)}$ in a clockwise sense. In (202), we then set $t = t_m^{(1)}$ and use the fact that

$$w_1'(t_m^{(1)}) = q_1 w_1(t_m^{(1)}), \quad (103)$$

which leads back to (101).

The integral formulas for (98) and (99) are really most suitable for numerical work when the distance from the discontinuity (i.e., the coast line) is small. In this case, the convergence of the n series in (96) is poor because of the smallness of $x - x_1$. Other representations also exist and their relative merit in numerical work has already been described.[1]

ACKNOWLEDGMENTS

I am extremely grateful to Mrs. Eileen Brackett for her help in preparing the manuscript. I also thank Professor E. Bahar for his very useful comments on this paper.

* Contribution from the Cooperative Institute of Research in Environmental Sciences, University of Colorado, Boulder, Colorado.

† This research was supported in part by Air Force Cambridge Research Laboratories under Contract PRO-CP-69-824.

[1] J. R. Wait, in *Advances in Radio Research*, J. A. Saxton, Ed. (Academic, New York, 1964), Vol. 1, pp. 157–217 [includes an extensive bibliography; errata list for this review paper is in IEEE Trans. Antennas Propagation **17**, 220 (1969)].

[2] P. C. Clemmow, Phil. Trans. Roy. Soc. (London) **A246**, 1 (1953).

[3] J. R. Thompson, Proc. Roy. Soc. (London) **A267**, 183 (1962).

[4] A. E. Heins and H. Feshbach, J. Math. & Phys. **26**, 143 (1947).

[5] J. Bazer and S. N. Karp, J. Res. NBS **66D**(3), 319 (1962).

[6] B. Noble, *The Wiener-Hopf Technique* (Pergamon, New York, 1958).

[7] F. W. J. Olver, "Bessel functions of integer order," in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, D.C., 1964).

[8] J. C. P. Miller, *The Airy Integral* (Cambridge U.P., Cambridge, England, 1946).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Spatially Homogeneous World Models*

István Ozsváth
The University of Texas at Dallas, Dallas, Texas

(Received 19 January 1970)

We obtain the field equations of Einstein for spatially homogeneous spaces as the Euler–Lagrange equations of a variational problem. We write these equations in Hamiltonian form and regularize them. In this way, we obtain a class of solutions without rotations. We derive, in particular, the Lagrangian function for the rotating model with the S^3 group first computed by Gödel. We suggest that the corresponding Hamiltonian equations can be regularized.

1. INTRODUCTION

It is the main objective of some cosmologists nowadays to treat the following outstanding problem of the relativistic cosmology: Consider the line element

$$ds^2 = dt^2 + A(\omega^1)^2 + B(\omega^2)^2 + C(\omega^3)^2 + 2D\omega^2\omega^3, \quad (1.1)$$

where ω^1, ω^2, and ω^3 are the invariant differential forms of the group S^3 satisfying the relations

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2, \quad (1.2)$$

and A, B, C, and D are functions of t only. We find the solution in form (1.1) to Einstein's field equations with dust such that the rotation and the expansion of the matter are different from zero. This is interesting not because the astronomers had discovered the rotation of the universe, but it is interesting from a theoretical point of view. This model probably would be the simplest world model with finite space part and with the most general motion of the "Weltsubstrat," that is, with nonvanishing translation, rotation, expansion, and shear.

We had this problem in mind when developing this paper, the structure of which is as follows: Using an idea of Weyl, we obtain Einstein's field equations for spatially homogeneous spaces as the Euler–Lagrange equations of a variational problem. More precisely, we obtain the vacuum field equations for all the groups and the field equations with incoherent matter for Class I groups only. We call Class I the Bianchi Type

This, in turn, is a consequence of the identity

$$\frac{w_1(t)}{w_1'(t) - q_2 w_1(t)} = \sum_n \frac{1}{(t_n^{(2)} - q_2^2)(t - t_n^{(2)})}, \quad (102)$$

which can be verified by integrating both sides over a contour in the complex t plane which encloses the poles at $t = t_n^{(2)}$ in a clockwise sense. In (202), we then set $t = t_m^{(1)}$ and use the fact that

$$w_1'(t_m^{(1)}) = q_1 w_1(t_m^{(1)}), \quad (103)$$

which leads back to (101).

The integral formulas for (98) and (99) are really most suitable for numerical work when the distance from the discontinuity (i.e., the coast line) is small. In this case, the convergence of the n series in (96) is poor because of the smallness of $x - x_1$. Other representations also exist and their relative merit in numerical work has already been described.[1]

ACKNOWLEDGMENTS

I am extremely grateful to Mrs. Eileen Brackett for her help in preparing the manuscript. I also thank Professor E. Bahar for his very useful comments on this paper.

* Contribution from the Cooperative Institute of Research in Environmental Sciences, University of Colorado, Boulder, Colorado.

† This research was supported in part by Air Force Cambridge Research Laboratories under Contract PRO-CP-69-824.

[1] J. R. Wait, in *Advances in Radio Research*, J. A. Saxton, Ed. (Academic, New York, 1964), Vol. 1, pp. 157–217 [includes an extensive bibliography; errata list for this review paper is in IEEE Trans. Antennas Propagation **17**, 220 (1969)].

[2] P. C. Clemmow, Phil. Trans. Roy. Soc. (London) **A246**, 1 (1953).

[3] J. R. Thompson, Proc. Roy. Soc. (London) **A267**, 183 (1962).

[4] A. E. Heins and H. Feshbach, J. Math. & Phys. **26**, 143 (1947).

[5] J. Bazer and S. N. Karp, J. Res. NBS **66D**(3), 319 (1962).

[6] B. Noble, *The Wiener-Hopf Technique* (Pergamon, New York, 1958).

[7] F. W. J. Olver, "Bessel functions of integer order," in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, D.C., 1964).

[8] J. C. P. Miller, *The Airy Integral* (Cambridge U.P., Cambridge, England, 1946).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Spatially Homogeneous World Models*

István Ozsváth
The University of Texas at Dallas, Dallas, Texas

(Received 19 January 1970)

We obtain the field equations of Einstein for spatially homogeneous spaces as the Euler–Lagrange equations of a variational problem. We write these equations in Hamiltonian form and regularize them. In this way, we obtain a class of solutions without rotations. We derive, in particular, the Lagrangian function for the rotating model with the S^3 group first computed by Gödel. We suggest that the corresponding Hamiltonian equations can be regularized.

1. INTRODUCTION

It is the main objective of some cosmologists nowadays to treat the following outstanding problem of the relativistic cosmology: Consider the line element

$$ds^2 = dt^2 + A(\omega^1)^2 + B(\omega^2)^2 + C(\omega^3)^2 + 2D\omega^2\omega^3, \quad (1.1)$$

where ω^1, ω^2, and ω^3 are the invariant differential forms of the group S^3 satisfying the relations

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2, \quad (1.2)$$

and A, B, C, and D are functions of t only. We find the solution in form (1.1) to Einstein's field equations with dust such that the rotation and the expansion of the matter are different from zero. This is interesting not because the astronomers had discovered the rotation of the universe, but it is interesting from a theoretical point of view. This model probably would be the simplest world model with finite space part and with the most general motion of the "Weltsubstrat," that is, with nonvanishing translation, rotation, expansion, and shear.

We had this problem in mind when developing this paper, the structure of which is as follows: Using an idea of Weyl, we obtain Einstein's field equations for spatially homogeneous spaces as the Euler–Lagrange equations of a variational problem. More precisely, we obtain the vacuum field equations for all the groups and the field equations with incoherent matter for Class I groups only. We call Class I the Bianchi Type

I, II, VIII, and IX groups, characterized by

$$d\omega^1 = 0, \quad d\omega^2 = 0, \quad d\omega^3 = 0, \tag{1.3}$$

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = 0, \quad d\omega^3 = 0, \tag{1.4}$$

$$d\omega^1 = \omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2, \tag{1.5}$$

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2. \tag{1.6}$$

Our reasons for this should be clear later.

We apply the general theory to the line element

$$ds^2 = dt^2 - (A\omega^1)^2 - (B\omega^2)^2 - (B\omega^3)^2 \tag{1.7}$$

with each of the groups (1.3)–(1.6). We write the Euler–Lagrange equations in Hamiltonian form. We regularize these equations; that is, we introduce a new variable τ by a suitable transformation on t such that these equations transform into an *analytic* system. This system is then easily solved by a computer, or one can think of the solution developed into convergent power series with respect to the regularizing parameter τ. As a second application, we derive the Lagrangian function for (1.1) with (1.2) first given by Gödel.[1] It is obvious that there are several ways to regularize the corresponding canonical equations; therefore, we can say that *the problem of the rotating universe can be solved by regularization*, in the same sense as Sundman solved the 3-body problem of the celestial mechanics. We do not give here, however, any explicit regularizing transformation, since there might be a "much better" one than the obvious one.

In closing, we refer to a remarkable talk delivered by Misner at the Cincinnati Conference.[2] Misner specializes the Arnowitt, Deser, and Misner formalism to type IX spaces in order to obtain the Einstein equations for

$$ds^2 = dt^2 - (A\omega^1)^2 - (B\omega^2)^2 - (C\omega^3)^2.$$

He writes the field equations in Hamiltonian form, introduces a new parameter instead of time, and discusses the singularities of the model. Our approach is similar. It is based on ideas of Weyl[3] and Gödel,[1] designed for spatially homogeneous spaces, and we think it is simpler. Concerning the introduction of a new parameter instead of the time, we follow Sundman,[4] as explained.

2. PRELIMINARIES

We consider the Lie group $M_4 = R \times G_3$, where R is the real line and G_3 is a 3-dimensional Lie group. Denoting by t the coordinate on R, we can introduce the vector fields

$$X_0 = \frac{\partial}{\partial t}, \quad X_a, \quad a = 1, 2, 3, \tag{2.1}$$

and the 1-forms

$$\omega^0 = dt, \quad \omega^a, \quad a = 1, 2, 3, \tag{2.2}$$

such that

$$\omega^\alpha(X_\beta) = \delta^\alpha{}_\beta, \quad \alpha, \beta \cdots 0, 1, 2, 3, \tag{2.3}$$

and X_a and ω^a are invariant under the left translations of G_3. One knows that the left-invariant vector fields of M_4 form a Lie algebra; that is,

$$[X_0, X_a] = 0, \quad a = 1, 2, 3, \quad [X_a, X_b] = C^f{}_{ab} X_f \tag{2.4}$$

or

$$d\omega^0 = 0, \quad d\omega^a = -\tfrac{1}{2} C^a{}_{bc} \omega^b \wedge \omega^c, \tag{2.5}$$

where $C^a{}_{bc}$ are the components of the structure constant tensor of the Lie algebra of G_3 with respect to the base (2.1) and (2.2). The C's satisfy

$$C^a{}_{bc} = -C^a{}_{cb} \tag{2.6}$$

and the Jacobi identities

$$C^a{}_{fb} C^f{}_{cd} + C^a{}_{fc} C^f{}_{db} + C^a{}_{fd} C^f{}_{bc} = 0. \tag{2.7}$$

We use X_α and ω^α to span the tensor algebra over M_4, that is, we specify tensor fields by giving their components with respect to these bases. We introduce on M_4 a connection by

$$\nabla_{X_\alpha} X_\beta = \Gamma_{\alpha\beta}{}^\gamma X_\gamma, \tag{2.8}$$

where $\Gamma_{\alpha\beta}{}^\gamma$ are the components of the connection with respect to (2.1). The curvature tensor field of the connection is given by

$$R(U, V)Y = \nabla_U \nabla_V Y - \nabla_V \nabla_U Y - \nabla_{[U,V]} Y,$$

and from that we have

$$R(X_\gamma, X_\delta)X_\beta = \nabla_{X_\gamma} \nabla_{X_\delta} X_\beta - \nabla_{X_\delta} \nabla_{X_\gamma} X_\beta - \nabla_{[X_\gamma, X_\delta]} X_\beta = R^\alpha{}_{\beta\gamma\delta} X_\alpha.$$

Following the roles of the covariant differentiation, we compute that

$$R^\alpha{}_{\beta\gamma\delta} = \Gamma_{\gamma\sigma}{}^\alpha \Gamma_{\delta\beta}{}^\sigma - \Gamma_{\delta\sigma}{}^\alpha \Gamma_{\gamma\beta}{}^\sigma - \Gamma_{\sigma\beta}{}^\alpha C^\sigma{}_{\gamma\delta} + X_\gamma \Gamma_{\delta\beta}{}^\alpha - X_\delta \Gamma_{\gamma\beta}{}^\alpha. \tag{2.9}$$

By requiring that the torsion tensor field

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$$

vanishes, that is,

$$\nabla_{X_\alpha} X_\beta - \nabla_{X_\beta} X_\alpha - [X_\alpha, X_\beta] = 0,$$

we obtain the following symmetry properties for the Γ's:

$$\Gamma_{\alpha\beta}{}^{\gamma} - \Gamma_{\beta\alpha}{}^{\gamma} = C^{\gamma}{}_{\alpha\beta}. \tag{2.10}$$

Using (2.10) we compute the components of the Ricci tensor field:

$$\begin{aligned} R_{\beta\gamma} &= R^{\alpha}{}_{\beta\gamma\alpha} \\ &= \Gamma_{\psi\beta}{}^{\phi}\Gamma_{\phi\gamma}{}^{\psi} - \Gamma_{\psi\phi}{}^{\psi}\Gamma_{\beta\gamma}{}^{\phi} + X_{\beta}\Gamma_{\psi\gamma}{}^{\psi} - X_{\psi}\Gamma_{\beta\gamma}{}^{\psi}. \end{aligned} \tag{2.11}$$

Introducing on M_4 a metric by the requirement that

$$g(X_\alpha, X_\beta) = \begin{bmatrix} 1 & 0 \\ 0 & \gamma_{ab}(t) \end{bmatrix} = \gamma_{\alpha\beta}(t), \quad \alpha, \beta = 0, 1, 2, 3, \tag{2.12}$$

that is,

$$ds^2 = dt^2 + \gamma_{ab}(t)\omega^a\omega^b, \quad a, b = 1, 2, 3, \tag{2.13}$$

M_4 becomes a pseudo-Riemannian space. Equation (2.13) excludes the possibility of lightlike t, but it is at our disposal to choose it timelike or spacelike. The metric (2.13) is left invariant under the transformations of G_3, but not invariant under M_4. We call these metrics spatially homogeneous, meaning that there exist global 3-dimensional hypersurfaces generated by G_3. The name also suggests that these hypersurfaces are spacelike. We make some remarks regarding this point later.

The requirement that the metric (2.13) should be covariant constant with respect to the connection (2.8), combined with the requirement (2.10), leads to the following equations:

$$\begin{aligned} &2g(\nabla_{X_\alpha} X_\beta, X_\gamma) \\ &\quad = X_\alpha g(X_\beta, X_\gamma) + X_\beta g(X_\gamma, X_\alpha) - X_\gamma g(X_\alpha, X_\beta) \\ &\qquad + g(X_\beta, [X_\gamma, X_\alpha]) + g(X_\gamma, [X_\alpha, X_\beta]) \\ &\qquad - g(X_\alpha, [X_\beta, X_\gamma]), \end{aligned}$$

that is,

$$\begin{aligned} \Gamma_{\alpha\beta\gamma} &\equiv \Gamma_{\alpha\beta}{}^{\sigma}\gamma_{\sigma\gamma} \\ &= \tfrac{1}{2}(X_\alpha\gamma_{\beta\gamma} + X_\beta\gamma_{\gamma\alpha} - X_\gamma\gamma_{\alpha\beta}) \\ &\quad + \tfrac{1}{2}(C_{\beta\gamma\alpha} + C_{\gamma\alpha\beta} - C_{\alpha\beta\gamma}), \end{aligned} \tag{2.14}$$

where

$$C_{\alpha\beta\gamma} = \gamma_{\alpha\sigma}C^{\sigma}{}_{\beta\gamma}. \tag{2.15}$$

Using (2.14), we observe that

$$X_a\gamma_{\alpha\beta} = 0, \quad a = 1, 2, 3, \tag{2.16}$$

and introduce the notation

$$X_0 f = \dot{f}, \tag{2.17}$$

where f is a function over M_4; we can write (2.8) explicitly with the γ's as

$$\begin{aligned} &\nabla_{X_0}X_0 = 0, \qquad \nabla_{X_0}X_a = K_a{}^bX_b, \\ &\nabla_{X_a}X_0 = K_a{}^bX_b, \quad \nabla_{X_a}X_b = -\tfrac{1}{2}\dot{\gamma}_{ab}X_0 + \Gamma_{ab}{}^cX_c, \end{aligned} \tag{2.18}$$

where

$$K_a{}^b = \tfrac{1}{2}\dot{\gamma}_{af}\gamma^{fb} \tag{2.19}$$

and

$$\Gamma_{ab}{}^c = \gamma^{cf}\tfrac{1}{2}(C_{bfa} + C_{fab} - C_{abf}). \tag{2.20}$$

We now substitute (2.18) into (2.11) and find that the components of the Ricci tensor field are given by

$$R_{00} = (K_f{}^f)^{\cdot} + K_f{}^gK_g{}^f, \tag{2.21}$$

$$R_{a0} = K_g{}^f\Gamma_{fa}{}^g - K_a{}^f\Gamma_{gf}{}^g \equiv K_g{}^fC^g{}_{fa} - K_a{}^fC^g{}_{gf}, \tag{2.22}$$

$$R_{ab} = \tfrac{1}{2}\ddot{\gamma}_{ab} - K_a{}^f\dot{\gamma}_{fb} + \tfrac{1}{2}\dot{\gamma}_{ab}(K_f{}^f) + R^*{}_{ab}, \tag{2.23}$$

where

$$R^*{}_{ab} = \Gamma_{fa}{}^g\Gamma_{gb}{}^f - \Gamma_{fg}{}^f\Gamma_{ab}{}^g \tag{2.24}$$

is the Ricci tensor field of the group space G_3. These expressions have been calculated by Taub[5] and Heckmann and Schücking.[6] Using the identity

$$\tfrac{1}{2}\gamma_{af}\ddot{\gamma}^{fb} = (K_a{}^b)^{\cdot} + 2K_a{}^fK_f{}^b, \tag{2.25}$$

we can compute the Ricci scalar R,

$$\tfrac{1}{2}R = (K_f{}^f)^{\cdot} + \tfrac{1}{2}[K_f{}^gK_g{}^f + (K_f{}^f)^2] + \tfrac{1}{2}R^*, \tag{2.26}$$

where

$$R^* = \gamma^{ab}R^*{}_{ab} \tag{2.27}$$

is the Ricci scalar of the group spaces G_3.

We consider Einstein's field equations with incoherent matter written in the form

$$R_{\alpha\beta} - \tfrac{1}{2}R\gamma_{\alpha\beta} = -\kappa\rho u_\alpha u_\beta, \tag{2.28}$$

$$u_\alpha u^\alpha = 1 \tag{2.29}$$

for the spaces (2.13), where

$$U = u^\alpha X_\alpha = uX_0 + u^aX_a \tag{2.30}$$

and

$$\mu = u_\alpha\omega^\alpha = u\,dt + u_a\omega^a \tag{2.31}$$

are the velocity vector field of the matter and the corresponding 1-form, respectively. The normalization (2.29) chooses the t lines to be timelike. One has to make this choice for physical and not for mathematical reasons. Choosing, instead of (2.29), the normalization

$$u_\alpha u^\alpha = -1, \tag{2.32}$$

we could extend our future discussions to spacelike t lines. The corresponding solutions, however, would represent stationary world models, where the density of the matter is a function of the spacelike coordinate t only. One is not looking systematically for such models without having a special reason. We would like to remark, however, that there are interesting special solutions for spacelike t lines in the case of the Bianchi Type VIII group if we include a nonvanishing Λ term into our discussions. These solutions are given by the line elements

$$\begin{aligned} ds^2 = dt^2 &+ (1-k)(\omega^1 \cos \beta t + \omega^2 \sin \beta t)^2 \\ &+ (1+k)(-\omega^1 \sin \beta t + \omega^2 \cos \beta t)^2 \\ &- (1+2k^2)(\omega^3)^2, \end{aligned} \tag{2.33}$$

where

$$\beta = \left(\frac{1-2k^2}{2(1+2k^2)}\right)^{\frac{1}{2}}, \quad \tfrac{1}{2} < |k| \le \frac{1}{2^{\frac{1}{2}}},$$

k a real parameter,

and

$$ds^2 = dt^2 + \tfrac{1}{2}(1+s)(\omega^1)^2 + \tfrac{1}{2}(1-s)(\omega^2)^2 - (\omega^3)^2, \tag{2.34}$$

where

$$|s| < 1, \quad s \text{ a real parameter,}$$

and

$$d\omega^1 = \omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2, \tag{2.35}$$

or in a special coordinate system

$$\begin{aligned} \omega^1 &= \cos x^3\, dx^1 = e^{x^1} \sin x^3\, dx^2, \\ \omega^2 &= -\sin x^3\, dx^1 + e^{x^1} \cos x^3\, dx^2, \\ \omega^3 &= e^{x^1}\, dx^2 + dx^3. \end{aligned} \tag{2.36}$$

Equations (2.33) and (2.24) are the Class II and Class III universes discovered by the author.[7] Equation (2.34) contains the famous Gödel cosmos[8] as a special case for $s = 0$. The speciality of (2.33) and (2.34) is that they are invariant under a 4-dimensional Lie group containing (2.35) as an invariant subgroup. As a consequence of that, the density of the matter is constant.

Coming back to our main line of reasoning, we list a few formulas for later use. The components of the tensor fields $\nabla_X U$ and $\nabla_X \mu(Y)$ are

$$u^\alpha{}_{:\beta} = X_\beta u^\alpha + \Gamma_{\beta\gamma}{}^\alpha u^\gamma \quad \text{and} \quad u_{\alpha:\beta} = X_\beta u_\alpha - \Gamma_{\beta\alpha}{}^\gamma u_\gamma, \tag{2.37}$$

as one easily sees following the roles of the covariant differentiation. All our subsequent formulas are consequences of (2.37). The equations of geodesic motion and the continuity equations are

$$u\dot{u} - \tfrac{1}{2}\dot{\gamma}_{fg} u^f u^g = 0, \tag{2.38}$$

$$u u_a - C^g{}_{fa} u^f u_g = 0, \tag{2.39}$$

$$(\rho u)^{\cdot} + \rho u (K_f{}^f) + \rho C^g{}_{gf} u^f = 0. \tag{2.40}$$

For later references, we write (2.29) as

$$(u)^2 + u_f u^f = 1. \tag{2.41}$$

Equation (2.38) is a consequence of (2.39) and (2.40). The components of the tensor of rotation are

$$\omega_{\alpha\beta} = \tfrac{1}{2}(X_\alpha u_\beta - X_\beta u_\alpha) - \tfrac{1}{2} C^\gamma{}_{\alpha\beta} u_\gamma, \tag{2.42}$$

that is,

$$\omega_{ab} = -\tfrac{1}{2} C^f{}_{ab} u_f, \quad \omega_{a0} = -\tfrac{1}{2}\dot{u}_a. \tag{2.43}$$

3. VARIATIONAL PRINCIPLE FOR VACUUM

Since the vacuum case already contains some essential features of our problem, for the sake of simplicity we start with this case. We write the vacuum field equations using (2.21), (2.22), (2.23), (2.25), and (2.26) in the following form:

$$R_{00} - \tfrac{1}{2}R = \tfrac{1}{2}(K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*) = 0, \tag{3.1}$$

$$R_{a0} = K_g{}^f C^g{}_{fa} - K_a{}^f C^g{}_{gf} = 0, \tag{3.2}$$

$$\begin{aligned} R_a{}^b - \tfrac{1}{2}R\delta_a{}^b = (K_a{}^b)^{\cdot} &- (K_f{}^f)\delta_a{}^b + (K_f{}^f)K_a{}^b \\ &- \tfrac{1}{2}[K_f{}^g K_g{}^f + (K_f{}^f)^2]\delta_a{}^b \\ &+ R^*{}_a{}^b - \tfrac{1}{2}R^*\delta_a{}^b = 0. \end{aligned} \tag{3.3}$$

These are Taub's equations[5] written in a slightly different form. Weyl writes in his famous book (Ref. 3, p. 251) while calculating the static spherically symmetric field for vacuum: "Wir nutzen das Wirkungsprinzip zunächst nur teilweise aus, indem wir annehmen, dass bei der Variation die zugrunde gelegte Normalform des ds^2 nicht zerstört wird; ··· bei solcher eingeschränkten Verwendung genügt es, das Wirkungsintegral für jene Normalform zu berechnen." These ideas apply in our case word for word.

The normal form for ds^2 is given in our case by (2.13). The action integral for this normal form is

$$\int \tfrac{1}{2} R g^{\frac{1}{2}}\, dx = \int_{t_1}^{t_2} \tfrac{1}{2} R^0 \gamma^{\frac{1}{2}}\, dt \int_G \omega^1 \wedge \omega^2 \wedge \omega^3, \tag{3.4}$$

where $\frac{1}{2}R$ is given by (2.26) and

$$\gamma = |\det(\gamma_{ab})|. \tag{3.5}$$

The first integral in (3.4) is extended between two fixed values of t, the second one over G_3 if compact, or over a part of it if otherwise, giving a finite constant

C which we normalize later. The action integral is, therefore,

$$J = c\int_{t_1}^{t_2} \gamma^{\frac{1}{2}}\{(K_f{}^f)^{\cdot} + \tfrac{1}{2}[K_f{}^g K_g{}^f + (K_f{}^f)^2] + \tfrac{1}{2}R^*\}\,dt. \tag{3.6}$$

Using the identity

$$(\gamma^{\frac{1}{2}})^{\cdot} = \gamma^{\frac{1}{2}} K_f{}^f \tag{3.7}$$

and integrating by parts and normalizing $c = 2$, we have

$$J = \int_{t_1}^{t_2} L\,dt = \int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*]\,dt. \tag{3.8}$$

But varying (3.8) with respect to the γ's is precisely the requirement to consider variations, which leaves the normal form (2.13) unchanged.

We now prove that

$$\delta J = 0 \tag{3.9}$$

gives the six field equations (3.3) as Euler–Lagrange equations. The proof is a straightforward calculation. We assume, following Siegel,[4] that our variational problem has a solution, and we consider a family of functions

$$\gamma_{ab}(\alpha; t), \quad -1 < \alpha < 1, \tag{3.10}$$

such that

$$\gamma_{ab}(0; t) = \gamma_{ab}(t) \tag{3.11}$$

is the solution of our variational problem. We construct with these functions the integral

$$J(\alpha) = \int_{t_1}^{t_2} L[\gamma_{ab}(\alpha; t), \dot\gamma_{ab}(\alpha; t)]\,dt, \quad -1 < \alpha < 1. \tag{3.12}$$

As a consequence of our assumption, $J(\alpha)$ assumes its extremum at $\alpha = 0$ and, therefore,

$$J'\big|_{\alpha=0} \equiv \frac{dJ(\alpha)}{d\alpha}\bigg|_{\alpha=0} = 0 \tag{3.13}$$

is equivalent to (3.9).

Carrying out our calculations, we see that

$$\gamma_{ab}(\alpha; t_1) = a_{ab} = \text{const}, \quad \gamma_{ab}(\alpha; t_2) = b_{ab} = \text{const}, \tag{3.14}$$

and that

$$\gamma'_{ab}(0; t) \text{ is arbitrary.} \tag{3.15}$$

The formula for $dJ/d\alpha$ reads as

$$\frac{dJ}{d\alpha} = \int_{t_1}^{t_2} (\gamma)^{\frac{1}{2}\prime}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*]\,dt + 2\int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[(K_f{}^g)' K_g{}^f - (K_f{}^f)(K_g{}^g)']\,dt + \int_{t_1}^{t_2} \gamma^{\frac{1}{2}}(\gamma^{fg})' R^*_{fg}\,dt. \tag{3.16}$$

We know that we do not have to compute $(R^*_{ab})'$. In order to facilitate the calculations, we compute the expressions

$$(\gamma^{fg})' = -\gamma^{fa}\gamma'_{ab}\gamma^{bg}, \tag{3.17}$$

$$(\gamma^{\frac{1}{2}})' = \tfrac{1}{2}\gamma^{\frac{1}{2}}\gamma'_{ab}\gamma^{ab}, \tag{3.18}$$

$$(K_f{}^g)' = \tfrac{1}{2}\dot\gamma'_{fa}\gamma^{ag} - K_f{}^a\gamma^{bg}\gamma'_{ab}, \tag{3.19}$$

$$(K_f{}^f)' = \tfrac{1}{2}\dot\gamma'_{ab}\gamma^{ab} - \gamma^{af}K_f{}^b\gamma'_{ab}. \tag{3.20}$$

Substituting into (3.16), we obtain

$$\frac{dJ}{d\alpha} = \int_{t_1}^{t_2} \gamma^{\frac{1}{2}}(\gamma^{af}K_f{}^b - \gamma^{ab}K_f{}^f)\dot\gamma'_{ab}\,dt - \int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[2\gamma^{af}K_f{}^g K_g{}^b - 2\gamma^{af}K_f{}^b(K_g{}^g) - \tfrac{1}{2}\gamma^{ab}(K_f{}^g K_g{}^f - K_f{}^f)^2 + R^{*ab} - \tfrac{1}{2}\gamma^{ab}R^*]\gamma'_{ab}\,dt. \tag{3.21}$$

We now evaluate the first integral in (3.21). Integrating by parts and using (3.7) and (3.14), we obtain

$$\int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[\gamma^{af}K_f{}^b - \gamma^{ab}(K_f{}^f)]\dot\gamma'_{ab}\,dt = -\int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[\gamma^{af}(K_f{}^b)^{\cdot} - \gamma^{ab}(K_f{}^f)^{\cdot} - 2\gamma^{af}K_f{}^g K_g{}^b + 3\gamma^{af}K_f{}^b(K_g{}^g) - \gamma^{ab}(K_f{}^f)^2]\gamma'_{ab}\,dt. \tag{3.22}$$

Substituting (3.22) into (3.21), we obtain

$$\frac{dJ}{d\alpha} = -\int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[\gamma^{af}(K_f{}^b)^{\cdot} - \gamma^{ab}(K_f{}^f)^{\cdot} + \gamma^{af}K_f{}^b(K_g{}^g) - \tfrac{1}{2}\gamma^{ab}[K_f{}^g K_g{}^f + (K_f{}^f)^2] + R^{*ab} - \tfrac{1}{2}R^*\gamma^{ab}]\gamma'_{ab}\,dt. \tag{3.23}$$

Substituting $\alpha = 0$ and using (3.13) and (3.15), we find that

$$-\gamma^{\frac{1}{2}}\gamma^{ta}\{(K_a{}^b)^{\cdot} - (K_g{}^g)^{\cdot}\delta_a{}^b + K_a{}^b(K_g{}^g) - \tfrac{1}{2}[K_f{}^g K_g{}^f + (K_f{}^f)^2] + R^*{}_a{}^b - \tfrac{1}{2}R^*\delta_a{}^b\} = 0, \tag{3.24}$$

and a glance at (3.3) proves our assertion. The Lagrangian of the vacuum problem is, therefore,

$$L = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*]. \tag{3.25}$$

In order to obtain the Euler–Lagrange equations in Hamiltonian form, we introduce

$$P_{ab} = \partial L/\partial\dot\gamma_{ab} \tag{3.26}$$

and define the Hamiltonian function by

$$H = (\partial L/\partial\dot\gamma_{ab})\dot\gamma_{ab} - L. \tag{3.27}$$

Since L is homogeneous of degree two in $\dot\gamma_{ab}$, we find that the Hamiltonian function is given by

$$H = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*], \tag{3.28}$$

and the Euler–Lagrange equations in Hamiltonian form are

$$\dot{\gamma}_{ab} = \partial H/\partial P_{ab}, \quad \dot{P}_{ab} = -\partial H/\partial \gamma_{ab}. \tag{3.29}$$

Since H does not depend explicitly on t, (3.28) has a constant value h for a solution of (3.29), that is,

$$H = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*] = h \tag{3.30}$$

is the energy integral. A glance at (3.1) shows that

$$h = 0. \tag{3.31}$$

This is the seventh field equation. Equations (3.2) are integrals of the other equations as Taub proves in Ref. 5. Therefore, the Taub equations can be written as

$$\delta\int_{t_1}^{t_2} \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*]\, dt = 0, \tag{3.32}$$

$$K_f{}^g C^f{}_{ga} - K_a{}^f C^g{}_{gf} = 0, \tag{3.33}$$

$$H = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*] = 0. \tag{3.34}$$

4. VARIATIONAL PRINCIPLE FOR INCOHERENT MATTER

We now consider the Einstein field equations with incoherent matter. Using (2.21), (2.22), (2.23), (2.25), and (2.26), we find that

$$R_{00} - \tfrac{1}{2}R = [K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*] = -\kappa\rho(u)^2, \tag{4.1}$$

$$R_{a0} = K_f{}^g C^f{}_{ga} - K_a{}^f C^g{}_{gf} = -\kappa\rho u u_a, \tag{4.2}$$

$$R_a{}^b - \tfrac{1}{2}R\delta_a{}^b = (K_a{}^b)^{\cdot} - (K_f{}^f)^{\cdot}\delta_a{}^b + (K_f{}^f)K_a{}^b - \tfrac{1}{2}[K_f{}^g K_g{}^f + (K_f{}^f)^2]\delta_a{}^b, \tag{4.3}$$

$$R^*{}_a{}^b - \tfrac{1}{2}R^*\delta_a{}^b = -\kappa\rho u_a u^b.$$

These are the Heckmann–Schücking equations[6] written in a slightly different form. We have, in addition,

$$(u)^2 + u_a u^a = 1, \tag{4.4}$$

$$u\dot{u}_a = C^f{}_{ga} u^g u_f, \tag{4.5}$$

$$(\rho u)^{\cdot} + (\rho u)[K_f{}^f + (C^g{}_{gf}u^f)/(1 - u_a u^a)^{\frac{1}{2}}] = 0. \tag{4.6}$$

Our problem is now to find the Lagrangian for the Heckmann–Schücking equations. Examining (4.6), we discover that the term

$$C^g{}_{gf}u^f \tag{4.7}$$

contains the vector

$$C^g{}_{gf} \tag{4.8}$$

obtained by contraction over two indices from the structure constant tensor of G_3. The term (4.7) vanishes if (4.8) vanishes. It is natural, therefore, to divide the 3-dimensional Lie groups, their Lie algebras, that is, into two different classes according to the vanishing or nonvanishing of the vector (4.8). These classes are the following:

$$\text{Class I: } C^g{}_{ga} = 0, \quad a = 1, 2, 3, \tag{4.9}$$

$$\text{Class II: } C^g{}_{ga} \neq 0, \quad a = 1, 2, 3. \tag{4.10}$$

Class I contains the groups of the following Bianchi types:

Class I: Type I, II, VIII, and IX. (4.11)

The structure of the Class II algebras is given by

$$[X_1, X_2] = 0, \quad [X_A, X_3] = C^B{}_A X_B, \quad A, B, \cdots = 1, 2, \tag{4.12}$$

or, alternatively,

$$d\omega^1 = -C^1{}_A\omega^A \wedge \omega^3, \quad d\omega^2 = -C^2{}_A\omega^A \wedge \omega^3, \quad d\omega^3 = 0. \tag{4.13}$$

Therefore, the Class II algebras are given by the different normal forms of the 2×2 real matrices $C^B{}_A$ with nonvanishing trace. These are[9]

Class II: Type III, IV, V, VI, VII. (4.14)

We write our variational principle for the Class I groups only. We can integrate (4.6) in this case to

$$\rho u = l/\gamma^{\frac{1}{2}} \geq 0, \tag{4.15}$$

where l is a constant. Writing (4.4) as

$$u = (1 - u_a u^a)^{\frac{1}{2}}, \tag{4.16}$$

we obtain from (4.15)

$$\rho = l/[\gamma(1 - u_a u^a)]^{\frac{1}{2}} \tag{4.17}$$

and

$$\rho(u)^2 = l[(1 - u_a u^a)/\gamma]^{\frac{1}{2}}. \tag{4.18}$$

We now claim that the Lagrangian of the Heckmann–Schücking equations for Class I groups is given by

$$L = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*] - 2\kappa l(1 - u_a u^a)^{\frac{1}{2}}, \tag{4.19}$$

and the Hamiltonian is

$$H = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*] + 2\kappa l(1 - u_a u^a)^{\frac{1}{2}}. \tag{4.20}$$

We obtain the Heckmann–Schücking equations as

$$\delta\int_{t_1}^{t_2} \{\gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*] - 2\kappa l(1 - u_f u^f)^{\frac{1}{2}}\}\, dt = 0, \tag{4.21}$$

$$K_f{}^g C^f{}_{ga} = -\kappa(l/\gamma^{\frac{1}{2}})u_a, \quad a = 1, 2, 3, \tag{4.22}$$

$$H = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 - R^*] + 2\kappa l(1 - u_f u^f)^{\frac{1}{2}} = h = 0, \tag{4.23}$$

and

$$\dot{u}_a = \frac{C^f{}_{ga}u^g u_f}{(1 - u_f u^f)^{\frac{1}{2}}}, \quad a = 1, 2, 3, \quad C^g{}_{ga} = 0!. \quad (4.24)$$

To prove this assertion, we compute the derivative $dF/d\alpha$ of

$$F(\alpha) = -2\kappa l \int_{t_1}^{t_2} (1 - u_a u^a)^{\frac{1}{2}}\, dt \quad \text{for} \quad \alpha = 0.$$

We imagine that the functions $\gamma_{ab}(\alpha; t)$ are substituted for $\gamma_{ab}[t]$:

$$\frac{dF}{d\alpha} = -\kappa l \int_{t_1}^{t_2} \frac{u^a u^b}{(1 - u_f u^f)^{\frac{1}{2}}} \gamma'_{ab}\, dt. \quad (4.25)$$

Using our earlier results, we see that

$$-(\gamma^{\frac{1}{2}})[R^{ab} - \tfrac{1}{2}R\gamma^{ab}] - [\kappa l/(1 - u_f u^f)^{\frac{1}{2}}]u^a u^b = 0. \quad (4.26)$$

Dividing by $-\gamma^{\frac{1}{2}}$ and using (4.17), we find that (4.26) is equivalent to (4.3). It follows, exactly as before, that the Hamiltonian function (4.20) has to be constant for the solutions. Dividing (4.20) by $2\gamma^{\frac{1}{2}}$, using (4.18) and (4.1), we see that the energy constant has to be zero. Equations (4.22) and (4.24) are consistent with the other equations (see Ref. 6).

5. SOME REMARKS

Examining Eqs. (3.32)–(3.34), one sees that the variational principle has its full power in the vacuum case. One would treat the vacuum problem as a mechanical problem defined by the Lagrangian function

$$L = \gamma^{\frac{1}{2}}[K_f{}^g K_g{}^f - (K_f{}^f)^2 + R^*], \quad (5.1)$$

and reduce it with the help of the integrals

$$K_f{}^g C^f{}_{ga} - K_a{}^f C^g{}_{gf} = 0, \quad a = 1, 2, 3. \quad (5.2)$$

In case of dust, the situation is different. The Class I groups are preferred because (4.6) is then integrable. Furthermore, the power of (4.21)–(4.23) is limited by (4.24) in general. We see in the last section of this paper that, in the case of

$$ds^2 = dt^2 + A(\omega^1)^2 + B(\omega^2)^2 + C(\omega^3)^2 + 2D\omega^2\omega^3, \quad (5.3)$$

(4.24) is trivial and the method retains its full power and simplicity. In order to obtain the *general* Type VIII and IX models (the Type I and II models do not have rotation), one develops the above formulas for the line element

$$ds^2 = (dt + p_f\omega^f)^2 + \gamma_{ab}(t)\omega^a\omega^b, \qquad a, b, f, \cdots = 1, 2, 3, \quad (5.4)$$

with

$$\dot{p}_a = 0, \quad a = 1, 2, 3, \quad (5.5)$$

which replaces (4.24) as the geodesic condition. The expressions corresponding to (4.21)–(4.23) are naturally more involved.

Another remark refers to the Bianchi Type II group given by

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = 0, \quad d\omega^3 = 0. \quad (5.6)$$

We claim that there is no rotating solution for (5.6). The proof is trivial. From (4.22), we obtain

$$K_f{}^g C^f{}_{g1} = 0 = \kappa(l/\gamma^{\frac{1}{2}})u_1,$$

that is,

$$u_1 = 0. \quad (5.7)$$

Due to the special form of the structure constant tensor, it follows that

$$C^f{}_{ab}u_f \equiv C^1{}_{ab}u_1 = 0. \quad (5.8)$$

Using (4.24) and (5.8), one sees that

$$\dot{u}_a = 0, \quad a = 1, 2, 3. \quad (5.9)$$

Since the components of the rotation tensor are given by

$$\omega_{ac} = -\tfrac{1}{2}C^f{}_{ab}u_f, \quad \omega_{a0} = -\tfrac{1}{2}\dot{u}_a$$

[see (2.43)], we obtain

$$\omega_{ab} = 0, \quad \omega_{a0} = 0, \quad (5.10)$$

as claimed. Rotating solution for Class I groups is, therefore, possible only with Bianchi Type VIII and IX groups. We now go over to more serious applications.

6. A CLASS OF SOLUTIONS

We consider the line element

$$ds^2 = dt^2 - (A\omega^1)^2 - (B\omega^2)^2 - (B\omega^3)^2 \quad (6.1)$$

with each of the four different Class I groups; that is, with

$$d\omega^1 = 0, \quad d\omega^2 = 0, \quad d\omega^3 = 0, \quad (6.2)$$

or

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = 0, \quad d\omega^3 = 0, \quad (6.3)$$

or

$$d\omega^1 = \omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2, \quad (6.4)$$

or

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2, \quad (6.5)$$

respectively. It is *not known* by me whether or not the cases with (6.3) and (6.4) are in the literature, but (6.1) with (6.5) has been discussed. With vacuum this is the Taub solution and with incoherent matter it is discussed by Behr.[10] The case (6.1) with (6.2) is fully integrated by Schücking. Our aim is to compute the Hamiltonian function of these cases and write the field equations in Hamiltonian form and *regularize* these equations, in the way that Sundman regularized the 3-body problem of the celestial mechanics. The idea is as follows: One introduces by a suitable transformation

$$t = t(\tau) \quad \text{or} \quad \tau = \tau(t),$$

a new independent variable τ such that our field equations, as a system of first order ordinary differential equations with respect to τ, should be *analytic.* A system

$$x'_\kappa = f_\kappa(x_l), \quad \kappa, l, \cdots = 1, 2, \cdots, n,$$

is called analytic if the functions f_κ as functions of x_l are analytic. Having done that successfully, one considers the problem solved since everything else can be done by computers and by the application of the qualitative theory of differential equations.[11]

To demonstrate all of this by an example, we proceed with our problem. From (4.22), it follows that

$$u = 1, \quad u_a = 0, \quad a = 1, 2, 3, \tag{6.6}$$

and Eqs. (4.24) are trivially satisfied. The Lagrangian and Hamiltonian functions read as

$$L = -4\dot{A}\dot{B}B - 2A\dot{B}^2 + AB^2R^* - 2\kappa l \tag{6.7}$$

and

$$H = -4\dot{A}\dot{B}B - 2A\dot{B}^2 - AB^2R^* + 2\kappa l, \tag{6.8}$$

respectively, where the relevant Ricci scalars are

for Eq. (6.2)

$$R^* = 0, \tag{6.9}$$

for Eq. (6.3)

$$R^* = -A^2/2B^4, \tag{6.10}$$

for Eq. (6.4)

$$R^* = -(A^2 + 4B^2)/2B^4, \tag{6.11}$$

for Eq. (6.5)

$$R^* = -(A^2 - 4B^2)/2B^4. \tag{6.12}$$

We treat the case (6.12), that is, (6.1) with (6.5). The other cases can be obtained by making suitable changes.

The Hamiltonian function for (6.1) with (6.5) is given by

$$H = -4\dot{A}\dot{B}B - 2A\dot{B}^2 + (A^3/2B^2) - 2A + 2\kappa l. \tag{6.13}$$

We write the field equations in Hamiltonian form, that is, we define

$$P = \frac{\partial L}{\partial \dot{A}} = -4B\dot{B}, \quad Q = \frac{\partial L}{\partial \dot{B}} = -4\dot{A}B - 4A\dot{B}. \tag{6.14}$$

Solving (6.14) for $\dot{A}$ and $\dot{B}$, we obtain

$$\dot{A} = (AP/4B^2) - (Q/4B), \quad \dot{B} = -P/4B. \tag{6.15}$$

Substituting into (6.13), we obtain

$$H = (AP^2/8B^2) - (PQ/4B) + (A^3/2B^2) - 2A + 2\kappa l, \tag{6.16}$$

and the field equations in Hamiltonian form are

$$\dot{A} = \partial H/\partial P, \quad \dot{B} = \partial H/\partial Q, \quad \dot{P} = -\partial H/\partial A, \quad \dot{Q} = -\partial H/\partial B. \tag{6.17}$$

The first two equations are (6.15), and the second two are given by

$$\dot{P} = -(P^2/8B^2) - (3A^2/2B^2) + 2,$$
$$\dot{Q} = (AP^2/4B^3) - (PQ/4B^2) + A^3/B^3. \tag{6.18}$$

The energy integral reads as

$$H \equiv (AP^2/8B^2) - (PQ/4B) + (A^3/2B^2) - 2A + 2\kappa l = 0. \tag{6.19}$$

The form of (6.15) and (6.18) strongly suggests the introduction of the new variables

$$x = A/B, \quad y = P/B, \quad z = Q/B, \tag{6.20}$$

or

$$A = xB, \quad P = yB, \quad Q = zB. \tag{6.21}$$

Then, the equations read as

$$\dot{x}B = \tfrac{1}{2}xy - \tfrac{1}{4}z, \tag{6.22}$$

$$\dot{B}B = -\tfrac{1}{4}yB, \tag{6.23}$$

$$\dot{y}B = \tfrac{1}{8}y^2 - \tfrac{3}{2}x^2 + 2, \tag{6.24}$$

$$\dot{z}B = x(x^2 + \tfrac{1}{4}y^2), \tag{6.25}$$

and

$$B[\tfrac{1}{8}xy^2 - \tfrac{1}{4}yz + \tfrac{1}{2}x^3 - 2x] + 2\kappa l = 0. \tag{6.26}$$

Introducing a new independent variable τ by

$$\frac{df}{d\tau} \equiv f' = \dot{f}B, \tag{6.27}$$

our equations become

$$x' = \tfrac{1}{2}xy - \tfrac{1}{4}z, \tag{6.28}$$

$$B' = -\tfrac{1}{4}yB, \tag{6.29}$$

$$y' = \tfrac{1}{8}y^2 - \tfrac{3}{2}x^2 + 2, \tag{6.30}$$

$$z' = x(x^2 + \tfrac{1}{4}y^2), \tag{6.31}$$

where Eq. (6.29) is a consequence of (6.26), (6.28), (6.30), and (6.31). A more elegant way to do this would be to reverse the order of the operations. One should carry out the time transformation first with the help of a canonical transformation (see Ref. 4, 35). It is then obvious that one retains an energy integral and, therefore, one can leave (6.29) aside. And, as a second step, one would go over to the ratios.

We want to give our method for solving the problem: We integrate the *analytic* system

$$x' = \tfrac{1}{2}xy - \tfrac{1}{4}z, \quad y' = \tfrac{1}{8}y^2 - \tfrac{3}{2}x^2 + 2, \quad z' = x(x^2 + \tfrac{1}{4}y^2); \tag{6.32}$$

we compute B from (6.26)

$$B = 4\kappa l/(-\tfrac{1}{4}xy^2 + \tfrac{1}{2}yz - x^3 + 4x); \tag{6.33}$$

we compute A from

$$A = xB. \tag{6.34}$$

The cosmic time t is computed from

$$t = \int_{\tau_0}^{\tau} \frac{4\kappa l}{-\tfrac{1}{4}xy^2 + \tfrac{1}{2}yz - x^3 + 4x}\, d\sigma. \tag{6.35}$$

This method solves our problem in a way similar to Sundman's solution of the 3-body problem of the celestial mechanics.

The best way of visualizing (6.32) is to go into a 3-dimensional Euclidean space E_3 with the coordinates x, y, and z. The right-hand sides of (6.32) are the components of an analytic vector field V over E_3. V is nowhere singular; that is, the components of V vanish nowhere on E_3 simultaneously.

Through any point $P_0 = (x_0, y_0, z_0)$ of E_3, one can draw with the help of a computer one and only one integral curve C of (6.32). We then find, corresponding to each such line, a universe following the rest of our method. Singularities of the universe occur, for example, where C goes through the yz plane; that is, $x = 0$ since the geometrical meaning of x is the ratio of the axes of the universe. To find the answers to the arising questions, one should study Nemytskii and Stepanov.[11] Some numerical calculations will be made in a later paper. As a curiosity, we compute the Friedmann cosmos. Assuming

$$x = 1, \tag{6.36}$$

then Eqs. (6.32) reduce to

$$z = 2y, \tag{6.37}$$

$$y' = \tfrac{1}{2}(1 + \tfrac{1}{4}y^2), \tag{6.38}$$

which integrate to

$$y = 2tg\tfrac{1}{4}(\tau - \tau_0). \tag{6.39}$$

Then, (6.33) reads

$$B = \tfrac{4}{3}\kappa l \cos^2 \tfrac{1}{4}(\tau - \tau_0) \tag{6.40}$$

and

$$t = \tfrac{4}{3}\kappa l[\tfrac{1}{2}(\tau - \tau_0) + \sin \tfrac{1}{2}(\tau - \tau_0)]. \tag{6.41}$$

We now list the equations for the other cases:

(6.1) with (6.2),

$$x' = \tfrac{1}{2}xy - \tfrac{1}{4}z, \quad y' = \tfrac{1}{8}y^2, \quad z' = \tfrac{1}{4}xy^2,$$
$$B = 16\kappa l/[y(2z - xy)], \quad A = xB, \quad t = \int_{\tau_0}^{\tau} B\, d\sigma; \tag{6.42}$$

(6.1) with (6.3),

$$x' = \tfrac{1}{2}xy - \tfrac{1}{4}z, \quad y' = \tfrac{1}{8}y^2 - \tfrac{3}{2}x^2, \quad z' = x(x^2 + \tfrac{1}{4}y^2),$$
$$B = \frac{4\kappa l}{-\tfrac{1}{4}xy^2 + \tfrac{1}{2}yz - x^3}, \quad A = xB, \quad t = \int_{\tau_0}^{\tau} B\, d\sigma; \tag{6.43}$$

(6.1) with (6.4),

$$x' = \tfrac{1}{2}xy - \tfrac{1}{4}z, \quad y' = \tfrac{1}{8}y^2 - \tfrac{3}{2}x^2 - 2, \quad z' = x(x^2 + \tfrac{1}{4}y^2),$$
$$B = \frac{4\kappa l}{-\tfrac{1}{4}xy^2 + \tfrac{1}{2}yz - x^3 - 4x}, \quad A = xB, \tag{6.44}$$
$$t = \int_{\tau_0}^{\tau} B\, d\sigma.$$

As a curiosity, we remark that (6.42) can be integrated in closed form:

$$x = \frac{\alpha}{3(\tau_0 - \tau)^3} + \beta, \quad y = \frac{8}{\tau_0 - \tau},$$
$$z = \frac{4}{\tau_0 - \tau}\left[\frac{\alpha}{3(\tau_0 - \tau)^3} + 4\beta\right], \tag{6.45}$$
$$B = \frac{\kappa l}{12\beta}(\tau_0 - \tau)^2, \quad A = xB, \quad t = -\frac{\kappa l}{36\beta}(\tau_0 - \tau)^3,$$

where τ_0, α, and β are constants of integration. The corresponding solution is a Schücking solution.

7. THE ROTATING UNIVERSES

There is a challenging problem in the relativistic cosmology: Consider the line element

$$ds^2 = dt^2 + A(\omega^1)^2 + B(\omega^2)^2 + C(\omega^3)^2 + 2D\omega^2\omega^3, \tag{7.1}$$

where A, B, C, and D are functions of t only and the differential forms ω^1, ω^2, and ω^3 satisfy either

$$d\omega^1 = \omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2 \tag{7.2}$$

or

$$d\omega^1 = -\omega^2 \wedge \omega^3, \quad d\omega^2 = -\omega^3 \wedge \omega^1, \quad d\omega^3 = -\omega^1 \wedge \omega^2. \tag{7.3}$$

We find A, B, C, and D such that (7.1) satisfies Einstein's field equations with incoherent matter and such that the expansion and the rotation of the matter do not vanish.

This problem is challenging not because the astronomers discovered the rotation of the universe, but because (7.1) with (7.3) probably gives the simplest finite model where the "Weltsubstrat" has the most general motion, namely, translation, rotation, expansion, and shear. For the sake of definiteness, we restrict ourselves to (7.1) with (7.3). We see that

$$\gamma_{ab} = \begin{pmatrix} A & 0 & 0 \\ 0 & B & D \\ 0 & D & C \end{pmatrix},$$

$$\gamma^{ab} = \begin{pmatrix} \frac{1}{A} & 0 & 0 \\ 0 & \frac{C}{BC - D^2} & -\frac{D}{BC - D^2} \\ 0 & -\frac{D}{BC - D^2} & \frac{B}{BC - D^2} \end{pmatrix}, \quad (7.4)$$

and, therefore,

$$K_a{}^b = \begin{pmatrix} \frac{\dot{A}}{2A} & 0 & 0 \\ 0 & \frac{\dot{B}C - D\dot{D}}{2(BC - D^2)} & \frac{B\dot{D} - \dot{B}D}{2(BC - D^2)} \\ 0 & \frac{C\dot{D} - \dot{C}D}{2(BC - D^2)} & \frac{B\dot{C} - D\dot{D}}{2(BC - D^2)} \end{pmatrix}, \quad (7.5)$$

where a is the row and b is the column index. From (4.22) and (7.5), it follows that

$$\frac{(B - C)\dot{D} - (\dot{B} - \dot{C})D}{2(BC - D^2)} = -\kappa \frac{l}{\gamma^{\frac{1}{2}}} u_1,$$

$$u_2 = 0, \quad u_3 = 0, \quad (7.6)$$

and from (4.24), we obtain

$$\dot{u}_1 = 0. \quad (7.7)$$

Therefore,

$$u_\alpha = \left[\left(1 - \frac{V^2}{A}\right)^{\frac{1}{2}}, V, 0, 0\right], \quad (7.8)$$

where V is a constant. One might mention that the only nonvanishing component of the rotation tensor [see (2.43)] is given by

$$\omega_{23} = -\tfrac{1}{2}V, \quad (7.9)$$

and the length of the vector of rotation W defined by

$$W^\alpha = \tfrac{1}{2}\eta^{\alpha\beta\gamma\delta}\omega_{\beta\gamma}u_\delta \quad (7.10)$$

is given by

$$g(W, W) = A(\tfrac{1}{2}V)^2. \quad (7.11)$$

We write (7.6) for later references in the following form:

$$A[(B - C)\dot{D} - (\dot{B} - \dot{C})D] = 2\kappa l V \gamma^{\frac{1}{2}}, \quad (7.12)$$

where

$$\gamma = |A(BC - D^2)|. \quad (7.13)$$

One easily computes that

$$L = \gamma^{\frac{1}{2}}\left(-\frac{\dot{A}(BC - D^2)^{\cdot}}{2A(BC - D^2)} - \frac{A(\dot{B}\dot{C} - \dot{D}^2)}{2A(BC - D^2)} + R^*\right) - 2\kappa l\left(1 - \frac{V^2}{A}\right)^{\frac{1}{2}}. \quad (7.14)$$

The Ricci scalar of the group space is given by

$$R^* = \frac{2(A^2 + B^2 + C^2) - (A + B + C)^2 + 4D^2}{2A(BC - D^2)}. \quad (7.15)$$

There are two problems: (a) reduction of the mechanical system, defined by the Lagrangian function (7.14), with the help of the first integral (7.12); (b) regularization of the reduced system.

A. Reduction of the Mechanical System

The reduction of a system is a standard problem in the mechanics, and we found its solution following standard methods. Therefore, we give the results only. Consider the functions x, y, z, and w defined by

$$x = -A, \quad y = -\tfrac{1}{2}\{B + C + [4D^2 + (B - C)^2]^{\frac{1}{2}}\},$$
$$z = -\tfrac{1}{2}\{B + C - [4D^2 + (B - C)^2]^{\frac{1}{2}}\}, \quad (7.16)$$
$$w = \arctan(B - C/2D),$$

or the inverse transformations

$$A = -x, \quad B = \tfrac{1}{2}(y + z) - \tfrac{1}{2}(y - z)\sin w,$$
$$C = \tfrac{1}{2}(y + z) + \tfrac{1}{2}(y - z)\sin w, \quad (7.17)$$
$$D = -\tfrac{1}{2}(y - z)\cos w.$$

We show that (7.17) reduces our system defined by (7.14) and (7.12). We first compute the new form of (7.12). One sees that

$$yz = BC - D^2; \quad (7.18)$$

therefore, $xyz = -A(BC - D^2)$ and

$$\gamma^{\frac{1}{2}} = (xyz)^{\frac{1}{2}}. \quad (7.19)$$

One easily computes that

$$A[(B - C)\dot{D} - (\dot{B} - \dot{C})] = \tfrac{1}{2}x(y - z)^2\dot{w};$$

therefore, (7.12) reads as

$$\dot{w} = \frac{4\kappa l V}{x(y - z)^2}(xyz)^{\frac{1}{2}}. \quad (7.20)$$

It does not contain w! We now compute the Lagrangian function (7.14) in the new variables and find that it does not contain w. Using (7.18), we see that

$$-\frac{\dot{A}}{2A}\frac{(BC - D^2)^{\cdot}}{BC - D^2} = -\frac{1}{2}\frac{\dot{x}}{x}\left(\frac{\dot{y}}{y} + \frac{\dot{z}}{z}\right). \quad (7.21)$$

A straightforward calculation shows that

$$\dot{B}\dot{C} - \dot{D}^2 = \dot{y}\dot{z} - \left(\frac{y-z}{z}\right)^2(\dot{w})^2 = \dot{y}\dot{z} - \frac{4\kappa^2 l^2 V^2}{x(y-z)^2}yz;$$

therefore,

$$-\frac{\dot{A}}{2A}\frac{(\dot{B}\dot{C} - \dot{D}^2)}{(BC - D^2)} = -\frac{1}{2}\frac{\dot{y}}{y}\frac{\dot{z}}{z} + \frac{2\kappa^2 l^2 V^2}{x(y-z)^2}. \quad (7.22)$$

Another trivial calculation shows that

$$R^* = -\frac{2(x^2 + y^2 + z^2) - (x + y + z)^2}{2xyz}. \quad (7.23)$$

From (7.14), (7.19), (7.21), (7.22), and (7.23) it then follows that

$$L = -(xyz)^{\frac{1}{2}}\left(\frac{1}{2}\frac{\dot{x}}{x}\frac{\dot{y}}{y} + \frac{1}{2}\frac{\dot{y}}{y}\frac{\dot{z}}{z} + \frac{1}{2}\frac{\dot{z}}{z}\frac{\dot{x}}{x} + \frac{2(x^2+y^2+z^2) - (x+y+z)^2}{2xyz} - \frac{2\kappa^2 l^2 V^2}{x(y-z)^2}\right) - 2\kappa l\left(1 + \frac{V^2}{x}\right)^{\frac{1}{2}}. \quad (7.24)$$

The Lagrangian function has been first computed by Gödel.[1] (Also, see Ref. 12.) Defining

$$p = \frac{\partial L}{\partial \dot{x}}, \quad q = \frac{\partial L}{\partial \dot{y}}, \quad r = \frac{\partial L}{\partial \dot{z}}, \quad (7.25)$$

we compute the Hamiltonian function

$$H = \frac{1}{2(xyz)^{\frac{1}{2}}}\{2[(xp) + (yq)^2 + (zr)^2] - (xp + yq + zr)^2 + 2(x^2 + y^2 + z^2) - (x + y + z)^2\} - \frac{2\kappa^2 l^2 V^2}{x(y-z)^2}(xyz)^{\frac{1}{2}} + 2\kappa l\left(1 + \frac{V^2}{x}\right)^{\frac{1}{2}}. \quad (7.26)$$

The field equations are

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{y} = \frac{\partial H}{\partial q}, \quad \dot{z} = \frac{\partial H}{\partial r},$$

$$\dot{p} = -\frac{\partial H}{\partial x}, \quad \dot{q} = -\frac{\partial H}{\partial y}, \quad \dot{r} = -\frac{\partial H}{\partial z}, \quad (7.27)$$

where H is given by (7.26).

The method for finding a rotating universe is as follows: Find a solution of (7.27) for which

$$H = 0. \quad (7.28)$$

Then we obtain w from (7.20) by integration. Compute the components of the metric from (7.17). Examining the form of (7.17), one sees that seeking a solution via the ansatz

$$\gamma_{ab} = \begin{pmatrix} a & 0 & 0 \\ 0 & b + c\sin\alpha & c\cos\alpha \\ 0 & c\cos\alpha & b - c\sin\alpha \end{pmatrix},$$

where a, b, c, and α are unknown functions of time, is a naive but well-founded approach. We now consider our second problem.

B. Regularization of the Reduced System

We formulate this problem as follows: Introduce a new independent variable by a suitable transformation of t such that (7.27) is transformed into an *analytic* system. Examining (7.26), one had the strong impression that *the problem of the rotating universe can be solved by regularization.* One sees several ways and one has several suggestions; the strongest one is probably to study Siegel's book.[4]

Note added in proof: I am indebted to G. F. R. Ellis for bringing to my attention the following two papers,

G. F. R. Ellis and M. A. H. MacCallum, Commun. Math. Phys. **12**, 108 (1969),

S. Hawking, Monthly Notices Roy. Astron. Soc. **142**, 129 (1969),

and for the remark that there are two additional groups of Class I, namely, a special Type VI and a special Type VII group characterized by

$$d\omega^1 = \omega^3 \wedge \omega^1, \quad d\omega^2 = \omega^2 \wedge \omega^3, \quad d\omega^3 = 0$$

and

$$d\omega^1 = \omega^2 \wedge \omega^3, \quad d\omega^2 = \omega^3 \wedge \omega^1, \quad d\omega^3 = 0,$$

respectively.

ACKNOWLEDGMENT

I am indebted to E. L. Schücking for clarifying discussions and to H. Armstrong for typing the manuscript.

* This work was supported by the Air Force Office of Scientific Research under Grant AF-AFOSR-903-67 and by the National Aeronautics and Space Administration under NASA Grant No. NGL 44-004-001.

[1] K. Gödel, Proc. Intern. Congr. Math. **1**, 175 (1952).
[2] C. W. Misner, lecture given at the Relativity Conference in the Midwest (Cincinnati, 1969).
[3] H. Weyl, *Raum Zeit Materie*, 5 Aufl. (Springer, Berlin 1923) [Wissenschaftliche Buchgesellschaft, Darmstadt (1961)].
[4] C. L. Siegel, *Vorlesungen über Himmelsmechanik* (Springer, Berlin, 1956).
[5] A. H. Taub, Ann. Math. **53**, 472 (1951).
[6] O. Heckmann and E. Schücking, in *Gravitation*, edited by L. Witten (Wiley, New York, 1962).
[7] I. Ozsváth, J. Math. Phys. **6**, 590 (1965).
[8] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).
[9] A. Z. Petrov, *Einstein Spaces* (Pergamon, New York, 1969).
[10] C. Behr, Z. Astrophys. **54**, 268 (1962).
[11] V. V. Nemytskii and V. V. Stepanov, *Qualitative Theory of Differential Equations* (Princeton U.P., Princeton, N.J., 1960).
[12] F. B. Estabrook, H. D. Wahlquist, and C. G. Behr, J. Math. Phys. **9**, 497 (1968).

Dust-Filled Universes of Class II and Class III*

István Ozsváth
Division of Mathematics, The University of Texas at Dallas, Dallas, Texas 75320

(Received 2 March 1970)

I construct on the Lie group $R \times H^3$ two different families of left-invariant metrics which satisfy the Einstein field equations with incoherent matter, calling the Riemannian spaces M_4, obtained this way, Class II and Class III universes. We discuss the geometry of these universes.

1. INTRODUCTION

Schücking and I discussed recently[1] the physical and the geometrical properties of the finite rotating universe, the Class I solution, according to the terminology introduced by Farnsworth and Kerr.[2] The Class I solution is a family of left-invariant metrics on the Lie group $R \times S^3$ satisfying Einstein's field equations with dust.

In this paper, I discuss in a similar manner the Class II and Class III universes, which are two different families of metrics imposed on the same manifold, namely, one the Lie group $R \times H^3$. The Class IV universes given in Ref. 3 receive their treatment in a subsequent paper. The four classes exhaust all the possibilities of homogeneous dust solutions of Einstein's field equations as Refs. 2 and 3 show.

In order to keep this paper readable independently of Ref. 1, I repeat some general remarks made there and suggest that the reader glance at Ref. 1 too.

2. USEFUL THEOREMS AND FORMULAS

As a technical introduction we list some well-known theorems and formulas for later use.[4]

Given a 4-dimensional manifold M_4, we denote the vector fields on M_4 by $X, Y, Z, \cdots$ and the 1-forms by $\omega, \theta, \phi, \cdots$. M_4 and the tensor fields can be regarded as analytic. The exterior derivative of ω is given by

$$d\omega(X, Y) = \tfrac{1}{2}\{X\omega(Y) - Y\omega(X) - \omega([X, Y])\}. \quad (2.1)$$

We denote the basis for the vector fields by

$$X_0, X_1, X_2, X_3 \quad (2.2)$$

and that for the 1-forms by

$$\omega^0, \omega^1, \omega^2, \omega^3, \quad (2.3)$$

where

$$\omega^a(X_b) = \delta^a{}_b. \quad (2.4)$$

We introduce an affine connection on M_4 by

$$\nabla_{X_a}(X_b) = \Gamma_{ab}{}^c X_c. \quad (2.5)$$

The connection form is defined by

$$\omega^a{}_b = \Gamma_{cb}{}^a \omega^c. \quad (2.6)$$

The covariant differentiation ∇_X is a derivation of the algebra $T(M_4)$ of the tensor fields such that it preserves the type of the tensor field and commutes with all contractions. The covariant derivative of a vector field Y is given by

$$\nabla_X(Y) = \xi^a(X_a\eta^c + \eta^b\Gamma_{ab}{}^c)X_c, \quad (2.7)$$

where

$$X = \xi^a X_a \quad \text{and} \quad Y = \eta^b X_b. \quad (2.8)$$

The covarient derivative of a 1-form is

$$U(X, Y) = (\nabla_X\omega)(Y) = X\omega(Y) - \omega(\nabla_X(Y)). \quad (2.9)$$

The components of the tensor field U, defined above, are given by

$$U_{ab} = U(X_a, X_b) = X_a u_b - \Gamma_{ab}{}^c u_c, \quad (2.10)$$

where

$$u_b = \omega(X_b). \quad (2.11)$$

The Lie derivative of Y with respect to X is defined by

$$L_X(Y) = \nabla_X(Y) - \nabla_Y(X). \quad (2.12)$$

One defines the torsion tensor field by

$$T(X, Y) = \nabla_X(Y) - \nabla_Y(X) - [X, Y] \quad (2.13)$$

and the curvature tensor field by

$$R(X, Y)Z = \nabla_X\nabla_Y(Z) - \nabla_Y\nabla_X(Z) - \nabla_{[X,Y]}(Z). \quad (2.14)$$

The components of T and R are

$$T(X_b, X_c) = T^a{}_{bc} X_a, \quad R(X_c, X_d)X_b = R^a{}_{bcd} X_a. \quad (2.15)$$

Cartan's structure equations are

$$d\omega^a = -\omega^a{}_p \wedge \omega^p + \tfrac{1}{2}T^a{}_{pq}\,\omega^p \wedge \omega^q, \quad (2.16)$$

$$d\omega^a{}_b = -\omega^a{}_p \wedge \omega^p{}_b + \tfrac{1}{2}R^a{}_{bpq}\,\omega^p \wedge \omega^q. \quad (2.17)$$

We assume henceforth that

$$T = 0, \quad \text{that is,} \quad T^a{}_{bc} = 0. \tag{2.18}$$

We define the functions

$$C^a{}_{bc} = -C^a{}_{cb}, \quad a, b, c, \cdots = 0, 1, 2, 3, \tag{2.19}$$

by the equations

$$[X_b, X_c] = C^a{}_{bc} X_a; \tag{2.20}$$

then it follows, by using (2.1) and (2.4), that

$$d\omega^a = -\tfrac{1}{2} C^a{}_{pq}\, \omega^p \wedge \omega^q \tag{2.21}$$

and, from (2.16), (2.18) and (2.6), that

$$d\omega^a = -\tfrac{1}{2}(\Gamma_{pq}{}^a - \Gamma_{qp}{}^a)\omega^p \wedge \omega^q. \tag{2.22}$$

Therefore

$$C^a{}_{bc} = \Gamma_{bc}{}^a - \Gamma_{cb}{}^a. \tag{2.23}$$

We introduce a pseudo-Riemannian metric on M_4 by the nondegenerate tensor field

$$g(X, Y) = g(Y, X). \tag{2.24}$$

It is well known that on a pseudo-Riemannian manifold there exists one and only one affine connection such that

$$T = 0 \quad \text{and} \quad \nabla_Z g = 0, \tag{2.25}$$

that is,

$$\nabla_X(Y) - \nabla_Y(X) = [X, Y] \tag{2.26}$$

and

$$2g(\nabla_X(Y). Z) = Xg(Y, Z) + Yg(Z, X) - Zg(X, Y) + g(Y, [Z, X]) + g(Z, [X, Y]) - g(X, [Y, Z]). \tag{2.27}$$

Suppose that

$$g(X_a, X_b) = g_{ab} = \text{diag}\,(+1, -1, -1, -1), \tag{2.28}$$

in other words,

$$g = g_{ab}\omega^a\omega^b. \tag{2.29}$$

It follows then that

$$2g(\nabla_{X_a}(X_b), X_c) = g(X_b, [X_c, X_a]) + g(X_c, [X_a, X_b]) - g(X_a, [X_b, X_c]) \tag{2.30}$$

and, using the notation

$$\Gamma_{abc} = \Gamma_{ab}{}^d g_{dc}, \quad C_{abc} = g_{ad} C^d{}_{bc}, \tag{2.31}$$

we obtain

$$\Gamma_{abc} = \tfrac{1}{2}(C_{bca} + C_{cab} - C_{abc}). \tag{2.32}$$

Using (2.1) and (2.17), we have

$$\begin{aligned} d\omega^a{}_b(X_c, X_d) &= \tfrac{1}{2}(X_c\omega^a{}_b(X_d) - X_d\omega^a{}_b(X_c) - \omega^a{}_b([X_c, X_d])) \\ &= -\tfrac{1}{2}(\omega^a{}_p(X_c)\omega^p{}_b(X_d) - \omega^a{}_p(X_d)\omega^p{}_b(X_c)) + \tfrac{1}{2}R^a{}_{bcd}, \end{aligned}$$

and, therefore,

$$R^a{}_{bcd} = \Gamma_{cf}{}^a \Gamma_{db}{}^f - \Gamma_{df}{}^a \Gamma_{cb}{}^f - \Gamma_{fb}{}^a C^f{}_{cd} + X_c\Gamma_{db}{}^a - X_d\Gamma_{cb}{}^a. \tag{2.33}$$

It should be noted that the power of the formalism developed above lies in the freedom of choice for the basis X_0, X_1, X_2, X_3 of the vector fields or ω^0, ω^1, ω^2, ω^3 of the 1-forms, respectively [with the proviso (2.4)]. In the following we specialize our manifold M_4 and make a definite choice for the case most adequate for our problem. The steps are as follows: Suppose that the functions $C^a{}_{bc}$ are constants and satisfy the Jacobi identities. Then our pseudo-Riemannian manifold M_4 is a Lie group. Suppose that M_4 is simply connected. Then it is the universal covering group, uniquely defined by the Lie algebra (2.20) of the invariant vector fields X_0, X_1, X_2, X_3. The corresponding left-invariant 1-forms ω^0, ω^1, ω^2, ω^3 satisfy

$$d\omega^a = -\tfrac{1}{2}C^a{}_{bc}\omega^b \wedge \omega^c \tag{2.34}$$

and characterize M_4 equivalently.

We choose now, for the base of vector fields or of the 1-forms on M_4, the invariant vector fields X_0, X_1, X_2, X_3 or invariant 1-forms ω^0, ω^1, ω^2, ω^3, respectively.

The requirement (2.28), that the X_0, X_1, X_2, X_3 should be pseudo-orthonormal, defines the left-invariant metric. Or, equivalently,

$$g = g_{ab}\omega^a\omega^b. \tag{2.35}$$

Generally speaking, this choice of base and the formalism sketched above allows one to discuss many properties of the group or Riemannian space M_4 from a simple knowledge of the constants of structure $C^a{}_{bc}$. We do not have to specialize the coordinates and can perform many calculations without an explicit knowledge of the left-invariant forms. The most important formal consequence of the above choice is that the corresponding Γ's are constants [see (2.32)]. But, above all in importance, our results will be global results since the theory of Lie groups[5] assures us that these vector fields and forms exist globally. Since the Γ's are constants, (2.33) reduces to

$$R^a{}_{bcd} = \Gamma_{cf}{}^a \Gamma_{db}{}^f - \Gamma_{df}{}^a \Gamma_{cb}{}^f - \Gamma_{fb}{}^a C^f{}_{cd}. \tag{2.36}$$

The components of the Ricci tensor field are

$$R_{bc} = R^f{}_{bcf} = \Gamma_{fb}{}^g \Gamma_{gc}{}^f + C^g{}_{fg} \Gamma_{bc}{}^f. \tag{2.37}$$

The field equations are

$$G_{ab} + \Lambda g_{ab} = R_{ab} - \tfrac{1}{2}Rg_{ab} + \Lambda g_{ab} = -\kappa\rho u_a u_b, \quad u_a u^a = 1, \tag{2.38}$$

where

$$R = R^a{}_a. \tag{2.39}$$

After a trivial computation, we obtain

$$R_{ab} = -\kappa\rho u_a u_b + (\Lambda + \kappa\rho/2)g_{ab}, \quad u_a u^a = 1. \quad (2.40)$$

3. THE GROUP

Consider a 4-dimensional vector space over the field of real numbers. We denote the base vectors by

$$\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \quad (3.1)$$

and convert this vector space into an algebra by introducing the noncommutative multiplication by the following requirements:

$$\begin{aligned} &\mathbf{e}_0\mathbf{e}_\mu = \mathbf{e}_\mu\mathbf{e}_0 = \mathbf{e}_\mu, \quad \mu = 0, 1, 2, 3, \\ &\mathbf{e}_1\mathbf{e}_1 = -\mathbf{e}_0, \quad \mathbf{e}_2\mathbf{e}_2 = \mathbf{e}_0, \quad \mathbf{e}_3\mathbf{e}_3 = \mathbf{e}_0, \\ &\mathbf{e}_2\mathbf{e}_3 = -\mathbf{e}_3\mathbf{e}_2 = -\mathbf{e}_1, \quad \mathbf{e}_3\mathbf{e}_1 = -\mathbf{e}_1\mathbf{e}_3 = \mathbf{e}_2, \\ &\mathbf{e}_1\mathbf{e}_2 = -\mathbf{e}_2\mathbf{e}_1 = \mathbf{e}_3. \end{aligned} \quad (3.2)$$

We call this algebra Gödel's quaternion algebra and the vectors

$$\mathbf{a} = a^\mu \mathbf{e}_\mu \quad (3.3)$$

Gödel quaternions.[6]

Introducing the conjugate quaternion $\mathbf{a}^*$ by

$$\mathbf{a}^* = a^0\mathbf{e}_0 - a^i\mathbf{e}_i, \quad (3.4)$$

we have from (3.2) that

$$\begin{aligned} \mathbf{aa}^* &= [(a^0)^2 + (a^1)^2 - (a^2)^2 - (a^3)^2]\mathbf{e}_0 \\ &= (a^0)^2 + (a^1)^2 - (a^2)^2 - (a^3)^2. \end{aligned} \quad (3.5)$$

We identified here the subfield $a^0\mathbf{e}_0$ with the real field.

Consider now the normed Gödel quaternions, that is, quaternions $\mathbf{a}$ satisfying the condition

$$\mathbf{aa}^* = (a^0) + (a^1)^2 - (a^2)^2 - (a^3)^2 = 1. \quad (3.6)$$

They obviously form a group with respect to the quaternion multiplication. Identifying the vectors (3.1) with the unit vectors along the axes in a 4-dimensional pseudo-Euclidean space of signature

$$+ + - -, \quad (3.7)$$

or Euclidean space of coordinates a^0, a^1, a^2, a^3, denoted by E^4, we find that (3.6) is the equation of the sphere or hyperboloid H^3, respectively. The manifold H^3 with the quaternion multiplication (3.2) is a Lie group, which we denote also by H^3.

We want to obtain the left-invariant vector fields of H^3 in the coordinate system induced by the Cartesian coordinates of the imbedding E^4. We consider, therefore, in the point $\mathbf{e}_0$ on H^3 the three vectors

$$\mathbf{e}_0 + \epsilon\mathbf{e}_i, \quad i = 1, 2, 3, \quad (3.8)$$

tangential to H^3 and propagate them by the left translations over H^3 generating the three independent left-invariant vector fields mentioned above. Since

$$\mathbf{a} = \mathbf{ae}_0, \quad \mathbf{a} + \epsilon\boldsymbol{\omega}_i = \mathbf{a}(\mathbf{e}_0 + \epsilon\mathbf{e}_i), \quad (3.9)$$

we obtain

$$\boldsymbol{\omega}_i = \mathbf{ae}_i \quad (3.10)$$

as the vectors at $\mathbf{a}$, corresponding to $\mathbf{e}_i$ at $\mathbf{e}_0$. Defining the components $e_i{}^\mu$ of $\boldsymbol{\omega}_i$ by

$$\boldsymbol{\omega}_i = e_i{}^\mu\mathbf{e}_\mu, \quad (3.11)$$

we obtain, using (3.10), (2.3),and (3.3), the following expressions:

$$e_1{}^\mu = (-a^1, a^0, a^3, -a^2), \quad e_2{}^\mu = (a^2, a^3, a^0, a^1), \quad e_3{}^\mu = (a^3, -a^2, -a^1, a^0). \quad (3.12)$$

Therefore, the invariant vector fields

$$E_i = e_i^\mu \frac{\partial}{\partial a^\mu} \quad (3.13)$$

are given by

$$\begin{aligned} E_1 &= -a^1\frac{\partial}{\partial a^0} + a^0\frac{\partial}{\partial a^1} + a^3\frac{\partial}{\partial a^2} - a^2\frac{\partial}{\partial a^3}, \\ E_2 &= a^2\frac{\partial}{\partial a^0} + a^3\frac{\partial}{\partial a^1} + a^0\frac{\partial}{\partial a^2} + a^1\frac{\partial}{\partial a^3}, \\ E_3 &= a^3\frac{\partial}{\partial a^0} - a^2\frac{\partial}{\partial a^1} - a^1\frac{\partial}{\partial a^2} + a^2\frac{\partial}{\partial a^3}. \end{aligned} \quad (3.14)$$

Computing the commutator relations, we obtain

$$[E_2, E_3] = -2E_1, \quad [E_3, E_1] = 2E_2, \quad [E_1, E_2] = 2E_3. \quad (3.15)$$

Introducing for later use new vector fields

$$X_0 = -\tfrac{1}{2}E_1, \quad X_1 = -\tfrac{1}{2}E_3, \quad X_2 = -\tfrac{1}{2}E_2, \quad (3.16)$$

we obtain

$$[X_1, X_2] = -X_0, \quad [X_2, X_0] = X_1, \quad [X_0, X_1] = X_2. \quad (3.17)$$

If we represent the unit quaternions

$$\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$$

by the matrices

$$\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \quad \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}, \quad \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \quad \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}, \quad (3.18)$$

respectively, every Gödel quaternion

$$a^\mu\mathbf{e}_\mu \quad (3.19)$$

goes over to the matrix

$$A = \begin{pmatrix} a^0 + a^3 & a^1 + a^2 \\ -a^1 + a^2 & a^0 - a^3 \end{pmatrix} \tag{3.20}$$

with

$$(a^0)^2 + (a^1)^2 - (a^2)^2 - (a^3)^2 = 1, \tag{3.21}$$

and the quaternion muliplication goes over to the matrix multiplication. This expresses the well-known fact that H^3 is isomorphic to $SLG(2, R)$.

We introduce on H^3 a new coordinate system

$$x^0,\ x^1,\ x^2 \tag{3.22}$$

by the substitutions

$$\begin{aligned} a^0 &= \tfrac{1}{2}e^{\frac{1}{2}x^1}x^2 \cos \tfrac{1}{2}x^0 + \cosh \tfrac{1}{2}x^1 \sin \tfrac{1}{2}x^0, \\ a^1 &= \tfrac{1}{2}e^{\frac{1}{2}x^1}x^2 \sin \tfrac{1}{2}x^0 + \cosh \tfrac{1}{2}x^1 \cos \tfrac{1}{2}x^0, \\ a^2 &= \tfrac{1}{2}e^{\frac{1}{2}x^1}x^2 \sin \tfrac{1}{2}x^0 + \sinh \tfrac{1}{2}x^1 \cos \tfrac{1}{2}x^0, \\ a^3 &= -\tfrac{1}{2}e^{\frac{1}{2}x^1}x^2 \cos \tfrac{1}{2}x^0 + \sinh \tfrac{1}{2}x^1 \sin \tfrac{1}{2}x^0. \end{aligned} \tag{3.23}$$

(This is a two-parametric family of straight lines on H^3—x^0 and x^1 being the parameters—and x^2 is the coordinate along the lines.)

This coordinate system covers H^3 completely. The matrix A is given in this coordinate by

$$A = \begin{pmatrix} e^{\frac{1}{2}x^1} \sin \tfrac{1}{2}x^0 & e^{\frac{1}{2}x^1} \cos \tfrac{1}{2}x^0 \\ e^{\frac{1}{2}x^1}x^2 \sin \tfrac{1}{2}x^0 - e^{-\frac{1}{2}x^1} \cos \tfrac{1}{2}x^0 & e^{\frac{1}{2}x^1}x \cos^2 \tfrac{1}{2}x^0 + e^{-\frac{1}{2}x^1} \sin \tfrac{1}{2}x^0 \end{pmatrix}. \tag{3.24}$$

The left-invariant 1-forms of a matrix group whose general element is given by the matrix A can be obtained by computing

$$\omega = A^{-1}\, dA. \tag{3.25}$$

As shown, for instance, by Flanders,[7] all matrix elements of ω will be left-invariant 1-forms. Carrying out the computation indicated in (3.25), we obtain

$$\omega = \begin{pmatrix} -\tfrac{1}{2}(\cos x^0\, dx^1 + e^{x^1} \sin x^0\, dx^2) & -\tfrac{1}{2}\, dx^0 + \tfrac{1}{2} \sin x^0\, dx^1 - e^{x^1} \cos^2 \tfrac{1}{2}x^0\, dx^2 \\ \tfrac{1}{2}\, dx^0 + \tfrac{1}{2} \sin x^0\, dx^1 + e^{x^1} \sin^2 \tfrac{1}{2}x^0\, dx^2 & \tfrac{1}{2}(\cos x^0\, dx^1 + e^{x^1} \sin x^0\, dx^2) \end{pmatrix}. \tag{3.26}$$

We select from (3.26) the following left-invariant 1-forms:

$$\begin{aligned} \omega^0 &= dx^0 + e^{x^1}\, dx^2, \\ \omega^1 &= \cos x^0\, dx^1 + e^{x^1} \sin x^0\, dx^2, \\ \omega^2 &= -\sin x^0\, dx^1 + e^{x^1} \cos x^0\, dx^2 \end{aligned} \tag{3.27}$$

as the base for the 1-forms on H^3. The corresponding left-invariant vector fields, serving as the base for the vector fields on H^3, are given by

$$\begin{aligned} X_0 &= \frac{\partial}{\partial x^0}, \\ X_1 &= -\sin x^0 \frac{\partial}{\partial x^0} + \cos x^0 \frac{\partial}{\partial x^1} + e^{-x^1} \sin x^0 \frac{\partial}{\partial x^2}, \\ X_2 &= -\cos x^0 \frac{\partial}{\partial x^0} - \sin x^0 \frac{\partial}{\partial x^1} + e^{-x^1} \cos x^0 \frac{\partial}{\partial x^2}. \end{aligned} \tag{3.28}$$

These are the vector fields defined by (3.16), written in the coordinate system (3.22) as defined by the substitutions (3.23), as one can see easily by a straightforward computation.

We now consider the group

$$M_4 = R \times H^3, \tag{3.29}$$

where the coordinate x^3 is introduced on R and

$$X_3 = \frac{\partial}{\partial x^3}. \tag{3.30}$$

Therefore, the left-invariant vector fields on M_4,

$$X_0,\ X_1,\ X_2,\ X_3, \tag{3.31}$$

defined by (3.28) and (3.30), can be chosen for the base of the vector fields on M_4, and the left-invariant 1-forms

$$\omega^0,\ \omega^1,\ \omega^2,\ \omega^3, \tag{3.32}$$

defined by (3.27),and

$$\omega^3 = dx^3 \tag{3.33}$$

are the corresponding base for the 1-forms on M_4. The Lie algebra of the left-invariant vector fields on M_4 is given by

$$\begin{gathered} [X_1, X_2] = -X_0, \quad [X_2, X_0] = X_1, \quad [X_0, X_1] = X_2, \\ [X_a, X_3] = 0, \quad a = 0, 1, 2, \end{gathered} \tag{3.34}$$

or, correspondingly,

$$\begin{aligned} d\omega^0 &= \omega^1 \wedge \omega^2, & d\omega^1 &= -\omega^2 \wedge \omega^0, \\ d\omega^2 &= -\omega^0 \wedge \omega^1, & d\omega^3 &= 0. \end{aligned} \tag{3.35}$$

In the subsequent sections two different pseudo-Riemannian metrics, invariant under the left translations of the group M_4 and satisfying the Einstein equations (2.40), are introduced on M_4, which is by construction simply connected and, therefore, is the uniquely defined universal covering group of the Lie algebra (3.34). These manifolds are called the Class II and Class III universes. We discuss their properties.

4. THE METRIC OF THE CLASS II UNIVERSES

We construct the metric on M_4 as follows. We let

$$R > 0 \quad \text{and} \quad \tfrac{1}{2} < |k| \leq (2)^{-\frac{1}{2}} \tag{4.1}$$

be two real parameters and introduce a new basis in the Lie algebra (3.34) by the following substitutions:

$$\begin{aligned} Y_0 &= \frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}} X_0 + \frac{1}{R}\left(\frac{2(1-2k^2)}{4k^2-1}\right)^{\frac{1}{2}} X_3, \\ Y_1 &= \frac{2}{R(1-k)^{\frac{1}{2}}} X_1, \quad Y_2 = \frac{2}{R(1+k)^{\frac{1}{2}}} X_2, \\ Y_3 &= \frac{2}{R}\left(\frac{1-2k^2}{4k^2-1}\right)^{\frac{1}{2}} X_0 + \frac{4k^2}{R}\frac{1}{(4k^2-1)^{\frac{1}{2}}} X_3. \end{aligned} \tag{4.2}$$

We define the metric on M_4 by demanding that Y_0, Y_1, Y_2, Y_3 be pseudo-orthonormal, that is,

$$g(Y_a, Y_b) = g_{ab} = \operatorname{diag}(+1, -1, -1, -1). \tag{4.3}$$

In other words, we define the line element to be

$$ds^2 = (\theta^0)^2 - (\theta^1)^2 - (\theta^2)^2 - (\theta^3)^2, \tag{4.4}$$

where $\theta^0, \theta^1, \theta^2, \theta^3$ is the basis of the 1-forms, corresponding to Y_0, Y_1, Y_2, Y_3 and given by

$$\begin{aligned} \theta^0 &= Rk^2\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\omega^0 - R\left(\frac{1-2k^2}{2(4k^2-1)}\right)^{\frac{1}{2}}\omega^3, \\ \theta^1 &= R[\tfrac{1}{2}(1-k)]^{\frac{1}{2}}\omega^1, \quad \theta^2 = R[\tfrac{1}{2}(1+k)]^{\frac{1}{2}}\omega^2, \\ \theta^3 &= -\frac{R}{2}\left(\frac{1-2k^2}{4k^2-1}\right)^{\frac{1}{2}}\omega^0 + \frac{R}{2}\frac{1}{(4k^2-1)^{\frac{1}{2}}}\omega^3. \end{aligned} \tag{4.5}$$

After trivial computations we obtain

$$\begin{aligned} ds^2 = (\tfrac{1}{2}R)^2[&(1+2k^2)(\omega^0)^2 - (1-k)(\omega^1)^2 \\ &- (1+k)(\omega^2)^2 - (\omega^3)^2 - 2(1-2k^2)^{\frac{1}{2}}\omega^0\omega^3], \end{aligned} \tag{4.6}$$

where the ω's satisfy (3.35) and are given in our coordinate system by (3.27). We would like to make the following remarks to (4.6). Since the invariant vector field X_3 commutes with the other vector fields X_a, $a = 0, 1, 2$ [see (3.34)], X_3 is also a generator of M_4. Therefore, in a coordinate system where $X_3 = \partial/\partial x^3$, the ω's do not depend on x^3 [see (3.27) and (3.30)]. But, since X_3 is not hypersurface orthogonal, we cannot get rid of the "cross terms" in the metric. Consider now the vector field $K = \partial/\partial x^2$. One sees that $[K, X_a] = 0$, $a = 0, 1, 2, 3$; and that therefore K is also a generator of M_4; consequently, the ω's and the metric are independent of x^2. Since

$$g(K, K) = (\tfrac{1}{2}R)^2 k(2k - \cos 2x^0)e^{2x^1} > 0$$

for $\frac{1}{2} < |k|$, K is a timelike generator of M_4 and x^2 a timelike coordinate. Therefore, (4.6) in our coordinate system exhibits the fact that the metric is stationary. But it is not static, since there is no hypersurface orthogonal time like Killing vector field.

It will turn out that the vector field Y_0 is tangent to the world lines of the matter. We introduce now new coordinates

$$\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3 \tag{4.7}$$

by the substitutions

$$\begin{aligned} \tilde{x}^0 &= R[\tfrac{1}{2}(4k^2-1)]^{\frac{1}{2}}x^0, \quad \tilde{x}^1 = x^1, \quad \tilde{x}^2 = x^2, \\ \tilde{x}^3 &= -(1-2k^2)^{\frac{1}{2}}x^0 + x^3 \end{aligned} \tag{4.8}$$

or

$$\begin{aligned} x^0 &= \frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\tilde{x}^0, \quad x^1 = \tilde{x}^1, \quad x^2 = \tilde{x}^2, \\ x^3 &= \frac{1}{R}\left(\frac{2(1-2k^2)}{4k^2-1}\right)^{\frac{1}{2}}\tilde{x}^0 + \tilde{x}^3. \end{aligned} \tag{4.9}$$

We see that

$$Y = \frac{\partial}{\partial \tilde{x}^0}, \tag{4.10}$$

which shows that the matter is at rest with respect to the coordinates (4.7).

Carrying out straightforward calculations, we find that

$$\begin{aligned} \omega^0 &= \frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}} d\tilde{x}^0 + \exp(\tilde{x}^1)\, d\tilde{x}^2, \\ \omega^1 &= \cos\frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\tilde{x}^0\, d\tilde{x}^1 \\ &\quad + \exp(\tilde{x}^1)\sin\frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\tilde{x}^0\, d\tilde{x}^2, \\ \omega^2 &= -\sin\frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\tilde{x}^0\, dx^1 \\ &\quad + \exp(\tilde{x}^1)\cos\frac{1}{R}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\tilde{x}^0\, d\tilde{x}^2, \\ \omega^3 &= \frac{1}{R}\left(\frac{2(1-2k^2)}{4k^2-1}\right)^{\frac{1}{2}} d\tilde{x}^0 + d\tilde{x}^3. \end{aligned} \tag{4.11}$$

Substituting (4.11) into (4.6), we obtain the metric in the new coordinate system (4.7). Carrying out these computations, we see that the metric has the form

$$ds^2 = (d\tilde{x}^0)^2 + 2\tilde{p}_\alpha\tilde{\omega}^\alpha\, d\tilde{x}^0 + \tilde{g}_{\alpha\beta}\tilde{\omega}^\alpha\tilde{\omega}^\beta, \quad \alpha, \beta, \gamma, \cdots = 1, 2, 3, \tag{4.12}$$

where $\tilde{p}_\alpha$ and $\tilde{g}_{\alpha\beta}$ are functions of $\tilde{x}^0$ alone and $\tilde{\omega}^1$, $\tilde{\omega}^2$, $\tilde{\omega}^3$ are 1-forms given by

$$\tilde{p}_\alpha = \left(0, Rk^2\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}, R\left(\frac{1-2k^2}{2(4k^2-1)}\right)^{\frac{1}{2}}\right), \tag{4.13}$$

$$\tilde{g}_{\alpha\beta} = (\tfrac{1}{2}R)^2\begin{pmatrix} -1 + k\cos 2\tilde{x}^0 & k\sin 2\tilde{x}^0 & 0 \\ k\sin 2\tilde{x}^0 & k(2k - \cos 2\tilde{x}^0) & -(1-2k^2)^{\frac{1}{2}} \\ 0 & -(1-2k^2)^{\frac{1}{2}} & -1 \end{pmatrix}, \tag{4.14}$$

and

$$\tilde{\omega}^1 = d\tilde{x}^1, \quad \tilde{\omega}^2 = \exp(\tilde{x}^1)\, d\tilde{x}^2, \quad \tilde{\omega}^3 = d\tilde{x}^3, \tag{4.15}$$

respectively. The 1-forms (4.15) are the left-invariant 1-forms of the group of Bianchi type III since

$$d\tilde{\omega}^1 = 0, \quad d\tilde{\omega}^2 = \tilde{\omega}^1 \wedge \tilde{\omega}^2, \quad d\tilde{\omega}^3 = 0, \tag{4.16}$$

as one sees immediately. Our solution is, therefore, a special case of spatially homogeneous solutions admitting the group (4.16).

We introduce now another coordinate system

$$\bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3 \tag{4.17}$$

by the substitutions

$$\bar{x}^0 = x^0 - \frac{(1-2k^2)^{\frac{1}{2}}}{1+2k^2}x^3, \quad \bar{x}^1 = x^1, \quad \bar{x}^2 = x^2,$$
$$\bar{x}^3 = \left(\frac{2}{1+2k^2}\right)^{\frac{1}{2}} x^3 \tag{4.18}$$

or

$$x^0 = \bar{x}^0 + \left(\frac{1-2k^2}{2(1+2k^2)}\right)^{\frac{1}{2}}\bar{x}^3, \quad \bar{x}^1 = x^1, \quad \bar{x}^2 = x^2,$$
$$x^3 = \tfrac{1}{2}[(1+2k^2)]^{\frac{1}{2}}\bar{x}^3. \tag{4.19}$$

We will see that $\bar{x}^3 = \text{const}$ are the H^3 hypersurfaces.

Carrying out these coordinate transformations, we obtain from (4.2) the following expressions:

$$Y_0 = \frac{1}{R}\frac{4k^2}{1+2k^2}\left(\frac{2}{4k^2-1}\right)^{\frac{1}{2}}\bar{X}_0 + \frac{2}{R}\left(\frac{1-2k^2}{(4k^2-1)(1+2k^2)}\right)^{\frac{1}{2}}\bar{X}_3,$$
$$Y_1 = \frac{2}{R(1-k)^{\frac{1}{2}}}(\bar{X}_1\cos\beta\bar{x}^3 + \bar{X}_3\sin\beta\bar{x}^3),$$
$$Y_2 = \frac{2}{R(1+k)^{\frac{1}{2}}}(-\bar{X}_1\sin\beta\bar{x}^3 + \bar{X}_3\cos\beta\bar{x}^3), \tag{4.20}$$
$$Y_3 = \frac{2}{R}\frac{1}{1+2k^2}\left(\frac{1-2k^2}{4k^2-1}\right)^{\frac{1}{2}}\bar{X}_0 + \frac{4k^2}{R}\left(\frac{2}{(4k^2-1)(1+2k^2)}\right)^{\frac{1}{2}}\bar{X}_3,$$

where

$$\beta = [(1-2k^2)/2(1+2k^2)]^{\frac{1}{2}} \tag{4.21}$$

and

$$\bar{X}_0 = \frac{\partial}{\partial\bar{x}^0},$$
$$\bar{X}_1 = -\sin\bar{x}^0\frac{\partial}{\partial\bar{x}^0} + \cos\bar{x}^0\frac{\partial}{\partial\bar{x}^1} + \exp(-\bar{x}^1)\sin\bar{x}^0\frac{\partial}{\partial\bar{x}^2},$$
$$\bar{X}_2 = -\cos\bar{x}^0\frac{\partial}{\partial\bar{x}^0} - \sin\bar{x}^0\frac{\partial}{\partial\bar{x}^1} + \exp(-\bar{x}^1)\cos\bar{x}^0\frac{\partial}{\partial\bar{x}^2}, \tag{4.22}$$
$$\bar{X}_3 = \frac{\partial}{\partial\bar{x}^3}.$$

We now introduce a new basis for the Lie algebra of the left-invariant vector fields on M_4 by the following substitutions:

$$Z_0 = 2k^2\left(\frac{2}{(4k^2-1)(1+2k^2)}\right)^{\frac{1}{2}}Y_0 - \left(\frac{1-2k^2}{(4k^2-1)(1+2k^2)}\right)^{\frac{1}{2}}Y_3,$$
$$Z_1 = Y_1, \quad Z_2 = Y_2, \tag{4.23}$$
$$Z_3 = -\left(\frac{1-2k^2}{(4k^2-1)(1+2k^2)}\right)^{\frac{1}{2}}Y_0 + 2k^2\left(\frac{2}{(4k^2-1)(1+2k^2)}\right)^{\frac{1}{2}}Y_3.$$

Since (4.23) is a Lorentz transformation between the Z's and the Y's, the metrics defined by

$$g(Z_a, Z_b) = g_{ab} = \text{diag}(+1, -1, -1, -1) \tag{4.24}$$

and

$$g(Y_a, Y_b) = g_{ab} = \text{diag}(+1, -1, -1, -1) \tag{4.25}$$

are identical. After obvious substitutions we obtain

$$
\begin{aligned}
Z_0 &= \frac{2}{R(1+2k^2)^{\frac{1}{2}}}\bar{X}_0,\\
Z_1 &= \frac{2}{R(1-k)^{\frac{1}{2}}}(\bar{X}_1\cos\beta\bar{x}^3+\bar{X}_2\sin\beta\bar{x}^3),\\
Z_2 &= \frac{2}{R(1+k)^{\frac{1}{2}}}(-\bar{X}_1\sin\beta\bar{x}^3+\bar{X}_2\cos\beta\bar{x}^3),\\
Z_3 &= \frac{2}{R}\bar{X}_3,
\end{aligned}
\tag{4.26}
$$

where β and $\bar{X}_0$, $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$ are given by (4.21) and (4.22), respectively. The corresponding left-invariant 1-forms are

$$
\begin{aligned}
\phi^0 &= \tfrac{1}{2}R(1+2k^2)^{\frac{1}{2}}\bar{\omega}^0,\\
\phi^1 &= \tfrac{1}{2}R(1-k)^{\frac{1}{2}}(\bar{\omega}^1\cos\beta\bar{x}^3+\bar{\omega}^2\sin\beta\bar{x}^3),\\
\phi^2 &= \tfrac{1}{2}R(1+k)^{\frac{1}{2}}(-\bar{\omega}^1\sin\beta\bar{x}^3+\bar{\omega}^2\cos\beta\bar{x}^3),\\
\phi^3 &= \tfrac{1}{2}R\bar{\omega}^3.
\end{aligned}
\tag{4.27}
$$

As the consequence of all that, the line element

$$
\begin{aligned}
ds^2 &= (\tfrac{1}{2}R)^2(1+2k^2)(\bar{\omega}^0)^2\\
&\quad -(1-k)(\bar{\omega}^1\cos\beta\bar{x}^3+\bar{\omega}^2\sin\beta\bar{x}^3)^2\\
&\quad -(1+k)(-\bar{\omega}^1\sin\beta\bar{x}^3+\bar{\omega}^2\cos\beta\bar{x}^3)^2-(\bar{\omega}^3)^2
\end{aligned}
\tag{4.28}
$$

is the same as (4.6) but is in the new coordinate system, where $\bar{\omega}^0$, $\bar{\omega}^1$, $\bar{\omega}^2$, $\bar{\omega}^3$ are given by

$$
\begin{aligned}
\bar{\omega}^0 &= d\bar{x}^0+\exp(\bar{x}^1)\,d\bar{x}^2,\\
\bar{\omega}^1 &= \cos\bar{x}^0\,d\bar{x}^1+\exp\bar{x}^1\sin\bar{x}^0\,d\bar{x}^2,\\
\bar{\omega}^2 &= -\sin\bar{x}^0\,d\bar{x}^1+\exp(\bar{x}^1)\cos\bar{x}^0\,d\bar{x}^2,\\
\bar{\omega}^3 &= d\bar{x}^3.
\end{aligned}
\tag{4.29}
$$

To sum up our findings, we see that we used two different bases for the Lie algebra of the left-invariant vector fields on M_4, namely,

$$Y_0,\, Y_1,\, Y_2,\, Y_3 \tag{4.30}$$

and

$$Z_0, Z_1, Z_2, Z_3. \tag{4.31}$$

We shall see in the next section that (4.30) is intimately connected with the motion of the matter in our solutions. (4.31) is distinguished by the geometry of the 3-dimensional hypersurfaces corresponding to the normal subgroup H^3 of M_4, as we shall see in Sec. 6.

The three different coordinate systems employed differ as follows: In (4.6) with the coordinates x^0, x^1, x^2, x^3, the x^2 lines are the integral curves of the time-like generators of M_4; in (4.12) the $\bar{x}^0$ lines are the world lines of the matter, as we shall see in the next section; in (4.28) the $\bar{x}^3$ lines are perpendicular to the 3-dimensional hypersurfaces corresponding to the normal subgroup. In case of $k=\frac{1}{2}$ we obtain a cosmos filled with radiation.

5. MISCELLANEOUS RESULTS AND THE MOTION OF THE MATTER

Consider the basis Y_0, Y_1, Y_2, Y_3 defined by (4.2). The Lie algebra of M_4 is given in this basis by the following commutation relations:

$$
\begin{aligned}
[Y_1, Y_2] &= -\frac{4k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}Y_0\\
&\quad +\frac{2}{R}\left(\frac{1-2k^2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}Y_3,\\
[Y_2, Y_0] &= \frac{1-k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}Y_1,\\
[Y_0, Y_1] &= \frac{1+k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}Y_2,\\
[Y_0, Y_3] &= 0,\\
[Y_1, Y_3] &= -\frac{2(1+k)}{R}\left(\frac{1-2k^2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}Y_2,\\
[Y_2, Y_3] &= +\frac{2(1-k)}{R}\left(\frac{1-2k^2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}Y_1.
\end{aligned}
\tag{5.1}
$$

Using (2.32), we compute the components of the affine connection:

$$
\begin{aligned}
\Gamma_{012} &= -\frac{1-2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}},\\
\Gamma_{123} &= -\frac{1+2k}{R}\left(\frac{1-2k^2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}},\\
\Gamma_{120} &= -\frac{k(1+2k)}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}},\\
\Gamma_{231} &= -\frac{1-2k}{R}\left(\frac{1-2k^2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}},\\
\Gamma_{210} &= -\frac{k(1-2k)}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}},\\
\Gamma_{312} &= -\frac{1}{R}\left(\frac{1-2k^2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}.
\end{aligned}
\tag{5.2}
$$

Using (2.37), we see that the components of the Ricci tensor field are given by

$$
\begin{aligned}
R_{ab} = \operatorname{diag}\Bigg(&-\frac{4k^2}{R^2(1-k^2)},\frac{2(1-2k^2)}{R^2(1-k^2)},\\
&\frac{2(1-2k^2)}{R^2(1-k^2)},\frac{2(1-2k^2)}{R^2(1-k^2)}\Bigg).
\end{aligned}
\tag{5.3}
$$

Comparing with (2.40), we see that we indeed have a solution of the field equations, with

$$u_a = (1, 0, 0, 0) \tag{5.4}$$

and

$$\frac{\kappa\rho}{2\Lambda} = -(4k^2 - 1), \quad \Lambda = -\frac{1}{R^2(1-k^2)}. \tag{5.5}$$

The meaning of (5.4) is that Y_0 is the velocity vector field of the matter. Since $Y_0 = \partial/\partial\tilde{x}^0$ [see (4.10)], it follows that in (4.12) the $\tilde{x}^0$ lines are the world lines of the matter, as stated earlier.

In order to investigate the motion of the matter, we have to integrate the equations[8]

$$L_{Y_0}(Y) = \nabla_{Y_0}(Y) - \nabla_Y(Y_0), \tag{5.6}$$

where Y is perpendicular to Y_0, that is,

$$Y = \eta^a Y_a, \tag{5.7}$$

the summation extending over 1, 2, 3.

The vector Y is a vector perpendicular to a particle geodesic, and the tip of its arrow is in the neighboring particle geodesic.

Substituting (5.7) into (5.6), applying the rules of the covariant derivation, and introducing the notation

$$\dot{\eta}^a = Y_0\eta^a,$$

we obtain the equations

$$\dot{\eta}^a = C^a{}_{b0}\eta^b. \tag{5.8}$$

Using (5.1), we get

$$\begin{aligned} \dot{\eta}^1 &= \frac{1-k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}\eta^2, \\ \dot{\eta} &= -\frac{1+k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}\eta^1, \\ \dot{\eta}^3 &= 0. \end{aligned} \tag{5.9}$$

These equations describe the motion of the matter with respect to the 3-dimensional vector frame of the Y_a. As a consequence of these equations, we have

$$(1+k)\dot{\eta}^1\eta^1 + (1-k)\dot{\eta}^2\eta^2 = 0,$$

and by integration we obtain

$$[\eta^1/(1-k)^{\frac{1}{2}}]^2 + [\eta^2/(1+k)^{\frac{1}{2}}]^2 = A^2, \quad \eta^3 = B, \tag{5.10}$$

as the equation of the orbit for the neighboring particle. The orbits of the particles in the Y_a frame are, therefore, ellipses in the (Y_1, Y_2) plane. The main axes of the ellipses are in the Y_1 and Y_2 directions. The axes of the ellipse rotate around Y_3 with respect to the inertial compass. To see that, we determine the motion of the frame Y_a along the world lines of the matter. Using the formula

$$\dot{Y}_a \equiv \nabla_{Y_0}(Y_a) = \Gamma_{0a}{}^b\, Y_b$$

and (5.2), we obtain the equations

$$\begin{aligned} \dot{Y}_0 &= 0, \\ \dot{Y}_1 &= \frac{1-2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}} Y_2, \\ \dot{Y}_2 &= -\frac{1-2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}} Y_1, \\ \dot{Y}_3 &= 0. \end{aligned} \tag{5.11}$$

The content of these equations is that Y^3 is parallel propagated along the $\tilde{x}^0$ lines and Y_1 and Y_2 and that the axes of the ellipse (5.10) are rotating around Y_3 with respect to the parallel propagated frame.[9] The angular velocity of this rotation is given by

$$\omega_{Y\ \mathrm{frame}} = \frac{1-2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}.$$

This gives a characterization for the frame Y_0, Y_1, Y_2, Y_3 by the motion of the matter.

Another way to bring Y_1, Y_2, Y_3 in connection with the motion of the matter is to decompose the tensor field

$$U(X, Z) = (\nabla_X\theta^0)(Z) = X\theta^0(Z) - \theta^0(\nabla_X(Z))$$

into symmetric and skew-symmetric parts (θ^0 is the covariant tensor field, 1-form, corresponding to Y_0). The components of the tensor field U are given by

$$U_{ab} = U(Y_a, T_b) = -\Gamma_{ab}{}^0 \quad \text{[see (2.9) and (2.10)].}$$

The symmetric part, the tensor of shear σ, has the nonvanishing components

$$\sigma_{12} = \sigma_{21} = \frac{k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}. \tag{5.12}$$

The skew-symmetric part, the tensor of rotation w, has the nonvanishing components

$$w_{12} = -w_{21} = \frac{2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}. \tag{5.13}$$

The nonvanishing component of the rotation vector V, defined by

$$v^a = -\tfrac{1}{2}\eta^{abcd}u_a w_{cd}, \tag{5.14}$$

is given by

$$v^3 = \frac{2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}. \tag{5.15}$$

Therefore, writing the tensor fields in contravariant form, we obtain

$$\sigma = \frac{k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}(Y_1 \otimes Y_2 + Y_2 \otimes Y_1) \tag{5.16}$$

or

$$\sigma = \frac{k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}}[2^{-\frac{1}{2}}(Y_1+Y_2) \otimes 2^{-\frac{1}{2}}(Y_1+Y_2) - 2^{-\frac{1}{2}}(Y_1-Y_2) \otimes 2^{-\frac{1}{2}}(Y_1-Y_2)]; \quad (5.17)$$

that is, the eigenvalues of σ are

$$\pm\frac{k}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}} \quad (5.18)$$

and the corresponding eigenvectors are

$$2^{-\frac{1}{2}}(Y_1 \pm Y_2). \quad (5.19)$$

The vector of rotation is given by

$$V = \frac{2k^2}{R}\left(\frac{2}{(1-k^2)(4k^2-1)}\right)^{\frac{1}{2}} Y_3. \quad (5.20)$$

For no value of k, $\frac{1}{2} < |k| \leq 2^{-\frac{1}{2}}$, is the shear or rotation vanishing. The Gödel cosmos is therefore not contained in Class II. This concludes the characterization of the frame Y_0, Y_1, Y_2, Y_3 by the motion of the matter.

Using (2.36), we can calculate the components of the curvature tensor field and the Weyl tensor field C, which turns out to be of Type I and can be given by the nonvanishing components

$$\begin{aligned} C_{2323} &= -C_{1010} = \frac{2k^2}{3R^2(1-k^2)}, \\ C_{3131} &= -C_{2020} = \frac{2k^2}{3R^2(1-k^2)}, \\ C_{1212} &= -C_{3030} = -\frac{4k^2}{3R^2(1-k^2)}, \\ C_{2310} &= -C_{3120} = \frac{2k}{R^2(1-k^2)}[\tfrac{1}{2}(1-2k^2)]^{\frac{1}{2}}, \end{aligned} \quad (5.21)$$

defined by

$$C_{abcd} = C(W_a, W_b, W_c, W_d), \quad (5.22)$$

where

$$\begin{aligned} W_0 &= \frac{1}{(4k^2-1)^{\frac{1}{2}}} Y_0 - \left(\frac{2(1-2k^2)}{4k^2-1}\right)^{\frac{1}{2}}, \\ W_1 &= 2^{-\frac{1}{2}}(Y_1 - Y_2), \quad W_2 = 2^{-\frac{1}{2}}(Y_1+Y_2), \\ W_3 &= -\left(\frac{2(1-2k^2)}{4k^2-1}\right)^{\frac{1}{2}} Y_0 + \frac{1}{(4k^2-1)^{\frac{1}{2}}} Y_3. \end{aligned} \quad (5.23)$$

We notice that the Weyl vector fields W_0, W_1, W_2, and W_3 can be regarded as the eigenvector fields of the tensor of shear [see (5.13)]. W_0 and W_3 belong to the eigenvalue zero.

6. GEOMETRY OF THE SOLUTION

Consider the basis Z_0, Z_1, Z_2, Z_3 defined by (4.26). The Lie algebra of M_4 is given in this basis by the following commutation relations:

$$\begin{aligned} [Z_1, Z_2] &= -\frac{2}{R}\frac{1+2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}} Z_0, \\ [Z_2, Z_0] &= \frac{2}{R}\frac{1-k}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}} Z_1, \\ [Z_0, Z_1] &= \frac{2}{R}\frac{1+k}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}} Z_2, \\ [Z_0, Z_3] &= 0, \\ [Z_1, Z_3] &= -\frac{1+k}{R}\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}} Z_2, \\ [Z_2, Z_3] &= +\frac{1-k}{R}\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}} Z_1. \end{aligned} \quad (6.1)$$

The first three commutation relations show that Z_0, Z_1, Z_2 form together the basis of a 3-dimensional subalgebra of the Lie algebra of M_4. The second three commutation relations indicate that the subalgebra is an ideal. This ideal generates H^3.

Using (2.32), we compute the components of the affine connection

$$\begin{aligned} \Gamma_{012} &= -\frac{1}{R}\frac{1-2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}}, \\ \Gamma_{123} &= -\frac{k}{R}\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}}, \\ \Gamma_{120} &= -\frac{1}{R}\frac{1+2k+2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}}, \\ \Gamma_{231} &= \frac{k}{R}\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}}, \\ \Gamma_{210} &= \frac{1}{R}\frac{1-2k+2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}}, \\ \Gamma_{312} &= -\frac{1}{R}\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}}. \end{aligned} \quad (6.2)$$

From (6.2) we can read out a bit of geometry. Since

$$\nabla_{Z_a}(Z_a) = \Gamma_{aa}{}^{b} Z_b = 0, \quad a = 0, 1, 2, 3,$$

it follows that the vector fields Z_0, Z_1, Z_2, Z_3 are geodesic. Denoting $\nabla_{Z_3}(Z_a)$ by $\dot{Z}_a$, we have

$$\begin{aligned} \dot{Z}_0 &= 0, \\ \dot{Z}_1 &= \frac{1}{R}\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}} Z_2, \\ \dot{Z}_2 &= -\left(\frac{2(1-2k^2)}{(1-k^2)(1+2k^2)}\right)^{\frac{1}{2}} Z_1, \\ Z_3 &= 0. \end{aligned} \quad (6.3)$$

Therefore, it follows that Z_0 is parallel-propagated along the $\bar{x}^3$ lines and Z_1 and Z_2 rotate around Z_0 with respect to the parallel-propagated frame. What is the geometrical meaning of Z_0, Z_1, Z_2? These three vector fields are tangential to the 3-spaces $\bar{x}^3 = \text{const}$. Looking into the geometry of these 3-spaces, we see that the components of the affine connection and the Ricci tensor field are given by

$$\Gamma_{012} = -\frac{1}{R}\frac{1-2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}},$$
$$\Gamma_{120} = -\frac{1}{R}\frac{1+2k+2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}}, \quad (6.4)$$
$$\Gamma_{201} = -\frac{1}{R}\frac{1-2k+2k^2}{[(1-k^2)(1+2k^2)]^{\frac{1}{2}}},$$

and

$$R_{ab} = \text{diag}\left(-\frac{2(1+2k+2k^2)(1-2k+2k^2)}{R^2(1-k^2)(1+2k^2)}, \frac{2(1-2k^2)(1-2k+2k^2)}{R^2(1-k^2)(1+2k^2)}, \frac{2(1-2k^2)(1+2k+2k^2)}{R^2(1-k^2)(1+2k^2)}\right), \quad (6.5)$$

respectively; therefore, Z_0, Z_1, Z_2 are the eigenvector fields of the Ricci tensor field. The eigenvalues do not depend on $\bar{x}^3$ [see (6.5)]. The nonvanishing components of the curvature tensor field are

$$R_{1212} = \frac{(1-12k^4)}{R^2(1-k^2)(1+2k^2)},$$
$$R_{2020} = -\frac{(1+2k-2k^2)^2}{R^2(1-k^2)(1+2k^2)}, \quad (6.6)$$
$$R_{0101} = -\frac{(1-2k-2k^2)^2}{R^2(1-k^2)(1+2k^2)}$$

[see (2.36)]. These are also independent of $\bar{x}^3$.

Introducing the notations

$$\rho_0 = 1 - 12k^4,$$
$$\rho_1 = -(1+2k-2k^2)^2 = -4[k-\tfrac{1}{2}(1+\sqrt{3})]^2[k-\tfrac{1}{2}(1-\sqrt{3})]^2, \quad (6.7)$$
$$\rho_2 = -(1-2k-2k^2)^2 = -4[k+\tfrac{1}{2}(1+\sqrt{3})]^2[k+\tfrac{1}{2}(1-\sqrt{3})]^2$$

and constructing Fig. 1, we obtain an impression about the dependence of (6.6) on the parameter k. Changing the sign of k is equivalent to changing the one and two directions. ρ_0 remains unchanged under the switch of sign in k.

The geometric meaning of (6.6) is as follows: R_{1212}, R_{2020}, R_{0101} are the Gaussian curvatures of the geodesic surfaces spanned by the vectors Z_1Z_2, Z_2Z_0, and Z_0Z_1, respectively, at the point in question.

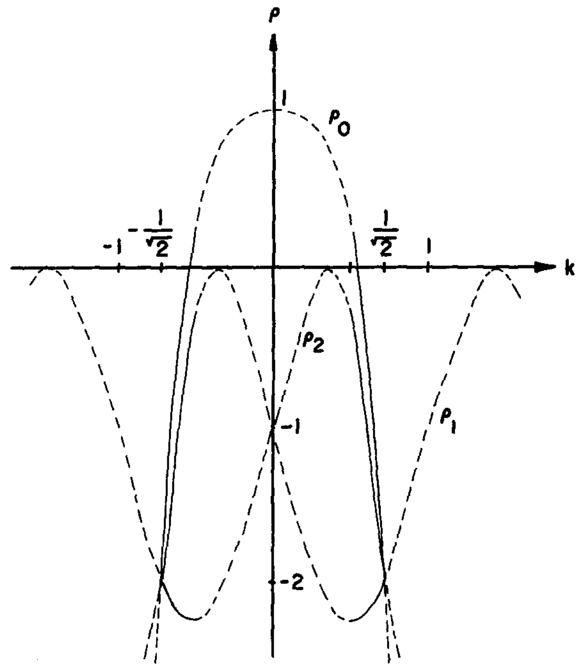


FIG. 1. ρ_0, ρ_1, ρ_2 are proportional to the Gaussian curvature of the geodesic 2-surfaces spanned by Z_1Z_2, Z_2Z_0, and Z_0Z_1, respectively.

From this it follows that the geometry of the hypersurfaces

$$ds^2 = (\tfrac{1}{2}R)^2[(1+2k^2)(\bar{\omega}^0)^2 - (1-k) \times (\bar{\omega}^1\cos\beta\bar{x}^3 + \bar{\omega}^2\sin\beta\bar{x}^3)^2 - (1+k) -(-\bar{\omega}^1\sin\beta\bar{x}^3 + \bar{\omega}^2\cos\beta\bar{x}^3)^2]$$

[see (4.28)] is independent of $\bar{x}^3$ and is given by the geometry of the space

$$ds^2 = (\tfrac{1}{2}R)^2[(1+2k^2)(\bar{\omega}^0)^2 - (1-k)(\bar{\omega}^1)^2 - (1+k)(\bar{\omega}^2)^2]. \quad (6.8)$$

We can think of the space-time (4.28) as a 1-parametric family of 3-dimensional hypersurfaces—$\bar{x}^3$ being the parameter. These hypersurfaces are generated by H^3 and all have the geometry of (6.8). The $\bar{x}^3$ lines are perpendicular to these hypersurfaces, which are embedded in (4.28) such that the Z_1 and Z_2 directions rotate around Z_0 as we move along the $\bar{x}^3$ lines. This is the geometrical content of (6.3). The $\bar{x}^3$ lines are spacelike; therefore, no physical observer can actually move along them.

It is probably interesting to point out the difference, or similarity between the Class I[1] and the Class II universes. We can think of the Class I universes as a 1-parametric family of 3-dimensional hypersurfaces—t being the parameter. These hypersurfaces are generated by S^3, and all have the geometry of

$$ds^2 = -(\tfrac{1}{2}R)^2[(1-k)(\bar{\omega}^1)^2 + (1+k)(\bar{\omega}^2)^2 + (1+2k^2)(\bar{\omega}^3)^2].$$

(The range of k is $|k| < \frac{1}{2}$ and

$$d\bar{\omega}^1 = -\bar{\omega}^2 \wedge \bar{\omega}^3,$$
$$d\bar{\omega}^2 = -\bar{\omega}^3 \wedge \bar{\omega}^1,$$
$$d\bar{\omega}^3 = -\bar{\omega}^1 \wedge \bar{\omega}^2.)$$

The t lines are perpendicular to these hypersurfaces, which are embedded in the Class I universes such that the Z_1 and Z_2 directions rotate around Z_3 as we move along the t lines. The t lines are timelike; therefore, physical observers can actually move along them.

We return now to the discussion of the Class II universes. One sees from Fig. 1 that the geometry of (6.8) is very simple at

$$k = 2^{-\frac{1}{2}}.$$

Our metric is then a special case of Class III universes, as we shall see later.

In order to verify our observation at the end of Sec. 4, we compute the components of the Ricci tensor field with respect to the Z's. Using (2.37) and (6.2), we obtain the following nonvanishing components:

$$\begin{aligned} R_{00} &= -\frac{2}{R^2}\frac{1+4k^4}{(1-k^2)(1+2k^2)}, \\ R_{03} &= \frac{4k^2}{R^2}\frac{[2(1-2k^2)]^{\frac{1}{2}}}{(1-k^2)(1+2k^2)}, \\ R_{11} &= R_{22} = \frac{2}{R^2}\frac{1-2k^2}{1-k^2}, \\ R_{33} &= \frac{4k^2}{R^2}\frac{1-2k^2}{(1-k^2)(1+2k^2)}. \end{aligned} \tag{6.9}$$

In the case of $k = \frac{1}{2}$, (6.9) takes the form

$$R_{00} = -20/9R^2, \qquad R_{03} = 8/9R^2,$$
$$R_{11} = R_{22} = 12/9R^2, \quad R_{33} = 4/9R^2. \tag{6.10}$$

Using (2.40), one sees that the field equations can be satisfied by

$$u_a = (1, 0, 0, -1),$$
$$\kappa\rho = 8/9R^2, \quad \Lambda = -16/9R^2. \tag{6.11}$$

Since $u^a u_a = 0$, one can interpret this model as filled with radiation having the energy density ρ and a Λ term. With these remarks, we close our discussions of the Class II universes.

7. THE METRIC OF THE CLASS III UNIVERSES

We consider the Lie group $R \times H^3$ as before and impose on it another metric as follows. Let

$$\rho > 0 \quad \text{and} \quad |s| < 1 \tag{7.1}$$

be two real parameters and introduce in the Lie algebra (3.34) a new basis by

$$\begin{aligned} Y_0 &= (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}} X_0, \\ Y_1 &= (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}[2/(1+s)]^{\frac{1}{2}} X_1, \\ Y_2 &= (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}[2/(1-s)]^{\frac{1}{2}} X_2, \\ Y_3 &= (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}} X_3, \end{aligned} \tag{7.2}$$

and we define the metric on M_4 by demanding that Y_0, Y_1, Y_2, Y_3 be pseudo-orthonormal, that is, that

$$g(Y_a, Y_b) = g_{ab} = \operatorname{diag}(+1, -1, -1, -1). \tag{7.3}$$

In other words, we define the line element to be

$$ds^2 = (\theta^0)^2 - (\theta^1)^2 - (\theta^2)^2 - (\theta^3)^2, \tag{7.4}$$

where $\theta^0, \theta^1, \theta^2, \theta^3$ form the corresponding basis of the left-invariant 1-forms, that is,

$$\begin{aligned} \theta^0 &= (2/\kappa\rho)^{\frac{1}{2}}\omega^0, \\ \theta^1 &= (2/\kappa\rho)^{\frac{1}{2}}[\tfrac{1}{2}(1+s)]^{\frac{1}{2}}\omega^1, \\ \theta^2 &= (2/\kappa\rho)^{\frac{1}{2}}[\tfrac{1}{2}(1-s)]^{\frac{1}{2}}\omega^2, \\ \theta^3 &= (2/\kappa\rho)^{\frac{1}{2}}\omega^3. \end{aligned} \tag{7.5}$$

After trivial substitutions, we obtain

$$\begin{aligned} ds^2 = (2/\kappa\rho)[(\omega^0)^2 &- \tfrac{1}{2}(1+s)(\omega^1)^2 \\ &- \tfrac{1}{2}(1-s)(\omega^2)^2 - (\omega^3)^2], \end{aligned} \tag{7.6}$$

where the ω's satisfy (3.35) and, if we use the coordinate system introduced before, can be given by (3.27) and (3.33). All the physical and geometrical investigation can be carried out in the frame of the Y's given by (7.2) and in the coordinate system introduced in Sec. 3, since the two different frames and the three different coordinate systems introduced in the case of the Class II universes coincide here, due to the simplicity of the line element (7.6).

At $s = 0$ we have the Gödel cosmos as we see in Sec. 8.

8. MISCELLANEOUS RESULTS AND THE MOTION OF THE MATTER

Consider the basis Y_0, Y_1, Y_2, Y_3 defined by (7.2). The Lie algebra of M_4 is given in this basis by the following commutation relations:

$$\begin{aligned} [Y_1, Y_2] &= -(\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}\frac{2}{(1-s^2)^{\frac{1}{2}}} Y_0, \\ [Y_2, Y_0] &= (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}\frac{1+s}{(1-s^2)^{\frac{1}{2}}} Y_1, \\ [Y_0, Y_1] &= (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}\frac{1-s}{(1-s^2)^{\frac{1}{2}}} Y_2, \\ [Y_a, Y_3] &= 0, \quad a = 0, 1, 2. \end{aligned} \tag{8.1}$$

Using (2.32), we compute the components of the affine connection

$$\Gamma_{012} = 0,\quad \Gamma_{120} = -(\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}\frac{1-s}{(1-s^2)^{\frac{1}{2}}},\quad \Gamma_{201} = -(\tfrac{1}{2}\chi\rho)^{\frac{1}{2}}\frac{1+s}{(1-s^2)^{\frac{1}{2}}}. \tag{8.2}$$

Using (2.37), we find that the only nonvanishing component of the Ricci tensor field is given by

$$R_{00} = -\kappa\rho. \tag{8.3}$$

Comparing this with (2.40), we see that

$$\Lambda = -\tfrac{1}{2}\kappa\rho \tag{8.4}$$

and

$$u_a = (1, 0, 0, 0). \tag{8.5}$$

The meaning of (8.5) is that Y_0 is the velocity vector field of the matter. Since $Y_0 = (\frac{1}{2}\kappa\rho)^{\frac{1}{2}}\partial/\partial x^0$, the $t = (2/\kappa\rho)^{\frac{1}{2}}x^0$-lines are the world lines of the matter.

In order to investigate the motion of the matter, we have to integrate Eqs. (5.6). Repeating the same reasoning as in Sec. 5 and using the same notations we find that the equations

$$\dot\eta^1 = (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}\frac{1+s}{(1-s^2)^{\frac{1}{2}}}\eta^2,\quad \dot\eta^2 = -(\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}\frac{1-s}{(1-s^2)^{\frac{1}{2}}}\eta^1,\quad \dot\eta^3 = 0 \tag{8.6}$$

describe the motion of the matter with respect to the 3-dimensional vector frame Y_1, Y_2, Y_3. The orbits of the neighboring particles are given by

$$\left(\frac{\eta^1}{(1+s)^{\frac{1}{2}}}\right)^2 + \left(\frac{\eta^2}{(1-s)^{\frac{1}{2}}}\right)^2 = A^2,\quad \eta^3 = B, \tag{8.7}$$

which are ellipses in the Y_1, Y_2 plane. The main axes of the ellipse lie in the Y_1 and Y_2 directions.

The axes of the ellipse do not rotate, since from the equations

$$\dot Y_a \equiv \nabla_{Y_0}(Y_a) = \Gamma_{0a}{}^b\, Y_b$$

and from (8.2) it follows that

$$\dot Y_0 = 0,\quad \dot Y_1 = 0,\quad \dot Y_2 = 0,\quad \dot Y_3 = 0.$$

Therefore the frame Y_1, Y_2, Y_3 is parallel-propagated along the x^0 lines and can therefore be chosen as the inertial compass. This gives a characterization for the frame Y_0, Y_1, Y_2, Y_3 by the motion of the matter.

Another way to bring Y_1, Y_2, Y_3 in connection with the motion of the matter is to compute the tensor of shear and the vector of rotation. Along the lines explained in Sec. 5 and using the same notation, we find that

$$\sigma = (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}s(1-s^2)^{-\frac{1}{2}}[2^{-\frac{1}{2}}(Y_1+Y_2)\otimes 2^{-\frac{1}{2}}(Y_1+Y_2) - 2^{-\frac{1}{2}}(Y_1-Y_2)\otimes 2^{-\frac{1}{2}}(Y_1-Y_2)], \tag{8.8}$$

that is, the eigenvalues of σ are

$$\pm (\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}s(1-s^2)^{-\frac{1}{2}} \tag{8.9}$$

and the corresponding eigenvectors are

$$2^{-\frac{1}{2}}(Y_1 \pm Y_2). \tag{8.10}$$

The vector of rotation is given by

$$V = -(\tfrac{1}{2}\kappa\rho)^{\frac{1}{2}}s(1-s^2)^{-\frac{1}{2}}Y_3. \tag{8.11}$$

The shear is vanishing for $s = 0$; therefore, we have the Gödel cosmos at this value of the parameter s as we already stated at the end of Sec. 7.

This concludes the characterization of the frame Y_0, Y_1, T_2, Y_3 by the motion of the matter.

Using (2.36), (2.37), (2.39) and the formulas

$$C_{abcd} = R_{abcd} - E_{abcd} - \tfrac{1}{12}Rg_{abcd},$$

where

$$E_{abcd} = \tfrac{1}{2}(S_{ad}g_{bc} - S_{ac}g_{bc} + g_{ad}S_{bc} - g_{ac}S_{bd}),\quad S_{ab} = R_{ab} - \tfrac{1}{4}Rg_{ab}, \tag{8.12}$$

and

$$g_{abcd} = g_{ad}g_{bc} - g_{ac}g_{bd},$$

we can calculate the components of the Weyl tensor field. The nonvanishing components are

$$C_{2323} = -C_{1010} = \tfrac{1}{6}\kappa\rho,\quad C_{3131} = -C_{2020} = \tfrac{1}{6}\kappa\rho,\quad C_{1212} = -C_{3030} = -\tfrac{1}{3}\kappa\rho.$$

This is a type I e^2 Weyl tensor[10] (type D) and Y_0, Y_1, Y_2, Y_3 are the Weyl vectors. We can give for (7.6) a similar description as we gave to (4.28) toward the end of Sec. 6.

We can think of the space-time (7.6) as a 1-parametric family of 3-dimensional hypersurfaces—x^3 being the parameter. These hypersurfaces are generated by H^3, and all have the same geometry. The x^3 lines are perpendicular to these hypersurfaces, which are embedded in (7.6) such that Z_0, Z_1, Z_2 are parallelly propagated along the x^3 lines. The x^3 lines are spacelike.

From this, one sees that (7.6) is intrinsically similar than (4.28). One can use these models to study the motion of rotation within the general theory of relativity.

In a forthcoming paper, we discuss singularly the Class IV universes which then exhaust all the possibilities for homogeneous universes with dust within the framework of general relativity.

ACKNOWLEDGMENTS

I am indebted to A. Trautman for a remark and to G. Cheshier and H. Armstrong for typing the manuscript.

* This work was supported by the Air Force Office of Scientific Research under Grant AF-AFOSR-903-67 and the National Aeronautics and Space Administration under NASA Grant No. NGL 44-004-001.

[1] I. Ozsváth and E. L. Schücking, Ann. Phys. (N.Y.) 55, 166 (1969).
[2] D. L. Farnsworth and R. P. Kerr, J. Math. Phys. 7, 1625 (1966).
[3] I. Ozsváth, J. Math. Phys. 6, 590 (1965).
[4] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962); S. Kobayshi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[5] C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, N.J., 1946); Vol. I; G. Hochschild, *The Structure of Lie Groups* (Holden-Day, San Francisco, 1965).
[6] K. Gödel, Rev. Mod. Phys. 21, 447 (1949).
[7] H. Flanders, *Differential Forms* (Academic, New York, 1963).
[8] F. A. E. Pirani, *Introduction to Gravitational Radiation Theory, Brandeis Lectures* (Prentice-Hall, Englewood Cliffs, N.J., 1965), p. 264.
[9] K. Gödel, Proc. Int. Cong. Math. 1, 175 (1950).
[10] P. Jordan, J. Ehlers, and W. Kundt, Abhandl. Akad. Wiss. (Mainz), Math.-Naturw. kl. No. 2 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Unitary Irreducible Representations of $SL(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$. I

E. Angelopoulos
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, Paris 5ème, France

(Received 24 February 1968; Revised Manuscript Received 13 November 1969)

The family of Lie groups $SL(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$ is studied from the mathematical point of view, i.e., their structure and unitary irreducible representations. The first steps in the study of representations are based on the Wigner–Mackey–Bruhat method. The homogeneous spaces and little groups are determined by introducing a (generally) pseudo-Hermitian metric on vector spaces of finite dimension. However, the presence of null vectors does not allow this approach to be an exhaustive one; there are cases where the stabilizer is a semidirect product of a reductive group by a nilpotent one. In such cases, results are available if the reductive subgroup is compact, but it seems that these results can be generalized. The connection is also examined, both in structure and in representations, between this family and creation–annihilation operators. A complete list of inequivalent classes of representations is given in terms of orbits and little groups for $n = 2, 3, 4, 6$.

INTRODUCTION

Since it was proved that compact Lie groups are an incomplete tool for treating the strong-interaction symmetries, several other types of groups have been proposed. Although most of them are semisimple groups, "inhomogeneous" groups have also been proposed.[1–7] Semisimple groups have the advantage of being a convenient tool for which there is an abundant mathematical literature. However, in spite of several hints, we do not know yet the symmetry group; nonsemisimple groups like the inhomogeneous special linear groups, especially those which contain the Poincaré group, may prove to be of great interest, and they appear as possible solutions to the external–internal unification problem. Thus, it is necessary to investigate their structure and their unitary irreducible representations in order to identify quantum numbers. Besides, there are other physical domains where such groups may be useful, for example, the classification of nuclear spectra and the creation–annihilation operators. The physical applications of this family will be treated in a subsequent article.

Here, we are interested in the mathematical study of these groups, i.e., the investigation of their structure and unitary irreducible representations (UIR's). The problem was solved by Wigner[8] for the Poincare group ($n = 2$). Then Mackey[9,10] and Bruhat[11] proved that the method of induced representations can be applied, with some topological restrictions, to any group containing a normal Abelian subgroup. Thus, when dealing with such groups, one needs to separate a vector space into orbits and to find the little groups that leave unchanged a point of each orbit, the choice of this point being arbitrary.

In the case of $G_n = SL(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$,[12] one obtains either reductive little groups or semidirect products of reductive groups by Abelian groups.[13] Explicit results are available for the representations of the first case.[14] The second case is more interesting: Further Wigner decomposition can generally be done only once more, because the next little groups' structure is more complicated (nilpotent subgroups take the place of Abelian ones). The investigation is carried on in some cases where the reductive subgroup is compact, with other methods consisting of the decomposition of the regular representation and the establishment of a complete correspondence between the representations of the reductive subgroup and the faithful representations of the whole semidirect product which occur in the regular representation.

On the other hand, a connection is established between Lie algebras of the family and creation–annihilation operators by means of the symplectic Lie algebra, which is identified to a subalgebra of the universal enveloping algebra generated by these operators. Any algebra of the family is isomorphic to a subalgebra

* This work was supported by the Air Force Office of Scientific Research under Grant AF-AFOSR-903-67 and the National Aeronautics and Space Administration under NASA Grant No. NGL 44-004-001.

[1] I. Ozsváth and E. L. Schücking, Ann. Phys. (N.Y.) 55, 166 (1969).
[2] D. L. Farnsworth and R. P. Kerr, J. Math. Phys. 7, 1625 (1966).
[3] I. Ozsváth, J. Math. Phys. 6, 590 (1965).
[4] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962); S. Kobayshi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[5] C. Chevalley, *Theory of Lie Groups* (Princeton U.P., Princeton, N.J., 1946); Vol. I; G. Hochschild, *The Structure of Lie Groups* (Holden-Day, San Francisco, 1965).
[6] K. Gödel, Rev. Mod. Phys. 21, 447 (1949).
[7] H. Flanders, *Differential Forms* (Academic, New York, 1963).
[8] F. A. E. Pirani, *Introduction to Gravitational Radiation Theory, Brandeis Lectures* (Prentice-Hall, Englewood Cliffs, N.J., 1965), p. 264.
[9] K. Gödel, Proc. Int. Cong. Math. 1, 175 (1950).
[10] P. Jordan, J. Ehlers, and W. Kundt, Abhandl. Akad. Wiss. (Mainz), Math.-Naturw. kl. No. 2 (1960).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Unitary Irreducible Representations of $SL(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$. I

E. Angelopoulos
Institut Henri Poincaré, 11 rue Pierre et Marie Curie, Paris 5ème, France

(Received 24 February 1968; Revised Manuscript Received 13 November 1969)

The family of Lie groups $SL(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$ is studied from the mathematical point of view, i.e., their structure and unitary irreducible representations. The first steps in the study of representations are based on the Wigner–Mackey–Bruhat method. The homogeneous spaces and little groups are determined by introducing a (generally) pseudo-Hermitian metric on vector spaces of finite dimension. However, the presence of null vectors does not allow this approach to be an exhaustive one; there are cases where the stabilizer is a semidirect product of a reductive group by a nilpotent one. In such cases, results are available if the reductive subgroup is compact, but it seems that these results can be generalized. The connection is also examined, both in structure and in representations, between this family and creation–annihilation operators. A complete list of inequivalent classes of representations is given in terms of orbits and little groups for $n = 2, 3, 4, 6$.

INTRODUCTION

Since it was proved that compact Lie groups are an incomplete tool for treating the strong-interaction symmetries, several other types of groups have been proposed. Although most of them are semisimple groups, "inhomogeneous" groups have also been proposed.[1–7] Semisimple groups have the advantage of being a convenient tool for which there is an abundant mathematical literature. However, in spite of several hints, we do not know yet the symmetry group; nonsemisimple groups like the inhomogeneous special linear groups, especially those which contain the Poincaré group, may prove to be of great interest, and they appear as possible solutions to the external–internal unification problem. Thus, it is necessary to investigate their structure and their unitary irreducible representations in order to identify quantum numbers. Besides, there are other physical domains where such groups may be useful, for example, the classification of nuclear spectra and the creation–annihilation operators. The physical applications of this family will be treated in a subsequent article.

Here, we are interested in the mathematical study of these groups, i.e., the investigation of their structure and unitary irreducible representations (UIR's). The problem was solved by Wigner[8] for the Poincare group ($n = 2$). Then Mackey[9,10] and Bruhat[11] proved that the method of induced representations can be applied, with some topological restrictions, to any group containing a normal Abelian subgroup. Thus, when dealing with such groups, one needs to separate a vector space into orbits and to find the little groups that leave unchanged a point of each orbit, the choice of this point being arbitrary.

In the case of $G_n = SL(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$,[12] one obtains either reductive little groups or semidirect products of reductive groups by Abelian groups.[13] Explicit results are available for the representations of the first case.[14] The second case is more interesting: Further Wigner decomposition can generally be done only once more, because the next little groups' structure is more complicated (nilpotent subgroups take the place of Abelian ones). The investigation is carried on in some cases where the reductive subgroup is compact, with other methods consisting of the decomposition of the regular representation and the establishment of a complete correspondence between the representations of the reductive subgroup and the faithful representations of the whole semidirect product which occur in the regular representation.

On the other hand, a connection is established between Lie algebras of the family and creation–annihilation operators by means of the symplectic Lie algebra, which is identified to a subalgebra of the universal enveloping algebra generated by these operators. Any algebra of the family is isomorphic to a subalgebra of some symplectic algebra: One can thus establish both structural and spectral connections among infinitesimal operators. The application of these results to field theory will be treated in a subsequent article.

Section I deals with the group structure and *TCP* generalization; the action of these operators is studied both in abstract group structure and in representations. The study of UIR's of proper G_n begins in Sec. II, where Wigner's method is used in two successive steps. In Sec. III, special cases of decomposition are examined, where Wigner's method can be applied. Section IV deals with cases in which difficulties due to nilpotent subgroups occur. Creation–annihilation operators are treated in Sec. V.

It is important to note the following: To avoid confusion, we specify that, when we speak of representations of a Lie group, we mean UIR's unless the contrary is specified; the description of representations is given up to those of reductive little groups, such as $U(p, q)$ or $GL(r, \mathbf{C})$, or their direct products.[14]

I. GENERALITIES (TRIVIAL STRUCTURAL REMARKS)

A. G_n and Poincaré Subgroups

The group G_n is the semidirect product of the unimodular complex linear group $SL(n, \mathbf{C})$ by the Abelian group $\mathbf{R}(n^2)$ of n^2 real translations. The action of $SL(n, \mathbf{C})$ on $\mathbf{R}(n^2)$ is

$$(A)_{\alpha\beta} \to \Lambda^{\gamma}_{\alpha}\bar{\Lambda}^{\delta}_{\beta}(A)_{\gamma\delta}, \quad \Lambda \in SL(n, \mathbf{C}), \quad A \in \mathbf{R}(n^2).$$

In matrix notation, the generic element of G_n can be written

$$g = \begin{pmatrix} \Lambda & A\Lambda^{*-1} \\ 0 & \Lambda^{*-1} \end{pmatrix},$$

where Λ and A are $n \times n$ complex matrices, Λ is unimodular, and A is Hermitian (or A is skew-Hermitian). The group law is

$$(A, \Lambda)(A', \Lambda') = (A + \Lambda A'\Lambda^*, \Lambda\Lambda').$$

There are many ways to choose a subgroup of G_n isomorphic to the universal covering of the inhomogeneous Lorentz group $\mathcal{T}$. However, we shall immediately put in evidence the most interesting choice in the case $n = 2n'$. In fact, $\mathbf{R}(4n^2) = \mathbf{R}(4) \times \mathbf{R}(n'^2)$, that is, every Hermitian matrix of $2n'$ lines and $2n'$ columns can be written as a finite sum of the generators $\sigma_\mu \times \alpha_i$, where α_i is an $n' \times n'$ Hermitian matrix and σ_μ is an ordinary Pauli matrix. Similarly, the matrices $\lambda \times \lambda'$ [where $\lambda \in \mathfrak{gl}(2, \mathbf{R})$, $\lambda' \in \mathfrak{gl}(n', \mathbf{R})$ and such that $Sp\lambda \cdot Sp\lambda' = 0$] span $\mathfrak{sl}(2n', \mathbf{R})$. The complexification of this algebra gives $\mathfrak{sl}(2n', \mathbf{C})$ and, if one puts the equivalence relations

$$(i\lambda) \times (\lambda') = (\lambda) \times (i\lambda') = i(\lambda \times \lambda'),$$
$$(i\lambda) \times (i\lambda') = -\lambda \times \lambda',$$

one finds again the commutation relations of $\mathfrak{g}_{2n}$, by using those of $\mathfrak{g}_2$ and $\mathfrak{g}_{n'}$. The Poincaré subalgebra is generated by $\{a_i \times 1, b_i \times 1; \sigma_\mu \times 1\}$, and its commutator in $\mathfrak{gh}(2n', \mathbf{C})$ is the subalgebra isomorphic to $\mathfrak{su}(n')$, spanned by $\{1 \times u\}$. (Here a_i and b_i are generators of the Lorentz Lie algebra.)

B. Involutive External Automorphisms of G_n

In view of further physical applications, it is important to investigate whether our group G_n can be imbedded into a larger group $\tilde{G}_n$ containing "parity," that is, whether $\tilde{G}$ is an extension of $\mathbf{Z}_2$ by G, since we require the operator parity to be an involution and to correspond to the physical parity for the Poincaré group ($n = 2$), or a particular Poincaré subgroup of G_n. For this purpose, we use the method introduced in Ref. 6; $\tilde{G}$ is a semidirect product of $\mathbf{Z}_2 = \{1, s\}$ by G. The operator parity is $P = \mathrm{Ad}\, s$ and it induces an automorphism on the Lie algebra $\mathfrak{g}$. Therefore, we investigate the different possibilities for $\mathrm{Ad}\, s \in \mathrm{Aut}\, \mathfrak{g}$. In order to determine $\mathrm{Aut}\, \mathfrak{g}$ (by a method similar to that used in Ref. 15), we first remark that $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{t}$ (considered as vector spaces), with $\mathfrak{k} = \mathfrak{sl}(n, \mathbf{C})$, $\mathfrak{t} = \mathbf{R}(n^2)$, $\mathfrak{t}$ being also a characteristic ideal of $\mathfrak{g}$. We can thus write, for every $F \in \mathrm{Aut}\, \mathfrak{g}$,[16]

$$F = (f, \Phi, \psi),$$

with

$$\begin{aligned} f &\in \mathrm{Aut}\, \mathfrak{t} = GL(n^2, \mathbf{R}), \\ \Phi &\in \mathrm{Aut}\, \mathfrak{k} = (SL(n, \mathbf{C})/\mathbf{Z}_n) \cdot (\mathbf{Z}_2 \times \mathbf{Z}_2) \\ &= \mathrm{Int}\, \mathfrak{k} \cdot (\mathbf{Z}_2 \times \mathbf{Z}_2), \\ \psi &\in \mathfrak{L}(\mathfrak{k}; \mathfrak{t}) = \mathfrak{L}(\mathbf{R}(2n^2 - 2); \mathbf{R}(n^2)). \end{aligned}$$

The group law establishes a relation between Φ and f: Let $p \in \mathrm{Hom}\,(\mathfrak{sl}(n, \mathbf{C}); \mathfrak{gl}(n^2, \mathbf{R}))$ be such that $p(K) \cdot T = KT + TK^*$; f must belong to the normalizer N of $p(\mathfrak{k})$ in $GL(n^2, \mathbf{R})$, because

$$f \cdot p(K) \cdot f^{-1} = p(\Phi(K));$$

thus, f defines Φ as an automorphism of $\mathfrak{k}$. Calculating that, if $n > 2$, the external automorphism $K \to -K^*$ cannot be expressed by any $\mathrm{Ad}\, f \in N$, we easily check that N is the direct product of $\mathrm{Aut}\, \mathfrak{k}/\mathbf{Z}_2$ by the multiplicative group of real numbers.

Using the group law again, one can prove that

$$\psi(S) = p(\Phi(S)) \cdot A = \Phi(S)A + A[\Phi(S)]^*,$$

where A is a fixed $n \times n$ matrix. One can thus write

$$\mathrm{Aut}\, \mathfrak{g} = (\mathbf{R}^* \times \mathrm{Aut}\, \mathfrak{k}/\mathbf{Z}_2) \cdot \mathbf{R}(n^2) = (\mathbf{R}^* \times \mathbf{Z}_2) \cdot \mathrm{Int}\, \mathfrak{g}$$

and

$$\begin{aligned} F_{\lambda,j,\Lambda,A}(K; T) &= (\Lambda \cdot jK \cdot \Lambda^{-1}; \lambda\Lambda \cdot jT \cdot \Lambda^* \\ &\quad + p(\Lambda \cdot jK \cdot \Lambda^{-1}) \cdot A) \\ &= \mathrm{Ad}\,(-A, \Lambda) \cdot (jK; \lambda jT), \end{aligned}$$

j being either 1 or J, with $JK = \bar{K}$ and $JT = \bar{T}$.

Searching for an involutive automorphism F, we see that λ must be equal to ± 1 and that there always exists an A' such that

$$F_{\pm 1,j,\Lambda,A} = F_{\pm 1,j,\Lambda,0} \cdot F_{1,1,1,A'} .$$

Thus, we can only investigate semisimple automorphisms, which leave globally invariant $\mathfrak{t}$ and $\mathfrak{k} = \text{Ad}\, A'(\mathfrak{k}') \cong \mathfrak{sl}(n, \mathbf{C})$.

We now consider the case $j = 1$. Then $\text{Ad}\, \Lambda^2 = I$ and $\Lambda^2 = \exp(2ik\pi/n) \cdot I$. Λ is semisimple and may always be put into the diagonal form $I_p \oplus (-I_{2q})$ or $\exp(i\pi/n) \cdot [I_p \oplus (-I_{2q+1})]$. Then λ must be -1, since we want P to be external. The positive and negative eigenspaces of P on the semisimple part both have a complex structure, $\mathfrak{k}_+$ being a subalgebra [see (A) below].

In the case $j = J$, we have $\text{Ad}\,(\Lambda\bar{\Lambda}) = I$, and λ may be either 1 or -1; the choice of λ affects only the eigenvalues of the eigenvectors of P belonging to $\mathfrak{t}$, while P defines in $\mathfrak{k}$ an external real form[17] $\mathfrak{k}_+$— [see (B) and (C)].

We now write the results in terms of positive and negative eigenspaces of P:

(A) $\mathfrak{k}_+ = \mathfrak{s}(\mathfrak{gl}(p, \mathbf{C}) + \mathfrak{gl}(q, \mathbf{C}))$,
$\qquad \mathfrak{k}_- = \mathbf{C}(p \times q) + \mathbf{C}(q \times p)$,
$\qquad \mathfrak{t}_+ = \mathbf{C}(p \times q), \qquad \mathfrak{t}_- = \mathbf{R}(p^2) + \mathbf{R}(q^2)$,

(B) $\mathfrak{k}_+ = \mathfrak{sl}(n, \mathbf{R}), \qquad \mathfrak{t}_- = i \cdot \mathfrak{sl}(n, \mathbf{R})$,
$\qquad \mathfrak{t}_\pm = \mathbf{R}(\tfrac{1}{2}n(n+1)), \quad \mathfrak{t}_\pm = \mathbf{R}(\tfrac{1}{2}n(n-1))$,

(C) $\mathfrak{k}_+ = \mathfrak{su}^*(2n'), \qquad \mathfrak{k}_- = i \cdot \mathfrak{su}^*(2n'), \quad 2n' = n$,

$$\mathfrak{t}_\pm = \left\{\begin{pmatrix} H & -A \\ A & \bar{H} \end{pmatrix}\right\}, \quad \mathfrak{t}_\mp = \left\{\begin{pmatrix} H & S \\ S & -\bar{H} \end{pmatrix}\right\},$$

where H, A, and S are $n' \times n'$ complex matrices, with H Hermitian, A skew-symmetric, and S symmetric.

C. Generalized *PCT*

Having defined P, we may as well define (external) operators T and C, in order to generalize the PCT assignment. We impose $PCT = e^{i\varphi}$, C and T both being of square unity and being represented by an antiunitary operator whenever P is represented by a unitary operator. Evidently, we need to define only one of them, say T, C being then defined by $C = PTe^{i\varphi}$. Following again the method of Flato and Sternheimer,[6] we have two choices. In the first one, we take T to be the principal anti-automorphism, which turns $\mathfrak{g}$ to its opposite Lie algebra, having again the possibility of combining it with the involution which has value $+1$ on $\mathfrak{k}$ and -1 on $\mathfrak{t}$. The second choice consists of defining T, not by its action on the algebra's structure, but on the representations of the group $\tilde{G}$. We consider two complex conjugate representations U and $\bar{U}$ of $\tilde{G}$: T is then defined by

$$U(Tg) \cdot \varphi = \bar{U}(g) \cdot \bar{\varphi}, \quad g \in \tilde{G}, \text{ for every vector } \varphi.$$

Both definitions give an antiunitary operator which is not an element of $\tilde{G}$. We have thus given a formal extension of the TCP assignment to the family of groups G_n which contain the inhomogeneous Lorentz group.

D. Representations of $\tilde{G}$

To determine representations of $\tilde{G}$ knowing those of G, let $\check{g} = sgs$. We can thus associate couples of representations of G; that is, with U we associate the representation $\check{U}$, $\check{U}(g) = U(sgs)$. The representations U and $\check{U}$ are, in general, inequivalent. This always happens when U is unitary irreducible and not reduced to unity on the Abelian subgroup. In this case, we proceed as follows: We "double" the Hilbert space of U, and we put

$$U(g) = \begin{pmatrix} U(g) & 0 \\ 0 & U(sgs) \end{pmatrix}, \quad \text{if } g \in G,$$
$$= \begin{pmatrix} 0 & U(gs) \\ U(sg) & 0 \end{pmatrix}, \quad \text{if } f \in sG.$$

Thus, we obtain UIR's of G in which the discrete element s is represented by

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

These representations provide a useful tool for the definition of P, C, and T on known representation spaces of G.

II. ORBITS AND STABILIZERS

A. First Decomposition

The method used here to determine representations of G is that used by Wigner[8] for the Poincaré group. Mackey[9,10] and Bruhat[11] have proved since then that this method can be generalized to any semidirect product of a reductive Lie group by an Abelian one. We thus have to decompose the dual to the Abelian subgroup into orbits and then find the representations of the stabilizer of one point of each orbit.

In case of G_n, the dual $\hat{\mathbf{R}}(n^2)$ is isomorphic to $\mathbf{R}(n^2)$; if both are represented by $n \times n$ Hermitian matrices, the bilinear form is

$$\langle \alpha, A \rangle = Sp\,\alpha A, \quad \alpha \in \hat{\mathbf{R}}(n^2), \quad A \in \mathbf{R}(n^2). \qquad (1)$$

$SL(n, \mathbf{C})$ acts on $\hat{\mathbf{R}}(n^2)$ according to

$$\alpha \cdot \Lambda = \Lambda^* \alpha \Lambda, \quad \Lambda \in SL(n, \mathbf{C}). \qquad (2)$$

It is then easy to verify that an orbit is determined by $|\det \alpha| = M^n$ and the triplet $(p, q, r) = \text{sign}\, \alpha$, the three integers denoting, respectively, the number of positive, negative, and null eigenvalues of α.

Taking for standard element in each orbit the matrix

$$(I_p \oplus (-I_q) \oplus 0_r), \quad \text{if} \quad M = 0,$$

and

$$M \cdot (I_p \oplus (-I_q)), \quad \text{if} \quad M \neq 0,$$

we have the following result.

Proposition 1: The little group corresponding to the orbit $\{M; (p, q, r)\}$ depends only on (p, q, r) and is

$$S(U(p,q) \times GL(r, \mathbf{C})) \cdot \mathbf{C}(r(p+q)).$$

Remark: By $S(G^k \times G'^{k'})$, when speaking of matrix groups, we mean the subgroup of $G \times G'$ determined by $(\det g)^k \cdot (\det g')^{k'} = 1$.

We immediately obtain the following corollaries.

Corollary 1: Every nondegenerate representation of G_n is determined by a positive number M, a partition of n into two integers (p, q), and a representation of $SU(p,q)$.

Corollary 2: Most degenerate representations of G_n are all the representations of $SL(n, \mathbf{C})$.

B. Second Decomposition

When $r(p + q) \neq 0$, the little group G' is not semisimple but is still a semidirect product of a reductive by an Abelian group. So we still apply the same method to determine its representations.

G' is a matrix group: any $g' \in G'$ can be written

$$g' = \begin{pmatrix} u & 0 \\ zu & \mu \end{pmatrix}, \quad u \in U(p,q), \quad \mu \in GL(r, \mathbf{C}),$$

$$z \in \mathfrak{L}(\mathbf{C}^{p+q}; \mathbf{C}^r), \quad \det u \cdot \det \mu = 1.$$

Taking the dual Abelian group isomorphic to be $\mathfrak{L}(\mathbf{C}^r; \mathbf{C}^{p+q})$, we have the bilinear form

$$\langle \zeta, z \rangle = \tfrac{1}{2} Sp(z\zeta + \zeta^* z^*). \tag{3}$$

The action of the homogeneous part on the dual space is

$$\zeta \xrightarrow{g'} u^{-1}\zeta\mu.$$

To decompose $\mathfrak{L}(\mathbf{C}(p + q); \mathbf{C}(r))$ into orbits of G', we observe that $u^{-1}\zeta\mu$ denotes the same linear mapping as ζ, after changes of bases in $\mathbf{C}(r)$ and $\mathbf{C}(p + q)$, the latter being a (p, q)-pseudo-unitary one. Rank ζ is an invariant. Moreover, there is a metric induced on Im ζ by the (p, q) metric of $\mathbf{C}(p + q)$: There exists a basis of $p_1 + q_1 + s_1$ mutually orthogonal vectors in Im ζ, p_1 of them having a positive square, q_1 a negative one, and s_1 a null one. The triplet (p_1, q_1, s_1) is an invariant.

Remark: The following inequalities obviously hold:

$$\begin{aligned} \text{Rank } \zeta = p_1 + q_1 + s_1 &\leq \min(p + q, r), \\ p_1 + s_1 &\leq p, \\ s_1 &\leq \min(p, q), \\ q_1 + s_1 &\leq q. \end{aligned}$$

Furthermore, if rank $\zeta = r \leq p + q$, we may associate with every ζ a vector ζ_0, the components of which are the minors of rank r of ζ. μ acts on ζ_0 like a scalar of modulus one, and u^{-1} acts according to the rth fundamental (nonunitary) representation of $SU(p,q)$,[18] both keeping invariant a pseudo-Hermitian product in $\mathbf{C}(\binom{p+q}{r})$. The metric tensor is then

$$g(i_1, i_2, \cdots, i_r; j_1, j_2, \cdots, j_r) = g(i_1; j_1) \cdot g(i_2; j_2) \cdots g(i_r; j_r),$$

with

$$i_1 < i_2 < \cdots < i_r.$$

There are no other invariants characterizing the orbits. We can thus state a theorem.

Theorem 1: Orbits of $\mathfrak{L}(\mathbf{C}(r); \mathbf{C}(p + q))$ are characterized by sign $\zeta = (p_1, q_1, s_1)$ and the real positive number $M_0 = |(\zeta_0 \mid \zeta_0)|$. M_0 is always zero if rank $\zeta < r$.

Remark 1: If $p + q = r$, $M_0 = |\det \zeta|^2$.

Remark 2: It is easy to check that orbits which differ only on M_0 have isomorphic stabilizers.

To determine the little group, we first find out its Lie algebra. Putting $u = \exp U$ and $\mu = \exp M$, we find

$$\begin{bmatrix} V_1 & -E^* & +E^* & 0 \\ EI_1 & \tfrac{1}{2}(\Lambda_1 - \Lambda_1^*) + iH & \tfrac{1}{2}(\Lambda_1 + \Lambda_1^*) - iH & FI_1' \\ EI_1 & \tfrac{1}{2}(\Lambda_1 + \Lambda_1^*) + iH & \tfrac{1}{2}(\Lambda_1 - \Lambda^*) - iH & FI_1' \\ 0 & -F^* & F^* & V_1' \end{bmatrix} \qquad M = \begin{bmatrix} V_1 & 0 & 0 \\ E & \Lambda_1 & 0 \\ X & Y & M_1 \end{bmatrix}$$

with

$$\begin{aligned} &V_1 \in \mathfrak{su}(p_1, q_1), && E \in \mathbf{C}((p_1 + q_1)s_1), \\ &V_1' \in \mathfrak{su}(p_1', q_1'), && F \in \mathbf{C}((p_1' + q_1')s_1), \\ &\Lambda_1 \in \mathfrak{gl}(s_1, \mathbf{C}), && X \in \mathbf{C}((p_1 + q_1)r_1), \\ &M_1 \in \mathfrak{gl}(r_1, \mathbf{C}), && Y \in \mathbf{C}(s_1 \cdot r_1), \\ & && H \in \mathbf{R}(s_1^2)\ \text{(Hermitian)}, \end{aligned}$$

$$I_1 = I(p_1) \oplus (-I(q_1)), \quad I_1' = I(p_1') \oplus (-I(q_1')),$$

$$p_1' = p - p_1 - s_1, \quad q_1' = q - q_1 - s_1,$$

$$r_1 = r - p_1 - s_1 - q_1,$$

$$2SpV_1 + SpV_1' + Sp(2\Lambda_1 - \Lambda_1^*) + SpM_1 = 0.$$

On the other hand, it is easy to find the discrete center of the little group; it is $\mathbf{Z}_n$.

Theorem 2: The stabilizer G'' of the orbit (p_1, q_1, s_1) of G' is the inessential extension of a reductive group R by a nilpotent one N; R is the group

$$S(U^2(p_1, q_1) \times U(p_1', q_1') \times GL^{2;1}(s_1; \mathbf{C}) \times GL(r_1; \mathbf{C}));$$

N is a central extension of the Abelian group A by the Abelian group B such that

$$\begin{aligned} A &= \mathfrak{L}(\mathbf{C}(p_1 + q_1), \mathbf{C}(s_1)) \\ &\quad \otimes \mathfrak{L}(\mathbf{C}(p_1' + q_1'), \mathbf{C}(s_1)) \otimes \mathfrak{L}(\mathbf{C}(s_1), \mathbf{C}(r_1)), \\ B &= \mathfrak{L}(\mathbf{C}(p_1 + q_1), \mathbf{C}(r_1)) \otimes \mathbf{R}(s_1^2). \end{aligned}$$

The group law in N is

$$\begin{aligned} n \cdot n' &= (H, X, Y, E, F)(H', X', E', F') \\ &= (H + H' + \tfrac{1}{2}(EI_1E'^* - E'I_1E^*) \\ &\quad + \tfrac{1}{2}(FI_1'F'^* - F'I_1'F^*), X + X' \\ &\quad + \tfrac{1}{2}(Y E' - Y'E), Y + Y', E + E', F + F'). \end{aligned}$$

The law in G'', expressed as a semidirect product, is

$$\begin{aligned} &(v_1, v_1', \lambda_1, \mu_1) \cdot (H, X, Y, E, F) \\ &\quad = (\lambda_1 H \lambda_1^*, \mu_1 X v_1^{-1}, \mu_1 Y \lambda_1^{-1}; \lambda_1 E v_1^{-1}, \lambda_1 F v_1'^{-1}) \\ &\quad\quad \cdot (v_1, v_1', \lambda_1, \mu_1). \end{aligned}$$

Remark: The exponent 2;1 means that, in the determinant condition, one must take $\det(\lambda^2 \lambda^{*-1})$.

C. Third Decomposition

To study representations of group G'' is much more tedious (in the general case) than in previous decompositions, because the normal subgroup of the semidirect product is no longer Abelian but, generally, nilpotent. The extension structure of G'' is expressed in this diagram:

$$\begin{array}{ccccccccc} & & 1 & & 1 & & 1 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 1 & \longrightarrow & B & \longrightarrow & N & \longrightarrow & A & \longrightarrow & 1 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 1 & \longrightarrow & B & \longrightarrow & G'' & \longrightarrow & G''/B & \longrightarrow & 1 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 1 & \longrightarrow & R & \longrightarrow & R & \longrightarrow & 1 \\ & & & & \downarrow & & \downarrow & & \\ & & & & 1 & & 1 & & \end{array}$$

The maximal Abelian normal subgroup is, in general, $B \otimes \mathfrak{L}(\mathbf{C}(s_1); \mathbf{C}(r_1))$; the action of G'' on the dual is

$$\begin{pmatrix} \hat{H} \\ \hat{X} \\ \hat{Y} \end{pmatrix} \to \begin{pmatrix} \lambda_1^* \hat{H} \lambda_1 \\ u_1^{-1} \hat{X} \mu_1 \\ \lambda_1^{-1}(\hat{Y} - E \cdot \hat{X})\mu_1 \end{pmatrix}.$$

Orbits of $\hat{H}$ and $\hat{X}$ have been calculated in Secs. II.A and II.B. There is, however, a difference: when rank $\hat{H} = s_1$ and rank $\hat{X} = r_1$, no determinants nor pseudonorms are conserved, since λ_1 and μ_1 are not unimodular matrices but only obey $|\det \lambda_1 \cdot \det \mu_1| = 1$. Thus, we have only the following invariant quantity:

$$\|\hat{H}\| \times \|\hat{X}\|^2,$$

where $\|\hat{H}\| = |\det \hat{H}|$ and $\|\hat{X}\|$ is the $\binom{p_1+q_1}{s_1}$ pseudonorm of $\hat{X}$ as defined in (2).

Orbits of $\hat{Y}$ depend on those of $\hat{X}$. Let $\mathbf{C}(r_1) = E_1 \oplus E_0$ with $E_0 = \operatorname{Ker} \hat{X}$; then $\hat{Y}$ is the sum of its restrictions to E_1 and E_0, i.e., $\hat{Y} = \Psi_1 + \Psi_0$. Then it is easy to see that rank Ψ_0 is the only discrete invariant characterizing orbits of $\hat{Y}$. There is no continuous invariant if rank $\hat{X}$ + rank $\Psi_0 < r_1$. However, such invariants may arise in special cases, as, for instance, when $\hat{X} = 0$ and $s_1 = r_1$, where $\|\hat{H}\hat{Y}\|$ is conserved.

The stabilizers will be of the form $G''' = R' \cdot (AB)$, where R' is a reductive group, determined as before. But here appears the following fact: In the extension, the maximal Abelian subgroup is no longer the second factor of a semidirect product, since its complementary is not a subgroup of G'''. Thus, one cannot consider R' by itself nor $R' \cdot B = G'''/A$ and apply again the Mackey–Bruhat procedure. If one tries to apply it to the whole group G''', one finds the same maximal Abelian normal subgroup, the orbits being points and the stabilizer being G''' itself. So far, no method has been devised that gives all representations of such groups.

We must note, however, that these difficulties arise when n becomes large. Few cases occur for $n \leq 6$, and the decomposition problems for larger n seem to present no physical interest. So we consider some special cases in the next section, trying to cover all cases for small n.

III. SPECIAL CASES OF DECOMPOSITION

In this section, we consider some cases where further decomposition may be carried on; i.e., N is Abelian.

A. Orbits without Null Vectors

Since difficulties arise for the orbits of Sec. II.B which contain null vectors but not their positive and negative components, we shall consider orbits where such vectors do not occur, i.e., $s_1 = 0$. Then G'' is the group

$$S(\underline{U}^2(p_1, q_1) \times U(p_1', q_1') \times \underline{GL}(r_1)) \cdot \mathbf{C}(r_1(p_1 + q_1)),$$

where only the underlined subgroups act on

$$\mathbf{C}(r_1(p_1 + q_1))$$

(and where p_1' may be smaller than q_1'). The decomposition is already given by Sec. II.B and (r_1, p_1, q_1) is decomposed into $p_2, q_2, p_2', q_2', s_2, r_2$. In the same way, (r, p, q) gives rise to $p_1, q_1, p_1', q_1', s_1, r_1$. If we take only orbits for which $s_2 = 0$, for the same reasons as above, and repeat this procedure until we arrive at a reductive stabilizer, we obtain the series

$$s(U(p_1', q_1') \times U^2(p_2', q_2') \times \cdots \times U^k(p_k', q_k') \times U^{k+1}(p_k, q_k) \times GL(r_k)),$$

such that

$$n = r_k + \sum j \cdot (p_j' + q_j') + (k + 1) \cdot (p_k + q_k).$$

We distinguish two series inside this formula: series (1) if $r_k \neq 0$, and series (2) if $r_k = 0$.

A stabilizer belonging to (1) determines only one orbit, since this notation makes a distinction between $U(p_j', q_j')$ and $U(q_j', p_j')$ for $p_j' \neq q_j'$, and thus all successive ranks and signatures are given. On the other hand, to the stabilizers belonging to (2) there corresponds a continuous spectrum of orbits since the $\left(\begin{smallmatrix} p_k + q_k \\ p_{k-1}+q_{k-1}\end{smallmatrix}\right)$ pseudonorm varies throughout all strictly positive real values and does not affect the stabilizer. So we have:

Theorem 3: There are two series of UIR's of G. Representations of series (1) are in 1-to-1 correspondence with those of the little group (1). Representations of the second series are determined by those of the little group (2) and by a strictly positive number.

Remark: p_j' and q_j' may be zero, but $p_k + q_k \neq 0$ (respectively, $r_k \neq 0$) in case (2) [respectively (1)]. k may take any value from 0 to $n - 1$ (provided we put $p_0 = p$, $q_0 = q$, $r_0 = r$, $p_0' = q_0' = 0$).

B. Orbits with Null Vectors

Another case where the difficulties due to nilpotency drop out is when $p_1 = q_1' = q_1 = q' = 0$ (then $p = q = s \leq r$); there are different results for $s_1 = r_1$ and $s_1 < r_1$.

Case A: $s_1 = r_1$: The group G'' is then

$$\mathbf{Z}_3 \times SL(s_1; \mathbf{C}) \cdot \mathbf{R}(s_1^2),$$

since the determinant condition gives

$$\det \lambda_1 = \exp(\tfrac{2}{3} ik\pi).$$

We have immediately:

Theorem 4: If $n = 3n'$, to every representation of G_n' there correspond three representations of G_n.

Remark: The same type of little group may occur as a direct factor after the same steps of decomposition described in Sec. III.A are carried out, the other direct factor being reductive.

Case B: $s_1 < r_1$: Let $(\Delta, \Psi) = (\hat{H}, \hat{Y})$ be an element of the dual Abelian subgroup. It transforms by

$$\begin{pmatrix} \Delta \\ \Psi \end{pmatrix} \to \begin{pmatrix} \lambda_1^* \Delta \lambda_1 \\ \lambda_1^{-1} \Psi \mu_1 \end{pmatrix}.$$

Rank Δ, sign Δ, and rank Ψ are discrete invariants, but, since Δ and Ψ do not vary independently, orbits cannot yet be characterized. So we consider the orbit of Ψ when Δ is a fixed (p', q', r') Hermitian form. As a result, λ_1 must be of the form

$$\lambda_1 = P_0 \lambda_1 P_0 + P_0 \lambda_1 P_1 + P_1 \lambda_1 P_1,$$

where $P_1 \lambda_1 P_1$ belongs to a $U(p', q')$ subgroup [in particular, $|\det(P_1 \lambda_1 P_1)| = 1$] and P_0 and P_1 are Hermitian idempotents such that

$$P_0 + P_1 = I, \quad P_0 \Delta P_0 = 0, \quad P_1 \Delta P_1 = \Delta.$$

Since $\Psi = P_0 \Psi + P_1 \Psi$, let us consider the orbit of $P_1 \Psi$ first:

$$P_0 \Psi \to (P_0 \lambda_1^{-1} P_0) \cdot (P_0 \Psi) \cdot \mu_1 + (P_0 \lambda_1^{-1} P_1) \cdot (P_1 \Psi) \cdot \mu_1,$$
$$P_1 \Psi \to (P_1 \lambda_1^{-1} P_1) \cdot (P_1 \Psi) \cdot \mu_1.$$

According to previous results (see Sec. II.B), rank and signature of the vector space spanned by the column vectors of $P_1 \Psi$ are invariant. Calculating the corresponding little group by introducing projectors Q_0 and Q_1, such that $Q_0 + Q_1 = I$, $P_1 \Psi Q_1 = P_1 \Psi$, and $P_1 \Psi Q_0 = 0$, one obtains, in particular, $Q_1 \mu_1 Q_0 = 0$, and, thus,

$$P_0\Psi Q_0 \to (P_0\lambda_1^{-1}P_0)\cdot(P_0\Psi Q_0)\cdot(Q_0\mu_1 Q_0)$$

and rank $P_0\Psi Q_0$ is a constant, while $P_0\Psi Q_1$ varies arbitrarily.

Searching for continuous spectra of invariants, we observe that $|\det (P_0\lambda_1^{-1}P_0| = |\det (Q_0\mu_1 Q_0)| > 0$; so there is no continuous spectrum related to $(P_0\Psi Q_0)$. Examining $P_1\Psi$, we see that $|\det \mu_1|$ is arbitrary, unless $P_1 = 1$ (i.e., rank $\Delta = s_1$), in which case

$$|\det \mu_1| = \det |(P_1\lambda_1 P_1)| = 1.$$

Then there exists a pseudonorm for $P_1\Psi = \Psi_1$, if rank Ψ = rank $\mu_1 = r_1$. Finally, no continuous spectrum exists for Δ. Thus we have the following result.

Proposition: Orbits are characterized by sign Δ and
(a) if rank $\Delta = s_1$ and rank $\Psi = r_1$, by sign Ψ and the $\binom{r_1}{s_1}$-pseudonorm $\|\Psi\|$,
(b) if rank $\Delta = s_1$ and rank $\Psi < r_1$, by sign Ψ, and
(c) if rank $\Delta < s_1$, by sign $P_1\Psi$ and rank $P_0\Psi P_0$.

We shall not calculate explicitly the stabilizers, except in case $s_1 = 1$, sufficient for $n \leq 6$.

If $s_1 = 1$, the only cases for sign Δ are $+$, $-$, and 0. Sign $P_1\Psi$ is no longer relevant since all columns of Ψ are parallel to one of them and rank Ψ is either 0 or 1. In the last case, further decompositions are easily done until a zero rank for some ψ_i is obtained, in which case the stabilizer is reductive. We thus obtain three series of stabilizers.

Orbit		Stabilizer	
Sign Δ	Successive rank Ψ		
$+$	$\underbrace{1, 1, \cdots, 1}_{(k-1)\text{ times}}\,0$ $(1 \leq k \leq n-3)$	$S(U^{k+1;1}(1, 0) \times GL(n-k-2))$	(4)
0		$S(GL^{k+1;1}(1) \times GL(n-k-2))$	(5)
$-$		$S(U^{k+1;1}(0, 1) \times GL(n-k-2))$	(6)

If the successive decompositions give no zero rank for any Ψ_i, we obtain the following terms, which eventually depend on an additional positive parameter $\|\Psi\|$.

Orbit			Stabilizer		
Sign Δ	Rank Ψ	Norm Ψ			
$+$	$\underbrace{1, 1, \cdots, 1}_{n-3}$	$M \in R^+$	$S(U^{n-1;1}(1, 0))$	$\sim \mathbf{Z}^n$	(4′)
0		0	$S(GL^{n-1;1}(1))$		(5′)
$-$		$M \in R^+$	$S(U^{n-1;1}(0, 1))$		(6′)

Remark: Analogous series may occur after some steps of decomposition of type Sec. III.A.

IV. DECOMPOSITION IN NILPOTENT CASES

Here, we consider one case where there is "effective nilpotency." This is a starting point for more general nilpotent cases.

Let the first decomposition be $(p+1, 1, 1)$ or $(1, p+1, 1)$ and the second one $s_1 = 1$. We thus obtain G'' with the group law

$$(t, \zeta, w)(t', \zeta', w') = (t' + t - \tfrac{1}{2}i(\zeta^* w\zeta' e^{-i\vartheta} - \zeta'^* w\zeta e^{i\vartheta}), \zeta + w\zeta' e^{-i\vartheta}, ww'),$$

where

$$t, t' \in \mathbf{R}, \quad \zeta', \zeta \in \mathbf{C}^p, \quad w, w' \in U(p), \quad \det w = e^{-2i\vartheta}.$$

If we put $u = we^{-i\vartheta}$, we observe that

$$w = u \cdot (\det u)^{-1/n} = u \cdot e^{i\vartheta},$$

i.e., G'' is the n-fold covering of the group Γ_p defined by

$$(t', \zeta', u')(t, \zeta, u) = (t + t' - \tfrac{1}{2}i(\zeta'^* u'\zeta - \zeta^* u'^* \zeta'), \zeta' + u'\zeta, u'u).$$

Since G'' and Γ_p are locally isomorphic, we study the representations of Γ_p. Since G'' is the central extension $Z_n \to G'' \to \Gamma_p$, all the representations of G'' are ray representations of Γ_p. They are all obtained by allowing the character of the center $U(1)$ of the reductive subgroup $U(p)$ of Γ_p to be any integral multiple of $2\pi/n$. Thus, there correspond n representations of G'' to any representation of Γ_p.

After observing that unfaithful representations of Γ_p are either those of $U(p)$ or those of $U(p)\cdot \mathbf{C}^p$, which are known, we study the faithful representations that occur in the regular representation of Γ_p.

We proceed to the decomposition of the right regular representation. This immediately separates the representations which take different values on the center:

$$(t, \zeta, u)\colon F(t', \zeta', u') \to F(t + t' + \tfrac{1}{2}i(\zeta^* u'^* \zeta' - \zeta'^* u' \zeta), \zeta' + u'\zeta, u'u).$$

By the Fourier transform on $\mathbf{C}^p \otimes \mathbf{R} \approx \mathbf{R}(2p+1)$, we obtain

$$(t, \zeta, u)\colon \hat{F}(\lambda, z, u') \to \exp\,[i\lambda t + \tfrac{1}{2}i(\zeta^* z + z^*\zeta)] \times \hat{F}(\lambda, u^{-1}(z - i\lambda\zeta), u'u).$$

Different λ's give inequivalent representations. Therefore, we now consider $\lambda \neq 0$ to be a constant. We write $\hat{F}(\lambda, z, u') = \exp(-\tfrac{1}{2}z^*z\,|\lambda|^{-1})\cdot\Phi_\lambda(z, u')$; then we introduce a polynomial basis on the z's for Φ_λ, which spans a dense subspace of $L^2(\mathbf{C}^p)$.

Thus, we have

$$\Phi_\lambda(z, u') \to \exp\,[-\tfrac{1}{2}|\lambda|\,\zeta^*\zeta + \tfrac{1}{2}i(z^*\zeta(1+\epsilon) + \zeta^* z(1-\epsilon))]\cdot\Phi_\lambda(u^{-1}(z - i\lambda\zeta), u'u),$$

where $\epsilon = \lambda^{-1}\cdot|\lambda|$.

Now, to find a convenient polynomial basis, we search the expression of the infinitesimal generators of the nilpotent part. We take

$$\left.\begin{aligned} X_\mu &= \frac{\partial}{\partial\xi_\mu} = \frac{\partial}{\partial\zeta_\mu} + \frac{\partial}{\partial\bar\zeta_\mu},\\ Y_\mu &= \frac{\partial}{\partial\eta_\mu} = i\left(\frac{\partial}{\partial\zeta_\mu} - \frac{\partial}{\partial\bar\zeta_\mu}\right),\end{aligned}\right\}\ \text{or}\ \left\{\begin{aligned} 2Z_\mu &= X_\mu + iY_\mu,\\ 2\bar Z &= X_\mu - iY_\mu,\end{aligned}\right.$$

and obtain, for $\lambda > 0$,

$$Z_\mu\Phi_\lambda = i\lambda\frac{\partial\Phi_\lambda}{\partial\bar z_\mu},\quad Z_\mu\Phi_\lambda = -i\lambda\frac{\partial\Phi_\lambda}{\partial z_\mu} + i\bar z_\mu\Phi_\lambda,$$

and, for $\lambda < 0$,

$$Z_\mu\Phi_\lambda = i\lambda\frac{\partial\Phi_\lambda}{\partial\bar z_\mu} + iz_\mu\Phi_\lambda,\quad \bar Z_\mu\Phi_\lambda = -i\lambda\frac{\partial\Phi_\lambda}{\partial z_\mu}.$$

We now introduce the polynomial basis

$$P^{+|\lambda|}_{\mathbf{h},\mathbf{k}}(z, z^*) = (\mathbf{h}!\,\mathbf{k}!)^{-1}\left(\bar z - |\lambda|\frac{\partial}{\partial\bar z}\right)^{\mathbf{h}} z^{\mathbf{k}},$$

$$P^{-|\lambda|}_{\mathbf{h},\mathbf{k}}(z, z^*) = (\mathbf{h}!\,\mathbf{k}!)^{-1}\left(z - |\lambda|\frac{\partial}{\partial\bar z}\right)^{\mathbf{h}} \bar z^{\mathbf{k}},$$

where $\mathbf{h}$ and $\mathbf{k}$ are integral p-tuples. We recall the symbolic notation

$$|\mathbf{k}| = |(a, b, c, \cdots)| = a + b + c + \cdots,$$
$$\mathbf{k}! = (a, b, c, \cdots)! = a!\,b!\,c!\cdots,$$
$$z^{\mathbf{k}} = (z_1, z_2, z_3, \cdots)^{(a,b,c,\cdots)} = z_1^a\cdot z_2^b\cdot z_3^c\cdots,$$
$$\mathbf{k} \le \mathbf{k}' \Leftrightarrow a \le a' \text{ and } b \le b' \text{ and } c \le c' \text{ and } \cdots,$$
$$\mathbf{k} - \mathbf{k}' = (a - a', b - b', c - c', \cdots).$$

We can now write

$$\Phi_{\pm|\lambda|}(z, u) = \sum_{\mathbf{h},\mathbf{k}} \varphi_{\mathbf{h},\mathbf{k}}(u)\cdot P^{\pm|\lambda|}_{\mathbf{h},\mathbf{k}}(z, z^*),$$

where $\mathbf{h}$ and $\mathbf{k}$ take any value.

We calculate next the restriction of the regular representation to the subgroup $\mathbf{C}^p = \{(0, \zeta, 1)\}$. Let $\rho = |\lambda|$ and $P = P^{-\rho}$ (the case $P^{+\rho}$ can be automatically deduced by interverting z and $\bar z$). We can then write

$$P_{\mathbf{h},\mathbf{k}} = \sum_{\mathbf{m}} [\mathbf{m}!\,(\mathbf{k}-\mathbf{m})!\,(\mathbf{h}-\mathbf{m})!](-p)^{|\mathbf{m}|} z^{\mathbf{h}-\mathbf{m}}\bar z^{\mathbf{k}-\mathbf{m}},$$

where $0 \le \mathbf{m} \le \mathbf{h}$ and $0 \le \mathbf{m} \le \mathbf{k}$.

Carrying out the calculations, we obtain

$$P_{\mathbf{h},\mathbf{k}} \to \sum_{\mathbf{j},\mathbf{m}} [(\mathbf{j}-\mathbf{m})!\,\mathbf{m}!\,(\mathbf{h}-\mathbf{m})!]^{-1} \times \mathbf{j}!\,(i\bar\zeta)^{\mathbf{j}-\mathbf{m}}(i\rho\zeta)^{\mathbf{h}-\mathbf{m}}\cdot P_{\mathbf{j},\mathbf{k}},$$

with

$$0 \le \mathbf{j} \le \infty,\quad 0 \le \mathbf{m} \le \mathbf{h},\quad 0 \le \mathbf{m} \le \mathbf{j}.$$

That is,

$$P^{-|\lambda|}_{\mathbf{h},\mathbf{k}} \to \exp(\tfrac{1}{2}\lambda\zeta^*\zeta)\sum_{0\le\mathbf{j}\le\infty}\left(+i\bar\zeta + \frac{\partial}{\partial(-i\lambda\zeta)}\right)^{\mathbf{j}}\cdot(-i\lambda\zeta)^{\mathbf{h}}\cdot(\mathbf{h}!)^{-1}\cdot P^{-|\lambda|}_{\mathbf{j},\mathbf{k}},$$

$$P^{+|\lambda|}_{\mathbf{h},\mathbf{k}} \to \exp(-\tfrac{1}{2}\lambda\zeta^*\zeta)\sum_{0\le\mathbf{j}\le\infty}\left(-i\zeta + \frac{\partial}{\partial(+i\lambda\zeta)}\right)^{\mathbf{j}}\cdot(i\lambda\bar\zeta)^{\mathbf{h}}\cdot(\mathbf{h}!)^{-1}\cdot P^{+|\lambda|}_{\mathbf{j},\mathbf{k}}.$$

Thus, the infinitesimal generators of the nilpotent subgroup act only on the first p-tuple. Since this action is independent of the value of $\mathbf{k}$, we may investigate the irreducible subspaces under the whole group by fixing $\mathbf{h} = 0$.

We introduce the basis functions

$$\psi_{+|\lambda|}(z, u) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}}(u)P^{+|\lambda|}_{0,\mathbf{k}} = \sum_{\mathbf{k}} \varphi_{\mathbf{k}}(u)z^{\mathbf{k}}\cdot(\mathbf{k}!)^{-1},$$

$$\psi_{-|\lambda|}(z, u) = \sum_{\mathbf{k}} \varphi_{\mathbf{k}}(u)P^{-|\lambda|}_{0,\mathbf{k}} = \sum_{\mathbf{k}} \varphi_{\mathbf{k}}(u)\bar z^{\mathbf{k}}\cdot(\mathbf{k}!)^{-1}.$$

We consider $\lambda > 0$, i.e., holomorphic functions on $\mathbf{C}^p$, the case $\lambda < 0$ being obtained by changing z to $\bar z$ and u to $\bar u$.

Decomposing $\varphi_{\mathbf{k}}(u)$ according to the characters of $U(p)$, we have

$$\varphi_{\mathbf{k}}(u) = \sum_{N,h,j,m} D^N(\dot u)^j_h\cdot e^{im\omega/p}\cdot K^{\mathbf{k}}_{N,h,j,m},$$

where $e^{i\omega} = \det u$, $\dot u \in (SU(p)/\mathbf{Z}_p)$, $m \in \mathbf{Z}$, D^N being the UIR of $SU(p)/\mathbf{Z}_p$ with maximal weight N.

But the space generated by the $P^{+|\lambda|}$ (respectively, $P^{-|\lambda|}$), $|\mathbf{k}|$ being constant, is the space of the UIR of $SU(p)/\mathbf{Z}_p$, with maximal weight $(0, 0, \cdots, 0, |\mathbf{k}|)$ [respectively, $(|\mathbf{k}|, 0, \cdots, 0)$], the center $U(1)$ being represented by $e^{i\omega}\cdot I \to e^{-i|\mathbf{k}|\omega}$ (respectively, $e^{i|\mathbf{k}|\omega}$).

Thus, the problem of decomposing the regular representation into irreducible components is equivalent to a reduction of the tensor product of two irreducible representations of $SU(p)$. Since there is no restriction on the range of N, we have immediately the following result.

Theorem: The regular representation of Γ_p is a direct integral of inequivalent unitary representations, each one labeled by a nonzero real number (character of the center). Furthermore, each of them decomposes into a direct sum of irreducible representations. Equivalence classes of the latter are in 1-to-1 correspondence with UIR's of $U(p)$.

Remark: The theorem holds if we replace Γ_p by $\Gamma_{(p-q)\times q}$ and $U(p)$ by $U(p-q)\times U(q)$. $\Gamma_{(p-q)\times q}$ is defined by exactly the same group law as Γ_p, the reductive group being $U(p-q) \times U(q)$ instead of $U(p)$.

V. RELATION BETWEEN $\mathfrak{g}_n$, $\mathfrak{sp}(m, \mathbf{R})$, AND CREATION–ANNIHILATION OPERATORS

Here we study the connection of the inhomogeneous $\mathfrak{sl}(n, \mathbf{C})$ structure with creation–annihilation operators. In fact, the set of n pairs of these operators forms a nilpotent Lie algebra $\mathfrak{A}_n$ of dimension $2n + 1$ with the commutation relations

$$[A_i, A_j^*] = \delta_{ij} I, \quad [A_i, A_j] = [A_i^*, A_j^*] = 0,$$
$$[A_i, I] = [A_i^*, I] = 0.$$

If now we consider the universal enveloping algebra U_n of $\mathfrak{A}_n$ on the field $\mathbf{R}$ of real numbers and take its quotient by the center I, we find that the set of symmetrized elements of degree two forms itself a Lie algebra, which is isomorphic to $\mathfrak{sp}(n, \mathbf{R})$. One can immediately identify the correspondence between the generators of the two algebras by means of the commutation relations

$$[\tfrac{1}{2}(A_iA_j^* + A_j^*A_i), \tfrac{1}{2}(A_k^*A_h^* + A_h^*A_k^*)]$$
$$= \delta_{ik}\tfrac{1}{2}(A_h^*A_j^* + A_j^*A_h^*) + \delta_{ih}\tfrac{1}{2}(A_k^*A_j^* + A_j^*A_k^*),$$
$$[\tfrac{1}{2}(A_iA_j^* + A_j^*A_i), \tfrac{1}{2}(A_kA_h + A_hA_k)]$$
$$= -\delta_{jk}\tfrac{1}{2}(A_hA_j + A_jA_h) - \delta_{jh}\tfrac{1}{2}(A_iA_k + A_kA_i),$$
$$[\tfrac{1}{2}(A_iA_j^* + A_j^*A_i), \tfrac{1}{2}(A_kA_h^* + A_h^*A_k)]$$
$$= \delta_{ih}\tfrac{1}{2}(A_kA_j^* + A_j^*A_k) - \delta_{jk}\tfrac{1}{2}(A_iA_h^* + A_h^*A_i),$$
$$[\tfrac{1}{2}(A_iA_j + A_jA_i), \tfrac{1}{2}(A_k^*A_h^* + A_h^*A_k^*)]$$
$$= \delta_{ik}\tfrac{1}{2}(A_jA_h^* + A_h^*A_j) + \delta_{jk}\tfrac{1}{2}(A_iA_h^* + A_h^*A_i)$$
$$+ \delta_{ih}\tfrac{1}{2}(A_jA_k^* + A_k^*A_j) + \delta_{jh}\tfrac{1}{2}(A_iA_k^* + A_k^*A_i).$$

Remark: To simplify the notations, we do not show the division of every polynomial of degree 2 in A_i and A_j by I, which is, actually, the only correct way to write commutators in the quotient U_n/I: One cannot drop the division when dealing with representations. The correspondence to the matrix Lie algebra of $sp(n, \mathbf{R})$ is

$$2(E_j^i - E_{n+i}^{n+j}) \sim A_iA_j^* + A_j^*A_i,$$
$$2(E_{n+i}^j + E_{n+j}^i) \sim A_iA_j + A_jA_i,$$
$$2(E_i^{n+j} - E_j^{n+i}) \sim A_i^*A_j^* + A_j^*A_i^*.$$

On the other hand, since we can always express $\mathfrak{sl}(n, \mathbf{C}) \cdot \mathbf{R}(n^2)$ as a subalgebra of $\mathfrak{sl}(2n, \mathbf{R}) \cdot \mathbf{R}(4n^2)$, we can identify $\mathfrak{g}_n$ as a subalgebra of $\mathfrak{sp}(2n, \mathbf{R})$.[19] The generators of this subalgebra are

$$\left.\begin{aligned}2X_{ij} &= A_iA_j^* + A_j^*A_i + A_{n+i}A_{n+j}^* + A_{n+j}^*A_{n+i}\\ 2Y_{ij} &= A_iA_{n+j}^* + A_{n+j}^*A_i - A_{n+i}A_j^* - A_j^*A_{n+i}\end{aligned}\right\} \in \mathfrak{gl}(n, \mathbf{C}),$$

$$\left.\begin{aligned}2\sigma_{ij} &= A_iA_j + A_jA_i + A_{n+i}A_{n+j} + A_{n+j}A_{n+i}\\ 2i\alpha_{ij} &= A_iA_{n+j} + A_{n+j}A_i - A_{n+i}A_i - A_jA_{n+i}\end{aligned}\right\}$$

$\in \mathbf{R}(n^2)$, σ_{ij} symmetric, real,

$i\alpha_{ij}$ skew-symmetric pure imaginary.

To obtain $\mathfrak{sl}(n, \mathbf{C})$ instead of $\mathfrak{gl}(n, \mathbf{C})$, we must impose a trace condition to the generators, that is, X_{ii} and Y_{ii} do not belong to the subalgebra, but $X_{ii} - X_{jj}$ and $Y_{ii} - Y_{jj}$ do. On the other hand, if we want to consider the generalized Weyl Lie algebra $\mathfrak{W}_n$ instead of the generalized Poincaré $\mathfrak{g}_n$ (the former being the semidirect product of the latter by a real generator, the dilatation), all elements X_{ii} (but not Y_{ii}) appear in $\mathfrak{W}_n$. It is easy to derive representations of W_n from those of G_n, since the dilatation affects only the determinant of α and not its signature. So, if $\det \alpha \neq 0$, there will be one orbit for each signature; the little group may have any real number for determinant, so that the pseudonorm conservation of Secs. II and III drops. In fact, any representation of W_n consists of a direct integral over a class of inequivalent representations of G_n that have identical successive signatures and little groups.

The relation between the structure of $\mathfrak{A}_n$ and $\mathfrak{g}_n$ is now established; we want to examine its effects on representation theory.

Let $\mathcal{A}_n$ be the simply connected Lie group with Lie algebra $\mathfrak{A}_n$. According to the general theory of representations of nilpotent Lie groups,[20] all its representations are known. Applying the Mackey–Bruhat method, we see that they form two series. The first series is isomorphic to the set of all representations of the Abelian group $\mathbf{R}(2n)$, the character of the center being trivial. These representations are not faithful.

All faithful representations occur in the second series and are determined by the nontrivial characters of the center. Passing to the Lie algebra, we see that every representation is characterized by a nonzero, pure imaginary number $i\lambda$, which is the value of the Casimir I. Acting on $L^2(\mathbf{R}^n; dx)$, it may be put into one of two forms

$$U_\lambda(A_k) = \lambda \frac{\partial}{\partial x_k}, \qquad \bar{U}_\lambda(A_h) = \lambda \frac{\partial}{\partial x_h},$$
$$U_\lambda(A_k^*) = i x_k, \quad \text{or} \quad \bar{U}_\lambda(A_h^*) = -i x_h, \quad \text{with } \lambda > 0,$$
$$U_\lambda(I) = i\lambda, \qquad \bar{U}_\lambda(I) = -i\lambda.$$

Available results for $\mathfrak{sp}(n; \mathbf{R})$ are obtained by this series only. From U_λ, one obtains

$$2M_{kj} = A_j A_k^* + A_k^* A_j \to 2x_k \frac{\partial}{\partial x_j} + \delta_{jk},$$
$$2\Sigma_{kj} = A_j A_k + A_k A_j \to i\lambda \frac{\partial^2}{\partial x_j \partial x_k},$$
$$-2S_{kj} = A_j^* A_k^* + A_k^* A_j^* \to i\lambda^{-1} \cdot x_j x_k.$$

Putting $H_i = M_{ii}$ and $Y_i = \frac{1}{2}\Sigma_{ii}$, $X_i = \frac{1}{2}S_{ii}$ and calculating the value of the first invariant Casimir operator of $\mathfrak{sp}(n, \mathbf{R})$, we find

$$C = \sum_i H_i^2 + (X_i + Y_i)^2 - (X_i - Y_i)^2 + \sum_{j<k} M_{jk}M_{kj} + M_{kj}M_{jk} + S_{jk}\Sigma_{kj} + \Sigma_{kj}S_{jk} = -\tfrac{1}{4}(2n^2 + 1).$$

It is easy to verify that the generators of $\mathfrak{sp}(n, \mathbf{R})$ are expressed as skew-Hermitian operators. On the other hand, representations corresponding to different λ's are either unitarily equivalent or contragredient, as we see by transforming the expressions above by the operator $f(x) \to \lambda^{-\frac{1}{4}} f(\lambda^{-\frac{1}{2}} x)$.

Two contragredient representations are not equivalent. Finally, as the example for $n = 1$ shows, these representations are not irreducible, but split into two irreducible inequivalent components acting respectively on the Hilbert spaces $\mathcal{H}^+$ and $\mathcal{H}^-$. $\mathcal{H}^+$ (respectively, $\mathcal{H}^-$) is the closure of the space spanned by polynomials of even (respectively, odd) degree in n real variables, every one multiplied by the function $\exp(-\frac{1}{2}\Sigma x_i^2)$. Thus, we have the following statement.

Proposition 2: A necessary condition for a skew-Hermitian (Schur) irreducible representation of $\mathfrak{sp}(n, \mathbf{R})$ to appear as a direct summand in a "squared" UIR of $\mathcal{A}_n$ is that the Beltrami–Laplace operator C has the value $-\frac{1}{2}(n^2 + \frac{1}{2})$. In this case, the contragredient representation appears also as a direct summand.

An illustration of this result, which establishes that correspondence of structures need not be reflected by a correspondence in representations, is the case $n = 1$.

For $U_\lambda(\mathfrak{A}_1)$, we have

$$Af(x) = \lambda f'(x),$$
$$A^*f(x) = ixf(x),$$
$$If(x) = i\lambda f(x),$$

which gives, for $\mathfrak{sp}(1, \mathbf{R}) \sim \mathfrak{sl}(2, \mathbf{R})$,

$$Hf(x) = xf'(x) + \tfrac{1}{2}f(x),$$
$$Yf(x) = -\tfrac{1}{2}if''(x),$$
$$Xf(x) = -\tfrac{1}{2}ix^2 f(x),$$

and

$$C = H^2 + 2(XY + YX) = -\tfrac{3}{4}.$$

Let us call this representation dD and its contragredient $d\bar{D}$. Examining it rapidly, we see that:

(a) The vectors $H^k \cdot \exp(-\frac{1}{2}x^2)$, $X^k \cdot \exp(-\frac{1}{2}x^2)$, and $Y^k \cdot \exp(-\frac{1}{2}x^2)$ (respectively, $H^k \cdot [x \cdot \exp(-\frac{1}{2}x^2)]$, $X^k \cdot [x \cdot \exp(-\frac{1}{2}x^2)]$, and $Y^k \cdot [x \cdot \exp(-\frac{1}{2}x^2)]$) span the same dense subspace of the space $\mathcal{H}^+$ (respectively, $\mathcal{H}^-$), consisting of all even (respectively, odd) square-integrable functions of $\mathbf{R}$ into $\mathbf{C}$.

(b) The compact generator $Y - X = \frac{1}{2}i(x^2 - d^2/dx)$ is the energy operator for the 1-dimensional harmonic oscillator. Its spectrum is $(n + \frac{1}{2})$, n being odd for odd eigenfunctions and even for even eigenfunctions. Its period is 4π, while the unity element of the group $\mathfrak{sl}(2, \mathbf{R})$ is equal to $\exp[2\pi(Y - X)]$.

From (a) and (b), we can conclude that

Proposition 3: D is integrable to the direct sum of two inequivalent UIR's of the group $[\mathbf{Z}_2 \cdot SL(2, \mathbf{R})]$,[21] the twofold covering of $SL(2, \mathbf{R})$, having center $\mathbf{Z}_4$. The expression of the representation D in $\mathcal{H}^+ \oplus \mathcal{H}^-$ is

$$\exp tH: f(x) \to e^{\frac{1}{2}t} f(e^t x),$$
$$\exp tX: f(x) \to \exp(-\tfrac{1}{2}itx^2) f(x),$$
$$\exp tY: f(x) \to \mathcal{F}^{-1}[\exp(-\tfrac{1}{2}itx^2)\mathcal{F}f](x).$$

We remark that the integrability and the connection with the harmonic oscillator can be easily generalized for n dimensions.

Coming back to the general case, we notice that representations of $\mathfrak{g}_n$ provide us with the spectra of many generators of $\mathfrak{sp}(2n, \mathbf{R})$. But, since we cannot, in general, take the "square root" of a representation of $\mathfrak{sp}(2n, \mathbf{R})$, we are not able to determine the spectra of the skew-adjoint operators A_i and A_i^*. In particular, the operator $\sum A_i A_i^* + A_i^* A_i$ is contained in $\mathfrak{W}_n$ and its spectrum can be easily calculated for a given representation. In fact, in most interesting cases, its eigenvectors are homogeneous polynomials on the variables (since its form is $\sum x_i \cdot \partial/\partial x_i$), which could be expected, since we identify it with the generalized dilatation.

CONCLUSION

The main results can be summarized in the following way:

(1) A "geometrical" study of involutive automorphisms and antiautomorphisms of the groups G_n is done, thus allowing the formal definition of *TCP* operators on groups containing the Poincaré group. One must remark that the method used is not specific to the family G_n.

(2) Many series of representations of G_n are found, among them all those which enter the Plancherel formula. However, this formula is not explicitly given, because the Plancherel measure is not yet known for $SU(p, q)$ in general.

(3) The groups $\Gamma_{p\times q} = R \cdot N$ (with R compact-reductive, N nilpotent) are examined, and their regular representation is decomposed into irreducible parts. It seems that the method applied here, which makes use of the properties of a "good" polynomial basis on the infinitesimal generators of N, may also be successfully applied in cases of groups $\Gamma_{p,q}$ (R noncompact) and even more general cases of such semidirect products.

(4) Structural and spectral relations are established between creation–annihilation operators and Lie algebras of the family G_n. Such a treatment has the physical meaning of connecting the group-theoretical and the quantum field-theoretical points of view.

(5) A thorough physical application of these results has still to be done, in order to obtain useful information for group-theoretical problems in domains such as strong interactions or nuclear spectroscopy. We shall deal with these applications in forthcoming papers.

ACKNOWLEDGMENTS

I should like to thank Professor M. Flato for suggesting me the subject as well as for helpful discussions and criticism. I am also indebted to Dr. D. Sternheimer for helpful discussions and remarks.

APPENDIX

We give the list of inequivalent classes of representations for small n in terms of orbits and little groups. This list is complete. Representations are characterized by an orbit decomposition; in this way, two representations which are induced by inequivalent representations of the same little group and which belong to the same orbit are not separated by this classification. [For example, for $n = 2$, the "physical" representations of $\mathcal{P}$—that is, with strictly positive mass—which have mass M are classified in the same family regardless of their spin. The spin would appear only if one gave the list of all representations of the little group $SU(2)$].

Thus, the description ends at a stabilizer, the representations of which are supposed known. Apart from pseudounitary and linear groups and their direct products, groups like $\Gamma_{(p,q)}$ appear. $\Gamma_{(p,q)}$ is defined in the same manner as $\Gamma_p = \Gamma_{(p,0)} = \Gamma_{(0,p)}$, but its faithful representations are not yet calculated (except for $\Gamma_{(p,0)}$ described in Sec. IV).

Orbits may be characterized by a continuous number, $M > 0$. The significance of M is not specified in each case: It may denote the modulus of determinant or a pseudonorm, as it is described in Secs. II and III.

Orbits are always characterized by discrete quantities, like successive signatures and ranks of matrices. These are sometimes described by plus and minus signs or by integers denoting rank, but they are also described by the little group expressed adequately. In the next column, a simpler expression of the stabilizer is given [the notation $GL(k)$ is used for $GL(k, C)$]. The notation $A \wedge B$, where A and B are Abelian groups, denotes a central extension of A by B.

TABLE I. $n = 2$.

Series	Continuous character	Discrete quantity	Little group	Little group
(2)	M	$++$ $+-$ $--$	$SU(2, 0)$ $SU(1, 1)$ $SU(0, 2)$	$\sim SU(2)$[a] $\sim SU(1, 1)$[b] $\sim SU(2)$[c]
(2)	M	$+\quad +$ $-\quad -$	$S(U(1, 0)^2)$ $S(U(0, 1)^2)$	$\sim \mathbf{Z}_2^{\mathrm{d}}$
(1)		$+\quad 0$ $-\quad 0$	$S(U(1, 0) \times GL(1))$ $S(U(0, 1) \times GL(1))$	$\sim U(1)$[e]
(1)		0	$SL(2)$	$\sim SL(2)$

[a] Positive mass $\sqrt{M}$.
[b] Imaginary mass $\sqrt{-M}$.
[c] Negative mass $-\sqrt{M}$.
[d] Zero mass, continuous spin M.
[e] Zero mass, discrete spin.

TABLE II. $n = 3$.

Series	Continuous character	Discrete quantities	Discrete quantities (l.g.)	Little group
		$+++$	$SU(3,0)$	$SU(3)$
		$++-$	$SU(2,1)$	$SU(2,1)$
		$+--$	$SU(1,2)$	
		$---$	$SU(0,3)$	
		$++ \quad +$	$S(U(1,0) \times U^2(1,0))$	
(2)	M	$+- \quad -$	$S(U(1,0) \times U^2(0,1))$	$U(1)$
		$-+ \quad +$	$S(U(0,1) \times U^2(1,0))$	
		$-- \quad -$	$S(U(0,1) \times U^2(0,1))$	
		$+ \quad + \quad +$	$S(U^3(1,0))$	$\mathbf{Z}_3$
		$- \quad - \quad -$	$S(U^3(0,1))$	
		0	$SL(3)$	$SL(3)$
		$+ \quad 0$	$S(U(1,0) \times GL(2))$	$U(1)\cdot SL(2)$
		$- \quad 0$	$S(U(0,1) \times GL(2))$	
(1)		$++ \quad 0$	$S(U(2,0) \times GL(1))$	$U(2)$
		$+- \quad 0$	$S(U(1,1) \times GL(1))$	$U(1,1)$
		$-- \quad 0$	$S(U(2,0) \times GL(1))$	
		$+ \quad + \quad 0$	$S(U^2(1,0) \times GL(1))$	$U(1)$
		$- \quad - \quad 0$	$S(U^2(0,1) \times GL(1))$	
(3)		$+- \quad 1^0$	$S(GL^{2;1}(1))\cdot\mathbf{R}$	$\mathbf{Z}_3 \times \mathbf{R}$

TABLE III. $n = 4$.

Series	Continuous character	Discrete quantities	Discrete quantities (little group)	Little group	
(2)	M	$(p,q,0)$	$SU(p,q)$	$SU(p,q)$	$p+q=4$
(2)	M	$(p,q,1)\ 1^+$	$S(U(p-1,q) \times U^2(1,0))$	$SU(p-1,q) \times U(1)$	$p+q=3$
(2)	M	$(p,q,1)\ 1^-$	$S(U(p,q-1) \times U^2(0,1))$	$SU(q,p-1) \times U(1)$	
(1)		$(p,q,1)\ 0$	$S(U(p,q) \times GL(1))$	$U(p,q)$	
(2)	M	$(p,q,2)\ 2^{\pm\pm}$	$S(U^2(p,q))$	$\mathbf{Z}_2 \times SU(p,q)$	$p+q=2$
(2)	M	$(p,q,2)\ 1^+1^+$	$S(U(p-1,q) \times U^3(1,0))$	$U(1)$	
(2)	M	$(p,q,2)\ 1^-1^-$	$S(U(p,q-1) \times U^3(0,1))$	$U(1)$	
(1)		$(p,q,2)\ 1^+0$	$S(U(p-1,q) \times U^2(1,0) \times GL(1))$	$U(1) \times U(1)$	
(1)		$(p,q,2)\ 1^-0$	$S(U(p,q-1) \times U^2(0,1) \times GL(1))$		
(1)		$(p,q,2)\ 0$	$S(U(p,q) \times GL(2))$	$U(1)\cdot(SU(p,q) \times SL(2))$	
(2)	M	$(p,q,3)\ 1^{\pm}1^{\pm}1^{\pm}$	$S(U^4(p,q))$	$\mathbf{Z}_4$	$p+q=1$
(1)		$(p,q,3)\ 1^{\pm}1^{\pm}0$	$S(U^3(p,q) \times GL(1))$	$U(1)$	
(1)		$(p,q,3)\ 1^{\pm}0$	$S(U^2(p,q) \times GL(2))$	$U(1) \times SL(2)$	
(1)		$(p,q,3)\ 0$	$S(U(p,q) \times GL(3))$	$U(1)\cdot SL(3)$	
(1)		$(0,0,4)$	$SL(4)$	$SL(4)$	
—		$(p+1,q+1,1)\ 1^0$	$S(GL^{2;1}(1) \times U(p,q))\cdot(\mathbf{C} \wedge \mathbf{R})$	$\mathbf{Z}_4\cdot\Gamma_1$	$p+q=1$
		$(1,1,2)\ 1^0$	$S(GL^{2;1}(1) \times GL(1))\cdot(\mathbf{C} \times \mathbf{R})$ gives:		
(4)		$(1,1,2)\ 1^0\ +\ 0$	$S(U^{2;1}(1,0) \times GL(1))$	$U(1)$	
(5)		$(1,1,2)\ 1^0\ 0\ 0$	$S(GL^{2;1}(1) \times GL(1))$	$GL(1)$	
(6)		$(1,1,2)\ 1^0\ -\ 0$	$S(U^{2;1}(0,1) \times GL(1))$	$U(1)$	
(4′)	M	$(1,1,2)\ 1^0\ +\ 1$	$S(U^{3;1}(1,0))$	$\mathbf{Z}_4$	
(5′)		$(1,1,2)\ 1^0\ 0\ 1$	$S(GL^{3;1}(1))$		
(6′)	M	$(1,1,2)\ 1^0\ -\ 1$	$S(U^{3;1}(0,1))$		

TABLE IV. $n = 6$.

Series	Continuous character	Discrete quantities (little group)	Little group	
(2)	M	$SU(p, q)$	$SU(p, q)$	$p + q = 6$
(2)	M	$S(U(p-1, q) \times U^2(1, 0))$	$\mathbf{Z}_2 \cdot U(p-1, q)$	$p + q = 5$
(2)	M	$S(U(p, q-1) \times U^2(0, 1))$	$\mathbf{Z}_2 \cdot U(p, q-1)$	
(1)		$S(U(p, q) \times GL(1))$	$U(p, q)$	$p + q = 5$
(2)	M	$S(U(p', q') \times U^2(p_1, q_1))$	$SU(p', q') \times U(p_1, q_1)$	$p_1 + q_1 = 2$, $p' + q' = 2$
(2)	M	$S(U(p', q') \times U^3(p_1, q_1))$	$SU(p', q') \times U(1)$	$p' + q' = 3$, $p_1 + q_1 = 1$
(1)		$S(U(p', q') \times U^2(p_1, q_1) \times GL(1))$	$U(p', q') \times U(1)$	
(1)		$S(U(p, q) \times GL(2))$	$U(1) \cdot (SU(p, q) \times SL(2))$	$p + q = 4$
(2)	M	$S(U^2(p, q))$	$\mathbf{Z}_2 \times U(p, q)$	$p + q = 3$
(2)	M	$S(U(p', q') \times U^2(p'', q'') \times U^3(p_2, q_2))$	$U(1) \times U(1)$	$p_2 + q_2 = p'' + q'' = 1$, $p' + q' = 1$, $p_1 + q_1 = 2$
(1)		$S(U(p', q') \times U^2(p_1, q_1) \times GL(1))$	$U(1) \times U(p_1, q_1)$	
(2)	M	$S(U(p', q') \times U^4(p_1, q_1))$	$\mathbf{Z}_2 \cdot U(p', q')$	$p' + q' = 2$, $p_1 + q_1 = 1$
(1)		$S(U(p', q') \times U^3(p_1, q_1) \times GL(1))$	$U(p', q') \times U(1)$	
(1)		$S(U(p', q') \times U^2(p_1, q_1) \times GL(2))$	$\mathbf{Z}_2 \cdot [U(p', q') \times U(1) \cdot SL(2)]$	
(1)		$S(U(p, q) \times GL(3))$	$U(1) \cdot [SU(p, q) \times SL(3)]$	$p + q = 3$
(2)	M	$S(U^3(p, q))$	$\mathbf{Z}_3 \times U(p, q)$	$p + q = 2$
(2)	M	$S(U^2(p', q') \times U^4(p_1, q_1))$	$U(1) \times \mathbf{Z}_2$	$p' + q' = 1$, $p_1 + q_1 = 1$
(1)		$S(U^2(p', q') \times U^3(p_1, q_1) \times GL(1))$	$U(1) \times U(1)$	
(1)		$S(U^2(p, q) \times GL(2))$	$U(p, q) \times SL(2)$	$p + q = 2$
(2)	M	$S(U(p', q') \times U^5(p_1, q_1))$	$U(1)$	$p_1 + q_1 = p' + q' = 1$
(1)		$S(U(p', q') \times U^4(p_1 + q_1) \times GL(1))$	$U(1) \times U(1)$	
(1)		$S(U(p', q') \times U^3(p_1 + q_1) \times GL(2))$	$U(1) \times [U(1) \cdot SL(2)]$	
(1)		$S(U(p', q') \times U^2(p_1 + q_1) \times GL(3))$	$U(1) \times [U(1) \cdot SL(3)]$	
(1)		$S(U(p, q) \times GL(4))$	$U(1) \cdot [SU(p, q) \times SL(4)]$	$p + q = 2$
(2)	M	$S(U^6(p, q))$	$\mathbf{Z}_6$	$p + q = 1$
(1)		$S(U^5(p, q) \times GL(1))$	$U(1)$	
(1)		$S(U^4(p, q) \times GL(2))$	$U(1) \times SL(2)$	
(1)		$S(U^3(p, q) \times GL(3))$	$U(1) \times SL(3)$	
(1)		$S(U^2(p, q) \times GL(4))$	$\mathbf{Z}_2 \cdot [U(1) \cdot SL(4)]$	
(1)		$S(U(p, q) \times GL(5))$	$U(1) \cdot SL(5)$	
(1)		$SL(6)$	$SL(6)$	

Table IV(bis). $n = 6$.

Series	Continuous character	Discrete quantities		Discrete quantities (little group)		Little group	
—		$(p+1, q+1, 1)$		1^0	$S(GL^{2;1}(1) \times U(p,q)) \cdot (\mathbf{C}^3 \wedge \mathbf{R})$	$\mathbf{Z}_6 \cdot \Gamma_{(p,q)}$	$p+q=3$
—		$(p_1+p'+1, q_1+q'+1, 2)$		$2^{\pm 0}$	$S(GL^{2;1}(1) \times U(p',q') \times U^2(p_1,q_1)) \cdot (\mathbf{C}^2 \wedge \mathbf{R})$	$\mathbf{Z}_6 \cdot \Gamma_{1\times 1}$	$\begin{cases} p'+q'=1 \\ p_1+q_1=1 \end{cases}$
(3)		$(2, 2, 2)$		2^{00}	$S(GL^{2;1}(2)) \cdot \mathbf{R}^4$	$\mathbf{Z}_3 \times G_2$	
		$(p+1, q+1, 2)$		1^0	$S(GL^{2;1}(1) \times U(p,q) \times GL(1)) \cdot (\mathbf{C}^2 \wedge (\mathbf{C} \times \mathbf{R}))$ gives:		
—		rank 1:			$S(GL^{3;1}(1) \times U(p,q)) \cdot (\mathbf{C}^2 \wedge \mathbf{R})$	$\mathbf{Z}_6 \cdot \Gamma_{(p,q)}$	$p+q=2$
—		rank 0:			$S(GL^{2;1}(1) \times U(p,q) \times GL(1)) \cdot (\mathbf{C}^2 \wedge \mathbf{R})$	$U(1) \times [\mathbf{R}^+ \cdot \Gamma_{(p,q)}]$	
		$(p+1, q+1, 3)$		$2^{\pm 0}$	$S(GL^{2;1}(1) \times U^2(p,q) \times GL(1)) \cdot ((\mathbf{C} \times \mathbf{C}) \wedge (\mathbf{C} \times \mathbf{R}))$ gives		
—		rank 1:			$S(GL^{2;1}(1) \times U^3(p,q)) \times \mathbf{R}$	$U(1) \times \mathbf{Z}_3 \times \mathbf{R}$	$p+q=1$
		rank 0			$S(GL^{2;1}(1) \times U^2(p,q) \times GL(1)) \cdot (\mathbf{C} \wedge (\mathbf{C} \times \mathbf{R}))$ gives		
			rank 1		$S(GL^{3;1}(1) \times U^2(p,q)) \cdot (\mathbf{C} \wedge \mathbf{R})$	$\mathbf{Z}_2 \times \Gamma_1$	$p+q=1$
			rank 0		$S(GL^{2;1}(1) \times U^2(p,q) \times GL(1)) \cdot (\mathbf{C} \wedge \mathbf{R})$	$U(1) \cdot (\mathbf{R}^+ \cdot \Gamma_1)$	
(3)		$(p+1, q+1, 3)$		$2^{+-}1^0$	$S(GL^{3;2}(1) \times U(p,q)) \times \mathbf{R}$	$U(1) \times \mathbf{R}$	$p+q=1$
		$(p+1, q+1, 3)$		1^0	$S(GL^{2;1}(1) \times U(p,q) \times GL(2)) \cdot (\mathbf{C} \wedge (\mathbf{C} \times \mathbf{R}))$ gives:		
—		ranks		1, 1	$S(GL^{4;1}(1) \times U(p,q) \cdot (\mathbf{C} \wedge \mathbf{R})$	$\mathbf{Z}_6 \cdot \Gamma_1$	$p+q=1$
—				1, 0	$S(GL^{3;1}(1) \times U(p,q) \times GL(1)) \cdot (\mathbf{C} \wedge \mathbf{R})$	$U(1) \times (\mathbf{R}^+ \cdot \Gamma_1)$	
—				0	$S(GL^{2;1}(1) \times U(p,q) \times GL(2)) \cdot (\mathbf{C} \wedge \mathbf{R})$	$[U(1) \cdot SL(2)] \times [(\mathbf{Z}_4 \times \mathbf{R}^+) \cdot \Gamma_1]$	
(4)		$(1, 1, 4)$	2^{+-}	$1^0 + 0$	$S(U^{3;2}(1,0) \times GL(1))$	$U(1)$	
(5)		$(1, 1, 4)$	2^{+-}	$1^0\ 0\ 0$	$S(GL^{3;2}(1) \times GL(1))$	$GL(1)$	
(6)		$(1, 1, 4)$	2^{+-}	$1^0 - 0$	$S(U^{3;2}(0,1) \times GL(1))$	$U(1)$	
(4′)	M	$(1, 1, 4)$	2^{+-}	$1^0 + 1$	$S(U^{4;2}(1,0))$	$\mathbf{Z}_6$	
(5′)		$(1, 1, 4)$	2^{+-}	$1^0\ 0\ 1$	$S(GL^{4;2}(1))$		
(6′)		$(1, 1, 4)$	2^{+-}	$1^0 - 1$	$S(U^{4;2}(0,1))$		
(4)		$(1, 1, 4)$	1^0	+	$S(U^{k+1;1}(1,0) \times GL(4-k))$		$1 \leq k \leq 3$
(5)				0	$S(GL^{k+1;1}(1) \times GL(4-k))$		
(6)				−	$S(U^{k+1;1}(0,1) \times GL(4-k))$		
(4′)	M			+	$S(U^{5;1}(1,0))$	$\mathbf{Z}_6$	
(5′)				0	$S(GL^{5;1}(1))$		
(6′)	M			−	$S(U^{5;1}(0,1))$		

[1] B. Sakita, Phys. Rev. **136**, B1756 (1964).
[2] T. Fulton and J. Wess, Phys. Letters **14**, 57, 344 (1965).
[3] T. Fulton and J. Wess, Phys. Letters **15**, 177 (1965).
[4] W. Rûhl, H. Ruegg, and T. S. Santhanam, CERN Preprint 66/1106/5, 1966.
[5] M. Flato, *Symétries de type Lorentzien et interactions fortes* (Gauthier-Villars, Paris, 1967).
[6] M. Flato and D. Sternheimer, J. Math. Phys. **7**, 1932 (1966).
[7] V. G. Kadyshevskii and I. T. Todorov, Yad. Fiz. **3**, 135 (1966) [Sov. J. Nucl. Phys. **3**, 96 (1966)].
[8] E. Wigner, Ann. Math. **40**, 149 (1939).
[9] G. Mackey, Ann. Math. **55**, 101 (1952).
[10] G. Mackey, Ann. Math. **58**, 193 (1953).
[11] F. Bruhat, Bull. Soc. Math. France **84**, 97 (1956).
[12] The notation $\mathbf{R}(p)$ is employed indifferently for the Abelian Lie group or Lie algebra isomorphic to $\mathbf{R}^p$. Sometimes $\mathbf{C}(p)$ figures instead of $\mathbf{R}(2p)$, when we want to emphasize a complex structure on the algebra or group.
[13] E. Angelopoulos, Compt. Rend. Acad. Sci. Paris **263**, 400 (1966).
[14] For representations of pseudo-unitary groups see, e.g., M. I. Graev, Trudy Mosk. Mat. Ob. 7, 335 (1958). For linear groups C. Fronsdal (ICTP Preprint 66/51, Trieste, 1966) gives quite complete results, although the approach is not entirely rigorous; see also E. M. Stein, Ann. Math. **86**, 461 (1967).
[15] L. Michel, Ecole Polytechnique, Preprint A81, 366(39) Paris, 1965.
[16] The hypothesis of continuity of automorphisms is needed for the proof, as in Ref. 8; even the pure algebraic method of Ref. 15 makes implicitly use of such a hypothesis when passing from rational to real numbers.
[17] These results contain the case $n = 2$, since $SU(2) \sim SU^*(2)$ and $SU(1, 1) \sim SL(2, \mathbf{R})$, while for larger n no parity can be defined by $SU(p, q)$; thus $SU^*(2n)$ appears as a good generalization.
[18] That is, as defined in the Dynkin diagrams for Lie algebras, see E. B. Dynkin, Transl. Am. Math. Soc., Ser. 2, **6**, No. 3 (1960).
[19] C. D. Palev (ICTP Preprint 68/23, Trieste, 1968) gives other possibilities of inclusions, taking only bilinear combinations of A_i^*, A_j.
[20] L. Pukanszky, *Leçons sur les représentations des groupes* (Dunod, Paris, 1965); see also for results on Lie algebra, J. Dixmier, Représentations irréductibles des alg. de Lie nilpotentes, Anais Acad. Brasil Ciencias.
[21] This group has a quotient by $\mathbf{Z}_2$ isomorphic to $SL(2, \mathbf{R})$, but does not contain $SL(2, \mathbf{R})$ as a subgroup; that is why the representations D_+, D_-, $\bar{D}_+$, and $\bar{D}_-$ do not appear in the series of $SL(2, \mathbf{R})$ calculated by Bargmann.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Thermodynamic Limit of Time-Dependent Correlation Functions for One-Dimensional Systems

G. GALLAVOTTI
The Rockefeller University, New York, New York 10021

O. E. LANFORD III*
Department of Mathematics, University of California, Berkeley, California 94720

AND

JOEL L. LEBOWITZ†
Belfer Graduate School of Science, Yeshiva University, New York, New York 10033

(Received 9 March 1970)

We investigate the time evolution of the correlation functions of a nonequilibrium system when the size of the system becomes very large. At the initial time $t = 0$, the system is represented by an equilibrium grand canonical ensemble with a Hamiltonian consisting of a kinetic energy part, a pairwise interaction potential energy between the particles, and an external potential. At time $t = 0$ the external field is turned off and the system is permitted to evolve under its internal Hamiltonian alone. Using the "time-evolution theorem" for a 1-dimensional system with bounded finite-range pair forces, we prove the existence of infinite-volume time-dependent correlation functions for such systems, $\lim \rho_\Lambda(t; q_1, p_1; \cdots; q_n, p_n)$, as $\Lambda \to \infty$, where Λ is the size of the finite system. We also show that these infinite-volume correlation functions satisfy the infinite BBGKY hierarchy in the sense of distributions.

1. INTRODUCTION

The rigorous mathematical study of equilibrium statistical mechanics during the last decade has achieved many successes. This study concerns itself primarily with the properties of equilibrium systems in the thermodynamic limit, i.e., as the size of the system becomes infinite at fixed temperature and activity (or density). In particular, the existence and analyticity of the correlation functions at small values of the activity z has been proven for a wide class of interacting systems.[1] The existence and convexity of the free energy has been proven for an even larger class of systems at all values of the activity.[2]

The comparable mathematical investigation of the infinite-volume limit of nonequilibrium systems is much more difficult and has begun only recently. The results are restricted to 1-dimensional systems of particles interacting by smooth, finite-range pair forces, and they prove the existence for all times of a "regular" solution of Newton's equations of motion for a "regular" initial configuration. A regular configuration is, roughly speaking, one in which the number of particles in a unit interval and the magnitude of the momentum of any particle in that interval have a bound of the form $\delta \log R$, where R denotes the distance of the interval from the origin. It is shown in Ref. 3 that, at equilibrium, if either the activity is small or the interparticle potential is positive, the set of nonregular configuration has probability zero.

A question left open by these results is whether a state which at time $t = 0$ is described by a set of correlation functions can still be described by a set of correlation functions when $t \neq 0$.

In this paper we investigate this question and prove that, for certain classes of initial states, the time-evolving state is described by correlation functions and that these correlation functions satisfy the BBGKY hierarchy in the sense of distributions [see (2.9)].

The initial states we consider can be described as follows: Suppose that the system is in equilibrium at temperature β^{-1} and activity z under the influence of a pair potential and an external potential h which is localized in a finite region I_h. At time $t = 0$, we switch off the external field and the system begins to evolve. We prove that, if the activity is sufficiently small (i.e., if we are deep inside the gaseous phase), the system can always be described by a set of correlation functions which vary in time according to the BBGKY hierarchy. We are unable to prove even that the time-averaged correlation functions evolve toward the correlation functions which correspond to the equilibrium state at temperature β^{-1} and activity z (in absence of external field), as would be expected. We are, however, able to prove that the time-averaged correlation functions converge to a limit satisfying the stationary BBGKY hierarchy.

We note that initial states of the kind just described suffice, in principle, for the study of transport properties at low activity.

2. DESCRIPTION OF INITIAL STATE AND SUMMARY OF RESULTS

We consider a 1-dimensional system of identical particles of unit mass, interacting through a stable pair-potential $\Phi(q)$ which has finite range and is twice continuously differentiable. We denote

$$C = -\inf_{n, q_1, \cdots, q_n} \left(n^{-1} \sum_{i<j} \Phi(q_i - q_j) \right). \quad (2.1)$$

The condition of stability says precisely that $C < \infty$; it guarantees that the thermodynamic functions are well defined.[1]

The initial states we consider will be equilibrium states, at inverse temperature β and chemical potential μ, for an interaction coming from the pair potential Φ and an external potential $h(q)$ which is continuous and which vanish outside some bounded interval I_h. We will assume that the activity z [$= e^{\mu\beta}(2\pi/\beta)^{\frac{1}{2}}$, in units where Planck's constant is unity] is small enough so that the Mayer series converges, i.e.,

$$z < B(\beta)^{-1} \exp(-\beta C - 1), \quad (2.2)$$

where

$$B(\beta) = \int dq \, |\exp[-\beta\Phi(q)] - 1|. \quad (2.3)$$

Under these conditions, the thermodynamic limit for the correlation functions is known to exist for $h = 0$.[1,4]

Now let Λ denote a finite interval centered at the origin and containing I_h, and let

$$H_\Lambda(q_1, p_1; \cdots; q_n, p_n) = \tfrac{1}{2} \sum_{i=1}^{n} p_i^2 + \sum_{i<j} \Phi_\Lambda(q_i - q_j) \quad (2.4)$$

denote the Hamiltonian for a system of n particles, in the box Λ with periodic boundary conditions, interacting by the periodized 2-body interaction Φ_Λ. (By Φ_Λ we mean the potential obtained by periodizing Φ with respect to Λ. In order that Φ_Λ be unambiguous, we will assume that the length of Λ is at least twice the range of Φ.[3]) We let $\tilde{T}_\Lambda^t$ denote the time-evolution mapping in the periodic box Λ determined by the Hamiltonian H_Λ. Finally, we let

$$h(q_1, \cdots, q_n) = \sum_{i=1}^{n} h(q_i). \quad (2.5)$$

Here it is convenient to introduce a piece of notation. Instead of writing $(q_1, p_1; \cdots; q_n, p_n)$ for a point of $(R^2)^n$, we will write $(x)_n$. If

$$(x)_n = (q_1, p_1; \cdots; q_n, p_n)$$

and if

$$(y)_m = (q_1', p_1'; \cdots; q_m', p_m'),$$

we use $(x)_n \cup (y)_m$ to denote

$$(q_1, p_1; \cdots; q_n, p_n; q_1', p_1'; \cdots; q_m', p_m') \in (R^2)^{n+m}.$$

We also write $d(x)_n$ for $dq_1\, dp_1 \cdots dq_n\, dp_n$.

We now want to consider the following situation: We start at time $t = 0$ with the equilibrium state in the box Λ with inverse temperature β and chemical potential μ (and the external potential h as well as the interparticle potential Φ). We let the state evolve in time with $\tilde{T}_\Lambda^t$ (without the external potential); we write down the correlation functions for the time-evolved state; and we study their behavior as $\Lambda \to \infty$. Physically, this situation corresponds to having a system in equilibrium in the presence of an external potential h, turning off the external potential at $t = 0$, and watching the evolution of the correlations functions as $\Lambda \to \infty$.

Thus, we want to examine the correlation functions[5]:

$$\rho_\Lambda(t; (x)_n) = \frac{1}{\Xi_\Lambda} \sum_{m=0}^{\infty} \frac{e^{\beta\mu(m+n)}}{m!} \int_{(\Lambda\times R)^m} d(x')_m \times \exp\{-\beta(H_\Lambda + h)[\tilde{T}_\Lambda^{-t}((x)_n \cup (x')_m)]\}, \quad (2.6)$$

where

$$\Xi_\Lambda = \sum_{m=0}^{\infty} \frac{e^{\beta\mu m}}{m!} \int_{(\Lambda\times R)^m} d(x')_m \exp\{-\beta(H_\Lambda + h)[(x')_m]\}. \quad (2.7)$$

Our main results can be stated as follows:

(i) If h is nonnegative, the limit as $\Lambda \to \infty$ of $\rho_\Lambda(t; (x)_n)$ exists for all t and $(x)_n$.

(ii) If h is not assumed to be nonnegative, the limit as $\Lambda \to \infty$ of $\rho_\Lambda(t; (x)_n)$ exists in the sense of distributions in $(x)_n$ for each t; the limiting distribution is actually a locally square-integrable function of $(x)_n$. In either case, we will denote the limit by $\rho(t; (x)_n)$.

(iii) The infinite-volume correlation functions $\rho(t; (x)_n)$ satisfy the BBGKY hierarchy in the following form: For any C^1 function $f(x)_n$ of compact support on $(R^2)^n$, we let

$$\rho_t(f) = \int_{(R^2)^n} \rho(t; (x)_n) f(x)_n \, d(x)_n. \quad (2.8)$$

Then $\rho_t(f)$ is a differentiable function of t and

$$\frac{d}{dt} \rho_t(f) = \rho_t(\{H, f\}) - \rho_t(f_1), \quad (2.9)$$

where

$$\{H, f\}(q_1, p_1; \cdots; q_n, p_n) = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial p_i} \frac{\partial f}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial f}{\partial p_i} \right), \quad (2.10)$$

$$f_1(q_1, p_1; \cdots; q_{n+1}, p_{n+1}) = \sum_{i=1}^{n} \frac{\partial \Phi(q_i - q_{n+1})}{\partial q_i} \frac{\partial f}{\partial p_i}, \quad (2.11)$$

$$H(q_1, p_1; \cdots; q_n, p_n) = \tfrac{1}{2} \sum_{i=1}^{n} p_i^2 + \sum_{i<j} \Phi(q_i - q_j). \quad (2.12)$$

Equation (2.9) may be obtained from the standard formal BBGKY hierarchy[5] by multiplying by the test function $f(x)_n$, integrating over (x_n), and putting the q and p derivatives on the test function by integration by parts.

(iv) If $\rho_0(x)_n$ denotes the equilibrium correlations with no external potential, then for all n, $m > 0$

$$\begin{aligned}\lim_{a\to\infty} \rho(t;\, q_1, p_1;\, \cdots;\, q_n, p_n;\, & q_{n+1} + a, p_{n+1};\, \cdots;\, q_{n+m} + a, p_{n+m}) \\ &= \rho(t;\, q_1, p_1;\, \cdots;\, q_n, p_n) \\ &\times \rho_0(q_{n+1}, p_{n+1};\, \cdots;\, q_{n+m}, p_{n+m}). \quad (2.13)\end{aligned}$$

3. INFINITE SYSTEMS

Although we did not need the theory of actually infinite systems to formulate our results, the proofs depend on this theory. We will summarize in this section the main results that we need. For more details, see Refs. 1 or 3.

A locally finite configuration of particles is defined by giving a sequence (possibly finite) of positions and momenta (q_i, p_i) such that each bounded interval in R contains only finitely many q_i. However, since the particles are supposed to be identical, we identify configurations which differ only by the labeling of the particles. Thus, a configuration may be thought of as subset of phase space R^2 with multiplicity, where the subset is just the set of occupied points and the multiplicity of each point is the number of particles at the point. We will let $\mathfrak{X}$ denote the set of all such configurations. If X and Y are configurations belonging to $\mathfrak{X}$, we let $X \cup Y$ be the configuration obtained by adding the multiplicities for X and Y. Also, if $\Lambda \subset R$, and if $X \in \mathfrak{X}$, we let $X \cap \Lambda$ denote the configuration obtained from $\mathfrak{X}$ by omitting all particles whose positions are not in Λ. The set of configurations with all particles in Λ will be denoted by $\mathfrak{X}(\Lambda)$.

We will say that a function f on $\mathfrak{X}$ is *measurable in* Λ if

$$f(X) = f\,(X \cap \Lambda)$$

for all $X \in \mathfrak{X}$. There is a simple way to construct such functions. Let ψ be a function on R^2 such that $\psi(q, p) = 0$ for $q \notin \Lambda$. Then define

$$(\sum \psi)(X) = \sum_i \psi(q_i, p_i),$$

where X is determined by (q_i, p_i). If Λ is bounded, there are only finitely many nonzero terms in this sum. Clearly, $\sum \psi$ is measurable in Λ. We give $\mathfrak{X}$ the weakest topology such that $\sum \psi$ is continuous for all continuous $\psi(q, p)$ whose support in q is bounded. It can be convincingly argued that states of classical statistical mechanics should be identified with Borel probability measures on $\mathfrak{X}$ (see Ref. 6).

If Λ is a bounded open subset of R, the mapping $X \to X \cap \Lambda$ is Borel from $\mathfrak{X}$ to $\mathfrak{X}(\Lambda)$. A Borel measure γ on $\mathfrak{X}$ defines therefore a measure γ_Λ on $\mathfrak{X}(\Lambda)$, i.e., a sequence $\gamma_{n,\Lambda}$ of symmetric Borel measures on $(\Lambda \times R)^n$, $n = 0, 1, 2, \cdots$. If each $\gamma_{n,\Lambda}$ is absolutely continuous with respect to Lebesgue measure, we define *density distributions* $f_\Lambda(x)_n$ by

$$d\gamma_{n,\Lambda} = f_\Lambda(x)_n\, d(x)_n/n!, \quad (3.1)$$

where $f_\Lambda(x)_n/n!$ is the probability density of finding precisely n particles with position $q_1, \cdots, q_n$ in Λ and momenta $p_1, \cdots, p_n$.

For any symmetric continuous function ψ on $(R^2)^n$, with compact support, we define a function $\sum \psi$ on $\mathfrak{X}$ by

$$\sum \psi(X) = \sum_{i_1 < i_2 < \cdots < i_n} \psi(q_{i_1}, p_{i_1};\, \cdots;\, q_{i_n}, p_{i_n}), \quad (3.2)$$

where the configuration X is defined by (q_i, p_i). If γ is a measure on $\mathfrak{X}$ such that $\sum \psi$ is γ-integrable for all such $\psi(x)_n$, then

$$\psi \to \int d\gamma(\sum \psi)$$

is a positive linear functional on the space of continuous symmetric functions on $(R^2)^n$ of compact support, i.e., a symmetric measure on $(R^2)^n$. When this measure exists and is absolutely continuous with respect to Lebesgue measure, so that it can be written $\rho(x)_n\, d(x)_n/n!$, we say that $\rho(x)_n$ is the nth correlation function of γ. To recapitulate: The correlation function $\rho(x)_n$ is defined by the relation

$$\int \psi(x)_n \rho(x)_n \frac{d(x)_n}{n!} = \int \sum \psi(X)\, d\gamma(X). \quad (3.3)$$

It is not hard to see that, if γ has correlation functions of all orders, then density distributions exist and, for $q_1, \cdots, q_n \in \Lambda$,

$$\begin{aligned}\rho(q_1, p_1;\, \cdots;\, q_n, p_n) &= \sum_{m=0}^{\infty} \int_{(\Lambda\times R)^m} \frac{dq_1'\, dp_1' \cdots dq_m'\, dp_m'}{m!} \\ &\times f_{\Lambda,n+m}(q_1, p_1;\, \cdots;\, q_m', p_m'). \quad (3.4)\end{aligned}$$

Conversely, if there exists a constant η such that, for all Λ and n,

$$\int_{(\Lambda\times R)^n} d(x)_n \rho(x)_n \le [\eta V(\Lambda)]^n, \quad (3.5)$$

$V(\Lambda)$ being the length of the interval Λ, the density distributions can be reexpressed in terms of the correlation functions by

$$f_\Lambda(x)_n = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \int_{(\Lambda\times R)^m} d(y)_m \rho((x)_n \cup (y)_m). \quad (3.6)$$

It has been shown[1] that, for activities satisfying (2.2), the infinite-volume limits of the correlation functions for finite-volume equilibrium ensembles exist; they have the form

$$\rho(q_1, p_1; \cdots ; q_n, p_n) = \rho(q_1, \cdots, q_n) \exp [-\tfrac{1}{2}\beta(p_1^2 + \cdots + p_n^2)], \quad (3.7)$$

where, for some real number ξ,

$$\rho(q_1, \cdots, q_n) \le \xi^n \quad (3.8)$$

for all n, $q_1, \cdots, q_n$, and hence they satisfy an estimate of the form (3.5). Thus, a measure on $\mathfrak{X}$ may be reconstructed from this set of correlation functions; we denote this measure γ_0 and call it the infinite-volume equilibrium state. It follows easily from the estimates in Ref. 7 that, if ψ is a bounded Borel function on $\mathfrak{X}$ measurable in some bounded set, then

$$\int d\gamma_0 \psi = \lim_{\Lambda\to\infty} \int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)} \psi, \quad (3.9)$$

where $\gamma_{(\Lambda)}$ is the finite-volume grand canonical ensemble density [regarded as a probability measure on $\mathfrak{X}(\Lambda)$].

4. THE EVOLUTION THEOREM

In this section we summarize the results of Ref. 3 in a form convenient for our purposes.

(i) The existence of a solution of the equations of motion has not been established for arbitrary initial data in $\mathfrak{X}$, but only for a special set $\hat{\mathfrak{X}}$ of configurations. This set $\hat{\mathfrak{X}}$ may be written as the union of a family of subsets $\hat{\mathfrak{X}}_\delta$, where δ runs through the positive real numbers. We have

$$\hat{\mathfrak{X}}_\delta \supset \hat{\mathfrak{X}}_{\delta'}, \quad \delta \ge \delta'.$$

Each $\hat{\mathfrak{X}}_\delta$ is compact in $\mathfrak{X}$. The sets $\hat{\mathfrak{X}}_\delta$ are large in the sense that

$$\gamma_0(\hat{\mathfrak{X}}) = 1, \quad \text{i.e.,} \quad \lim_{\delta\to\infty} \gamma_0(\mathfrak{X} \setminus \hat{\mathfrak{X}}_\delta) = 0.$$

In fact, a slightly stronger statement is true

$$\lim_{\delta\to\infty} \gamma_{(\Lambda)}(\mathfrak{X}(\Lambda) \cap \hat{\mathfrak{X}}_\delta) = 1$$

uniformly in Λ for large Λ.

(ii) The crux of the existence of time evolution is contained in the following statement: There is a 1-parameter group of mappings T^t of $\hat{\mathfrak{X}}$ onto itself such that, for any continuous function ψ on $\mathfrak{X}$ which is measurable in some bounded interval,

$$\lim_{\Lambda\to\infty} \psi(\tilde{T}_\Lambda^t (X \cap \Lambda)) = \psi(T^t X)$$

for $X \in \hat{\mathfrak{X}}$. The convergence is uniform for X in any fixed $\hat{\mathfrak{X}}_\delta$ and t in any bounded interval.[8] For any fixed δ, $(t, X) \to T^t X$ is continuous from $R \times \hat{\mathfrak{X}}_\delta$ to $\mathfrak{X}$. If an appropriate labeling of the particles in

$$T^t X = (q_1(t), p_1(t), q_2(t), \cdots)$$

is chosen, then the $(q_i(t), p_i(t))$ solve the differential equations

$$\frac{dq_i(t)}{dt} = p_i(t), \quad \frac{dp_i(t)}{dt} = \sum_{j\ne i} \Phi'(q_i(t) - q_j(t)).$$

(iii) The mapping $(X, X') \to X \cup X'$, sending $\mathfrak{X} \times \mathfrak{X}$ to $\mathfrak{X}$, is continuous. If $X \in \hat{\mathfrak{X}}_\delta$ and $X' \in \hat{\mathfrak{X}}_{\delta'}$, then $X \cup X' \in \hat{\mathfrak{X}}_{\delta+\delta'}$.

(iv) Every point in $(R^2)^n$ determines a point in $\mathfrak{X}$ representing a configuration of exactly n particles. We will usually fail to distinguish between $(x)_n$ as a point of $(R^2)^n$ and the corresponding point in $\mathfrak{X}$. The mapping from $(R^2)^n$ to $\mathfrak{X}$ so defined is continuous and the image of each bounded set is contained in some $\hat{\mathfrak{X}}_\delta$.

(v) The equilibrium measure γ_0 on $\hat{\mathfrak{X}}$ (more precisely, the measure obtained by restricting γ_0 to $\hat{\mathfrak{X}}$) is invariant under T^t for all t; i.e., if $E \subset \hat{X}$ is a Borel set, then

$$\gamma_0(E) = \gamma_0(T^{-t}E).$$

5. INFINITE-VOLUME LIMITS OF TIME-DEPENDENT QUANTITIES

Proposition 1: Let ϕ and ψ be functions on $\mathfrak{X}$, both measurable in some bounded interval I. We assume ϕ to be continuous and ψ to be a Borel function. We also assume that, for some real number α,

$$|\phi(X)| \le \exp [\alpha N_I(X)], \quad (5.1)$$

$$|\psi(X)| \le \exp [\alpha N_I(X)] \quad (5.2)$$

[where $N_I(X)$ is the number of particles in the interval I for the configuration X]. Then

(i) $\psi(Y)\phi(T^t Y)$ is γ_0 integrable and

$$\int_{\mathfrak{X}} d\gamma_0(Y)\psi(Y)\phi(T^t Y) = \lim_{\Lambda\to\infty} \int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)}(Y)\psi(Y)\phi(\tilde{T}_\Lambda^t Y). \quad (5.3)$$

(ii) If ϕ is further assumed to be bounded, then, for any $(x)_n \in (R^2)^n$, $\psi(Y)\phi(T^t(Y \cup (x)_n))$ is γ_0 integrable, and

$$\int_X d\gamma_0(Y)\psi(Y)\phi(T^t(Y \cup (x)_n)) = \lim_{\Lambda\to\infty} \int_{X(\Lambda)} d\gamma_{(\Lambda)}(Y)\psi(Y)\phi(\tilde{T}^t_\Lambda(Y \cup (x)_n)). \quad (5.4)$$

The integral varies continuously with $(x)_n$.

Proof: There exists a real number ξ such that the nth correlation function of $\gamma_{(\Lambda)}$, $\rho_\Lambda(x)_n$ satisfies

$$\rho_\Lambda(x)_n \le \xi^n \exp\left(-\tfrac{1}{2}\beta \sum_{i=1}^n p_i^2\right),$$
$$(x)_n = (q_1, p_1; \cdots; q_n, p_n) \quad (5.5)$$

for all n, $(x)_n$, and all sufficiently large Λ. This inequality persists for the infinite-volume correlation functions. By (3.6), the probability of finding precisely N particles in I, with respect to any $\gamma_{(\Lambda)}$ or with respect to γ_0, is majorized by

$$(n!)^{-1}[\xi(2\pi/\beta)^{\frac{1}{2}}V(I)]^n \exp[\xi(2\pi/\beta)^{\frac{1}{2}}V(I)]. \quad (5.6)$$

It follows that $\exp[\alpha N_I(Y)]$ is square-integrable with respect to each $\gamma_{(\Lambda)}$ and with respect to γ_0 and that its square-integral has an upper bound which is independent of Λ. By (5.1) and (5.2), ϕ and ψ are both γ_0 square-integrable and, since γ_0 is invariant under T^t, $\psi \circ T^t$ is also γ_0 square-integrable. By the Schwarz inequality, then, $\psi(Y)\phi(T^tY)$ is γ_0-integrable. Similarly, if ϕ is bounded, $\psi(Y)\phi(T^t(Y \cup (x)_n))$ is γ_0 integrable. By 4(ii)–(iv), $T^t(Y \cup (x)_n)$ varies continuously with $(x)_n$; hence,

$$\int d\gamma_0(Y)\psi(Y)\phi(T^t(Y \cup (x)_n))$$

is a continuous function of $(x)_n$, by the Lebesgue dominated-convergence theorem.

Because of the boundedness of the square-integrals, replacing ϕ by $-\lambda$ if $\phi \le \lambda$, ϕ if $-\lambda \le \phi \le \lambda$, and λ if $\phi \ge \lambda$ with λ large, makes a change in

$$\int d\gamma_{(\Lambda)}\psi(\phi \circ T^t)$$

which is small uniformly in Λ. Hence, in proving (5.3), we can assume that ϕ is bounded. In this case, (5.3) is a special case of (5.4). In a similar way, we see that, in proving (5.4), ψ may also be assumed to be bounded.

To prove (5.4), assuming ψ bounded, we choose $\epsilon > 0$ and then choose δ large enough so that

$$\gamma_{(\Lambda)}(\mathfrak{X}(\Lambda)\setminus\hat{\mathfrak{X}}_\delta) < \epsilon, \quad \text{for all sufficiently large } \delta\Lambda, \quad (5.7)$$

and

$$\gamma_0(\mathfrak{X}\setminus\mathfrak{X}_\delta) < \epsilon;$$

this is possible by (i). Now, by 4(ii)–(iv), the mapping

$$Y \to \Phi(T^t(Y \cup (x)_n))$$

is continuous on $\hat{\mathfrak{X}}_\delta$, and $\hat{\mathfrak{X}}_\delta$ is compact in $\mathfrak{X}$. The collection of all functions on $\hat{\mathfrak{X}}_\delta$, which are restrictions of continuous functions on $\mathfrak{X}$ measurable in bounded intervals (the interval may vary with the function), is an algebra of continuous functions on $\hat{\mathfrak{X}}_\delta$ containing the constants and separating points. Hence, by the Stone–Weierstrass theorem,[9] there is a continuous function Φ_1 on $\mathfrak{X}$, measurable in some bounded interval, such that

$$|\phi_1(Y) - \phi(T^t(Y \cup (x)_n))| < \epsilon \quad (5.8)$$

for all $Y \in \hat{\mathfrak{X}}_\delta$. We can also assume

$$\|\phi_1\|_\infty \le \|\phi\|_\infty. \quad (5.9)$$

Because

$$\lim_{\Lambda\to\infty} \phi(\tilde{T}^t_\Lambda([Y \cup (x)_n] \cap \Lambda)) = \phi(T^t(Y \cup (x)_n))$$

uniformly for $Y \in \hat{\mathfrak{X}}_\delta$ (by 4(ii)) we have

$$|\phi_1(Y) - \phi(\tilde{T}^t_\Lambda(Y \cup (x)_n))| < \epsilon \quad (5.10)$$

for all sufficiently large Λ and all $Y \in \hat{\mathfrak{X}}_\delta \cap \mathfrak{X}(\Lambda)$.

Now

$$\begin{aligned}
&\left|\int_X d\gamma_0(Y)\psi(Y)\phi(T^t(Y \cup (x)_n)) - \int_{X(\Lambda)} d\gamma_{(\Lambda)}\psi(Y)\phi(\tilde{T}^t_\Lambda(Y \cup (x)_n))\right| \\
&\le \left|\int_X d\gamma_0(y)\psi(Y)[\phi(T^t(Y \cup (x)_n)) - \phi_1(Y)]\right| \\
&\quad + \left|\int_X d\gamma_0(Y)\psi(Y)\phi_1(Y) - \int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)}(Y)\psi(Y)\phi_1(Y)\right| \\
&\quad + \left|\int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)}(Y)\psi(Y)[\phi_1(Y) - \phi(\tilde{T}^t_\Lambda(Y \cup (x)_n))]\right|.
\end{aligned} \quad (5.11)$$

The first term on the right of (5.11) is majorized by

$$\begin{aligned}
&\left|\int_{\hat{\mathfrak{X}}_\delta} d\gamma_0(Y)\psi(Y)[\phi(T^t(Y \cup (x)_n)) - \phi_1(Y)]\right| \\
&\quad + \left|\int_{\mathfrak{X}\setminus\hat{\mathfrak{X}}_\delta} d\gamma_0(Y)\psi(Y)[\phi(T^t(Y \cup (x)_n)) - \phi_1(Y)]\right| \\
&\le \|\psi\|_\infty \epsilon + \epsilon\|\psi\|_\infty(2\|\phi\|_\infty) = \epsilon\|\psi\|_\infty(1 + 2\|\phi\|_\infty).
\end{aligned}$$

[We have used (5.7), (5.8), and (5.9).] Similar arguments show that the same quantity majorizes the the third term on the right of (5.11) provided that Λ

is large enough so that (5.10) holds. Finally, the middle term on the right of (5.11) approaches zero as $\Lambda \to \infty$ by (3.9). Hence, for Λ sufficiently large,

$$\Big|\int_{\mathfrak{X}} d\gamma_0(Y)\psi(Y)\phi(T^t(Y \cup (x)_n)) - \int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)}(Y)\psi(Y)\phi(\tilde{T}^t_\Lambda(Y \cup (x)_n))\Big| \le 3\epsilon \,\|\psi\|_\infty (1 + 2\,\|\phi\|_\infty).$$

Since $\epsilon > 0$ is arbitrary, (5.4) follows.

Corollary 1: Assume the external potential h is nonnegative, and let the time-dependent finite-volume correlation functions be defined as in (2.6). Then

$$\lim_{\Lambda\to\infty} \rho_\Lambda(t; (x)_n)$$

exists for all t and $(x)_n$ and the convergence is locally uniform in $(x)_n$. Furthermore,

$$\lim_{\Lambda\to\infty} \rho_\Lambda(t; (x)_n) = e^{\beta\mu n}\int_{\mathfrak{X}} d\gamma_0(Y) \times \exp[-\beta \textstyle\sum h(T^{-t}(Y \cup (x)_n))] \times \exp\{-\beta[H(x)_n + W((x)_n, Y)]\} \times \left(\displaystyle\int_{\mathfrak{X}} d\gamma_0(Y)e^{-\beta\Sigma h(Y)}\right)^{-1}, \quad (5.12)$$

where

$$W((x)_n, Y) = \sum_{i,j} \phi(q_i - q'_j);$$

$Y = (q'_i, p'_i)$, and the limit is a continuous function of $(x)_n$.

Proof: From the definition of $\rho_\Lambda(t; (x)_n)$ and the fact that

$$H_\Lambda \circ \tilde{T}^t_\Lambda = H_\Lambda$$

(conservation of energy), it follows that

$$\rho_\Lambda(t; (x)_n) = e^{\beta\mu n}\int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)}(Y) \times \exp[-\beta \textstyle\sum h(\tilde{T}^t_\Lambda(Y \cup (x)_n))] \times \exp\{-\beta[H(x)_n + W((x)_n, Y)]\} \times \left(\displaystyle\int_{\mathfrak{X}(\Lambda)} d\gamma_{(\Lambda)}(Y)e^{-\beta\Sigma h(Y)}\right)^{-1}; \quad (5.13)$$

the corollary then follows by straightforward application of Proposition 1.

Corollary 2: Let the time-dependent finite-volume correlation functions be defined as in (2.6), and let the external potential h be any continuous function of compact support. Let $f(x)_n$ be any continuous function of compact support on $(R^2)^n$. Then

$$\lim_{\Lambda\to\infty}\int \frac{d(x)_n}{n!}\rho_\Lambda(t; (x)_n)f(x)_n = \int d\gamma_0(Y)\exp[-\beta(\textstyle\sum h)(T^{-t}Y)](\sum f)(Y) \times \left(\displaystyle\int d\gamma_0(Y)e^{-\beta\Sigma h(Y)}\right)^{-1}; \quad (5.14)$$

moreover, there exist locally square-integrable functions $\rho(t; (x)_n)$ such that

$$\lim_{\Lambda\to\infty}\int \frac{d(x)_n}{n!}\rho_\Lambda(t; (x)_n)f(x)_n = \int \frac{d(x)_n}{n!}\rho(t; (x)_n)f(x)_n, \quad (5.15)$$

for all continuous f of compact support.

Proof: Again, by the definition of the finite-volume correlation functions and the conservation of energy, we have

$$\int \frac{d(x)_n}{n!}\rho_\Lambda(t; (x)_n)f(x)_n = \int d\gamma_{(\Lambda)}(Y)\exp[-\beta \textstyle\sum h(\tilde{T}^{-t}_\Lambda Y)]\sum f(Y) \times \left(\displaystyle\int d\gamma_{(\Lambda)}(Y)e^{-\beta\Sigma h(Y)}\right)^{-1}; \quad (5.16)$$

thus (5.14) follows from Proposition 1. The existence of the infinite-volume correlation functions $\rho(t; (x)_n)$ as locally square-integrable functions (and not merely as measures) follows from the fact, easily verified, that, if Ω is any bounded open set in $(R^2)^n$, the mapping $f \to \sum f$ from the space of continuous functions with support in Ω to the space of continuous functions on $\mathfrak{X}$ extends to a continuous mapping from $L^2(\Omega, d(x)_n)$ to $L^2(\mathfrak{X}, d\gamma_0)$. Hence, since $e^{-\Sigma h(\cdot)} \in L^2(\mathfrak{X}, d\gamma_0)$, the mapping

$$f \to \lim_{\Lambda\to\infty}\int \frac{d(x)_n}{n!} f(x)_n \rho_\Lambda(t; (x)_n)$$

extends to a continuous linear functional on $L^2(\Omega)$ and is therefore given by a function square-integrable on Ω.

6. THE BBGKY HIERARCHY

Theorem 1: Let ψ be a nonnegative function on $\mathfrak{X}$ with

$$\int \psi(Y)\, d\gamma_0(Y) = 1 \quad \text{and} \quad \int \psi(Y)^2\, d\gamma_0(Y) < \infty.$$

Let γ denote the probability measure $\psi(Y)\,d\gamma_0(Y)$ on $\mathfrak{X}$, and let γ^t be the time-evolved measure defined by

$$\int \phi(Y)\,d\gamma^t(Y) = \int \phi \circ T^t(Y)\,d\gamma(Y). \quad (6.1)$$

Then γ^t has correlation functions of all orders, and these correlation functions are locally square-integrable. Moreover, for any function f which is infinitely differentiable and of compact support on $(R^2)^n$, $\int d\gamma^t \sum f$ is a differentiable function of t, and

$$\frac{d}{dt}\int \sum f\,d\gamma^t = \int \sum(\{H,f\})\,d\gamma^t - \int \sum f_1\,d\gamma^t, \quad (6.2)$$

where the notation is defined in (2.10)–(2.12).

Proof: By a simple calculation, using the invariance of γ_0 under T^t, we have

$$d\gamma^t = (\psi \circ T^t)\,d\gamma_0,$$

$$\int d\gamma_0\,|\psi \circ T^t|^2 = \int d\gamma_0\,|\psi|^2 < \infty.$$

Thus γ^t is obtained from γ_0 by multiplication by a square-integrable function. The arguments used in the proof of Corollary 2 show that this implies that γ^t has locally square-integrable correlation functions of all orders. On the other hand,

$$\int \sum f\,d\gamma^t = \int \sum [f \circ T^t(Y)]\psi(Y)\,d\gamma_0(Y). \quad (6.3)$$

It follows readily from 4(ii) that, for any infinitely differentiable f and $Y \in \hat{\mathfrak{X}}$,

$$\frac{d}{dt}\sum f(T^tY) = \sum(\{H,f\})(T^tY) - \sum f_1(T^tY). \quad (6.4)$$

The right-hand side of this expression may be verified to be γ_0-square-integrable; hence, its absolute value is γ integrable, and the integral is a bounded function of t. The complement of $\hat{\mathfrak{X}}$ has γ-measure zero. Hence, by standard theorems about differentiation under the integral sign,

$$\frac{d}{dt}\int \sum f\,d\gamma^t = \int [\sum(\{H,f\}) - \sum f_1]\,d\gamma^t, \quad (6.5)$$

which is just Eq. (6.2).

Corollary 3: Equation (2.9) holds.

Proof: The $\rho(t;(x)_n)$ are the correlation functions of the measure obtained by evolving in time the measure

$$e^{-\beta\Sigma h(Y)}\,d\gamma_0(Y)\Big/\int d\gamma_0(Y')e^{-\beta\Sigma h(Y')}$$

[by Sec. 4(iii)], and $e^{-\beta\Sigma h(Y)}$ is γ_0-square-integrable.

7. CLUSTER PROPERTIES

Let τ_a denote the operation of translation by a, acting on $\mathfrak{X}$, i.e., $\tau_a(q_i,p_i) = (q_i + a, p_i)$. The equilibrium state γ_0 is invariant under τ_a and has strong cluster properties under the action of τ_a. For a detailed discussion of these cluster properties, see Ref. 1; we will need the following fact, easily deduced from the results in this reference: If ψ and ϕ are functions on $\mathfrak{X}$ which are γ_0-square-integrable, then

$$\lim_{|a|\to\infty}\int d\gamma_0(Y)\psi(Y)\phi(\tau_aY) = \int d\gamma_0(Y)\psi(Y)\int d\gamma_0(Y)\psi(Y). \quad (7.1)$$

Using this result, we will prove the following:

Theorem 2: Let the $\rho(t;(x)_n)$ be defined as in Corollary 2, and let $\rho_0(x)_n$ denote the nth correlation function of the equilibrium measure γ_0. Then, for continuous symmetric functions $f(x)_n$, $g(y)_m$ of compact support,

$$\lim_{|a|\to\infty}\int d(x)_n\,d(y)_m f(x)_n g(y)_m \rho\,(t;(x)_n \cup \tau_a(y)_m) = \int d(x)_n f(x)_n \rho(t;(x)_n)\int d(y)_m g(y)_m \rho_0(y)_m. \quad (7.2)$$

Proof: Let $g_a(y)_m = g(\tau_a(y)_m)$, and let $\mathfrak{X}_a$ be the function on $\mathfrak{X}$ defined by

$$\mathfrak{X}_a((q_i,p_i)) = \frac{1}{(n+m)!}\sum_{i_1,\cdots,i_{n+m}}{}' f(q_{i_1},p_{i_1};\cdots;q_{i_n},p_{i_n}) \times g_a(q_{i_{n+1}},\cdots,p_{i_{n+m}}), \quad (7.3)$$

where the sum $\sum'$ is to be taken over all $(n+m)$-tuples of distinct indices. Since f and g both have compact supports, for sufficiently large a,

$$f(q_{i_1},\cdots,p_{i_n})g_a(q_{i_{n+1}},\cdots,p_{i_{n+m}}) = 0$$

if any i_k, $1 \le k \le n$, is equal to any i_e, $n+1 \le e \le n+m$. Hence, for such a,

$$\chi_a((q_i,p_i)) = \frac{1}{(n+m)!}\sum_{i_1,\cdots,i_n}{}' f(q_i,\cdots,p_{i_n}) \times \sum_{i_{n+1}\cdots,i_{n+m}}{}' g_a(q_{i_{n+1}},\cdots,p_{i_{n+m}}),$$

$$\mathfrak{X}_a(Y) = \frac{n!\,m!}{(n+m)!}(\sum f)(Y)(\sum g)(\tau_aY). \quad (7.4)$$

If we let

$$\psi(Y) = e^{-\beta\Sigma h(Y)}\left(\int d\gamma_0(Y')e^{-\beta\Sigma h(Y')}\right)^{-1}, \quad (7.5)$$

then Corollary 2 and the definition of correlation functions gives

$$\frac{1}{(n+m)!}\int d(x)_n\,d(y)_m f(x)_n g(y)_m \rho\,(t;(x)_n \cup \tau_a(y)_m) = \int d\gamma_0(Y)\psi(T^{-t}Y)\mathfrak{X}_a(Y) = \frac{n!\,m!}{(n+m)!}\int d\gamma_0(Y)\psi(T^{-t}Y)(\sum f)(Y)(\sum g)(\tau_aY) \quad (7.6)$$

for large a. Since any power of ψ, $\sum f$, or $\sum g$ is γ_0-integrable, it follows from (7.1) that

$$\lim_{|a|\to\infty} \int d(x)_n\, d(y)_m f(x)_n g(y)_m \rho\,(t; (x)_n \cup \tau_a(y)_m)$$
$$= n!\, m! \left(\int d\gamma_0(Y)\psi(T^{-t}Y) \sum f(Y)\right) \times \left(\int d\gamma_0(Y) \sum g(Y)\right)$$
$$= \left(\int d(x)_n f(x)_n \rho(t; (x)_n)\right) \times \left(\int d(y)_m g(y)_m \rho_0(y)_m\right),$$

which is just (7.2).

8. REMARKS

We have seen that the infinite-volume correlation functions $\rho(t; (x)_n)$ are the correlation functions for a measure obtained by multiplying the equilibrium measure γ_0 by $\psi \circ T^{-t}$, where ψ is defined in (7.5). We may define time-averaged correlation functions

$$\bar\rho(T, (x)_n) = T^{-1} \int_0^T dt \rho(t; (x)_n); \qquad (8.1)$$

these are the correlation functions of the measure obtained by multiplying γ_0 by

$$\bar\psi_T = T^{-1} \int_0^T dt \psi \circ T^{-t}. \qquad (8.2)$$

By the mean ergodic theorem,[10] $\bar\psi_T$ converges in $L^2(\chi, d\gamma_0)$ to some limiting function $\bar\psi_\infty$ which has γ_0-integral unity and is invariant under T^t. As in the proof of Corollary 2, the measure $\bar\gamma_\infty$ obtained by multiplying γ_0 by $\bar\psi_\infty$ has locally square-integrable correlation functions $\bar\rho(x)_n$. Trivially, we have

$$\int \frac{d(x)_n}{n!} f(x)_n \bar\rho_\infty(x)_n =$$
$$\int d\gamma_0 \bar\psi_\infty \sum f = \lim_{T\to\infty} \int d\gamma_0 \bar\psi_T \sum f$$
$$= \lim_{T\to\infty} \int \frac{d(x)_n}{n!} f(x)_n \bar\rho(T; (x)_n), \qquad (8.3)$$

i.e.,

$$\bar\rho_\infty(x)_n = \lim_{T\to\infty} \bar\rho(T; (x)_n) \qquad (8.4)$$

in the sense of distributions.

Moreover, the measure $\bar\gamma_\infty$ is time invariant and is obtained by multiplying γ_0 by a square-integrable function. Hence (Theorem 1) its correlation functions must satisfy the stationary BBGKY hierarchy. We have thus shown that the time-averaged correlation functions tend, as $T \to \infty$, to stationary correlation functions. Unfortunately, we do not know that these stationary correlation functions are the equilibrium ones.[11] This would follow if it could be proved that the equilibrium measure γ_0 is ergodic with respect to T^t. The ergodicity of the low-activity equilibrium measures is the outstanding problem in the theory of the time evolution of 1-dimensional systems, and no serious attack has yet been made on it.

Recently, Ruelle[12] has shown that, for a large class of potentials (the so-called superstable potentials) and for arbitrary temperature and activity, the finite-volume correlation functions satisfy an inequality of the form (5.5), where ξ may be chosen to be independent of Λ. One can construct infinite-volume equilibrium measures by taking limits along subsequences of boxes converging to infinity; the equilibrium measures obtained in this way need not be unique (i.e., they may depend on the particular sequence of boxes chosen), but any one of them is concentrated on $\hat{\mathfrak{X}}$, invariant under T^t and has correlation functions satisfying (5.5). It is easy to see that all our results, except those in Sec. 7, extend with appropriate modifications to apply to states obtained by making local perturbations on these equilibrium states.

* Alfred E. Sloan Foundation Fellow, also supported in part by U.S. Office Naval Research, Contract N 00014-69-A-0200-1002.

† Supported in part by the U.S. Air Force Office Special Research, under Grant 68-1416.

1 D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).

2 J. L. Lebowitz and E. Lieb, Phys. Rev. Letters **22**, 631 (1969).

3 O. E. Lanford, Commun. Math. Phys. **9**, 126 (1968); **11**, 257 (1969).

4 The existence of the equilibrium correlation functions for $h \neq 0$ for the systems considered here is a consequence of our general results. It may also be proven independently for all dimensions.

5 See, for example, N. N. Bogoliubov, J. Phys. USSR **10**, 265 (1946) ("Problems of a Dynamical Theory in Statistical Physics," transl. E. Gora, Providence College, Providence, R.I., 1959.)

6 D. Ruelle, J. Math. Phys. **8**, 1657 (1967).

7 D. Ruelle, Ann. Phys. (N.Y.) **25**, 109 (1963).

8 The last statement, which is the heart of the evolution theorem, means in effect that, if we concentrate our attention on the motion (during a finite time t) of the particles of a given configuration which are initially in a certain finite interval, their motion will not be much affected by the particles initially very far away (the actual size of the "region of influence" will, of course, depend on t and δ). It is this intuitively reasonable statement which provides the key to our ability of controlling the dynamics of our system at least to the extent of proving the rather primitive results of this paper.

9 L. H. Loomis, *Abstract Harmonic Analysis* (Van Nostrand, Princeton, N.J., 1953).

10 See P. R. Halmos, *Lectures on Ergodic Theory* (Chelsea, New York, 1956).

11 The fact that the equilibrium correlation functions, for low activity, satisfy the BBGKY hierarchy even in higher dimensions and for more general potentials has been proven by G. Gallavotti, Nuovo Cimento **52b**, 208 (1968).

12 D. Ruelle, Commun. Math. Phys., to be published.

Stationary States for a Nonlinear Wave Equation

MELVYN S. BERGER*
Belfer Graduate School of Science, Yeshiva University, New York, New York 10033

(Received 10 February 1970)

Global variational methods are used to derive mathematically rigorous results on the existence and nonexistence of stationary states for some nonlinear wave equations (e.g., the nonlinear Schrödinger equation and the complex nonlinear Klein–Gordan equation). *A priori* estimates useful in the study of stability of stationary states are also derived.

INTRODUCTION

A typical nonlinear wave equation to be discussed here can be written

$$i\frac{\partial u}{\partial t} = \Delta u + f(|\mathbf{x}|, |u|^2)u, \tag{1}$$

where $\mathbf{x}$ is a point in $\mathbb{R}^3$, $f(|\mathbf{x}|, |u^2)$ is a real-valued function, jointly continuous in $\mathbf{x}$ and $|u|^2$ with $f(|\mathbf{x}|, s) = o(1)$ as $s \to \infty$, and Δ denotes the 3-dimensional Laplacian. A complex-valued function $u(t, \mathbf{x})$ is called a stationary state for (1) if

(a) $u(t, \mathbf{x})$ satisfies (1) on $\mathbb{R}^1 \times \mathbb{R}^3$ and

(b) $u(t, \mathbf{x}) = e^{i\lambda t}v(\mathbf{x})$ where λ is some real number and $v(\mathbf{x})$ is a smooth real-valued function defined on $\mathbb{R}^3$ tending to zero exponentially as $|\mathbf{x}| \to 0$, but not identically zero.

We will suppose that $f(|\mathbf{x}|, |u|^2) = k(|\mathbf{x}|)\,|u|^\sigma$ with $0 < \sigma < 4$ and $k(|\mathbf{x}|)$ a Lipschitz continuous positive bounded function [with $0 \le k_1 \le k(|\mathbf{x}|) \le k_2 < \infty$ for all $|\mathbf{x}|$]. Then we shall show that, for fixed $\lambda < 0$, Eq. (1) possesses a countably infinite number of distinct stationary solutions $u_n(\mathbf{x}, t) = e^{i\lambda t}v_n(x)$. On the other hand, if $\sigma \ge 4$ with $f(|\mathbf{x}|, |u|^2) = |u|^\sigma$ as above, we show that (1) possesses no stationary solutions for any λ. To demonstrate the existence of the desired stationary states, we shall find the stationary states of (1) on a sequence of spheres $S_N \equiv \{\mathbf{x} \mid |\mathbf{x}| \le N\}$ and find the desired result by letting $N \to \infty$. The negative result mentioned above is obtained by proving the existence of a simple "integral invariant" for solutions of (1) and showing that for $\sigma \ge 4$ only $u(\mathbf{x}, t) \equiv 0$ satisfies such an equation. Analogous results for the nonlinear Klein–Gordan equation

$$\frac{\partial^2 u}{\partial t^2} = \Delta u - m^2 u + f(|\mathbf{x}|, |u|^2)u, \quad m \ne 0, \tag{2}$$

are discussed in the Appendix.

1. A VARIATIONAL PRINCIPLE FOR STATIONARY STATES

Setting $u(\mathbf{x}) = e^{i\lambda t}v(x)$ in (1), we obtain the following equation for $v(x)$ and λ:

$$\Delta v + f(|\mathbf{x}|, v^2)v + \lambda v = 0. \tag{3}$$

By a result of Kato,[1] if $v(\mathbf{x}) \to 0$ exponentially as $|\mathbf{x}| \to \infty$, Eq. (3) has no nonzero solutions for $\lambda > 0$. Furthermore, if we assume $f(|\mathbf{x}|, v^2)$ is homogeneous of degree $s > 0$ in v^2 and define the function $F(\mathbf{x}, v^2)$ by the relations

$$\frac{\partial F}{\partial v} = f(|\mathbf{x}|, v^2)v, \quad F(0) = 0, \tag{4}$$

then the following lemma holds.

Lemma 1: Stationary states of (1) cannot exist for $\lambda > 0$. For fixed $\lambda \le 0$, the stationary states of (1) are (apart from a constant multiplier) critical points (which vanish exponentially at infinity) of the functional

$$G(u) = \int_{\mathbb{R}^3} F(|\mathbf{x}|, v^2)\, d\mathbf{x}$$

subject to the constraint

$$\int_{\mathbb{R}^3} (\nabla v)^2 + |\lambda|\, v^2 = R, \quad R \text{ a nonzero const.}$$

Proof: Indeed, any critical point described in the statement of the lemma vanishes exponentially at infinity, is nonzero, and satisfies an equation of the form

$$\Delta \bar{v} - |\lambda|\, \bar{v} + cf(|\mathbf{x}|, \bar{v}^2)\bar{v} = 0, \quad c \ne 0. \tag{5}$$

Since f is homogeneous of degree s, $v = c^{\frac{1}{2}s}\bar{v}$ satisfies (3).

As in the Introduction, we shall assume $f(|\mathbf{x}|, v^2) = k(|\mathbf{x}|)\,|v|^\sigma$. Then, in order to find radially symmetric stationary states $v = v(|\mathbf{x}|)$, we set $|\mathbf{x}| = r$ and

$rv(r) = w(r)$. Then w vanishes at $\mathbf{x} = \mathbf{0}$ and satisfies the ordinary differential equation

$$\frac{d^2w}{dr^2} - |\lambda|\, w + r^{-\sigma}k(r)\,|w|^\sigma\, w = 0. \qquad (6)$$

The analog of Lemma 1 can then be stated.

Lemma 2: The radially symmetric stationary states of (1) (apart from a constant multiplier) are in one-to-one correspondence with the critical points (which vanish exponentially at infinity) of the functional

$$G(w) = \int_0^\infty r^{-\sigma}k(r)\,|w|^{\sigma+2}\,dr,$$

subject to the constraint

$$\int_0^\infty (\dot{w}^2 + |\lambda|\, w^2)\,dr = R, \quad R = \text{const.}$$

2. CRITICAL POINTS ON BOUNDED INTERVALS

We approximate the critical points of Lemma 2 (which we denote by C_∞) by finding the critical points C_N of the functional

$$G_N(w) = \int_0^N r^{-\sigma}k(r)\,|w|^{\sigma+2}\,dr,$$

subject to the constraint

$$\int_0^N (\dot{w}^2 + |\lambda|\, w^2) = R,$$

over the admissible class H_N consisting of all absolutely continuous functions $w(r)$ with square-integrable derivative over $[0, N]$ such that $w(0) = w(N) = 0$. The class H_N is a Hilbert space with respect to the inner product

$$(u, v)_{H_N} = \int_0^N (\dot{u}\dot{v} + |\lambda|\, uv)\,dr.$$

Thus, the constraint

$$\int_0^N (\dot{u}^2 + |\lambda|\, u^2)\,dr = R$$

implies that we allow as admissible functions those functions u with $\|u\|^2_{H_N} = R$.

In order to establish the existence of critical points C_N in the class H_N, we prove the following result.

Lemma 3: (i) For each function $w(x) \in H_\infty$ and $0 \le \sigma \le 4$, $|\lambda| \ne 0$,

$$\left(\int_0^\infty r^{-\sigma}\,|w|^{\sigma+2}\,dr\right)^{2/(\sigma+2)} \le K_{\sigma,\lambda}\int_0^\infty (\dot{w}^2 + |\lambda|\, w^2)\,dr, \qquad (7)$$

where $K_{\sigma,\lambda}$ is a constant independent of w. Furthermore,

(ii) the mapping $\mathfrak{L}$ of $H_N \to L_{\sigma+2}[0, N]$ {the space of functions $u(x)$ whose $(\sigma + 2)$th powers are integrable over $[0, N]$} defined by $\mathfrak{L}u = ur^{-\sigma/(\sigma+2)}$ is compact provided that $0 \le \sigma < 4$, and N is finite.

Proof: (i) Since $C_0^\infty(0, \infty)$ is dense in H_∞, it suffices to prove (7) for $w \in C_0^\infty(0, \infty)$. Any function $w \in C_0^\infty(0, \infty)$ can be written $w(r) = ru(r)$ where $u(r) = u(|x|) \in C_0^\infty(\mathbb{R}^3)$. Now for any function $u \in C_0^\infty(\mathbb{R}^3)$, an inequality of Nirenberg[2] shows that

$$\|u\|_{L_p} \le K_p(\|\nabla u\|_{L_2})^\alpha(\|u\|_{L_2})^{1-\alpha}$$

for $2 \le p \le 6$ with $\alpha(p) = 3(\frac{1}{2} - p^{-1})$, where K_p is a constant independent of u. Since for $a, b > 0$ and $0 \le \alpha \le 1$, $a^\alpha b^{1-\alpha} \le (a^2 + b^2)^{\frac{1}{2}}$, we obtain

$$(\|u\|_{L_p})^2 \le K_p[(\|\nabla u\|_{L_2})^2 + (\|u\|_{L_2})^2]. \qquad (8)$$

Setting $u = w(r)/r$ in (8), since

$$\int_0^\infty \frac{w^2}{r^2}\,dr \le 4\int_0^\infty \dot{w}^2\,dr,$$

we find, for $p = \sigma + 2$ with $0 \le \sigma \le 4$,

$$\left[\int_0^\infty \left(\frac{w}{r}\right)^{\sigma+2} r^2\,dr\right]^{2/(\sigma+2)} \le K_{\sigma,2}\left(\int_0^\infty (\dot{w}^2 + |\lambda|\, w^2)\,dr\right),$$

where $K_{\sigma,2} = K_{\sigma+2}\max(10, |\lambda|^{-1})$.

(ii) To prove the compactness of the map $\mathfrak{L}: H_N \to L_{\sigma+2}[0, N]$, we employ Kondrachev's lemma for functions $u(x)$ (with square-integrable gradient) defined on the ball $\Sigma_N = \{x \mid |x| \le N\}$. Again, if $w \in C_0^\infty[0, N]$ we may write $w(r) = ru(r)$ for some $u \in C_0^\infty(\Sigma_N)$. To show that the map $\mathfrak{L}$ is compact, it suffices to show that, for any sequence $w_n \in C_0^\infty[0, N]$ with H_N norms uniformly bounded, $\mathfrak{L}w_n = w_n r^{-\sigma/(\sigma+2)}$ is strongly convergent in $L_{\sigma+2}[0, N]$. Now as w_n has uniformly bounded H_N norms, it follows that $u_n = w_n/r$ and ∇u_n have uniformly bounded L_2 norms; hence by Kondrachev's lemma $u_n = w_n/r$ is strongly convergent in L_p for $2 \le p < 6$. Setting $p = \sigma + 2$ with $0 \le \sigma < 4$, we obtain

$$\lim_{m,n\to\infty} \|u_n - u_m\|_{L_{\sigma+2}} = \lim_{m,n\to\infty}\left(\int_0^N \left|\frac{w_n}{r} - \frac{w_m}{r}\right|^{\sigma+2} r^2\,dr\right)^{1/(\sigma+2)} = 0.$$

So that, as $n, m \to \infty$,

$$\|\mathfrak{L}u_n - \mathfrak{L}u_m\|_{L_{\sigma+2}} = \left(\int_0^N \left|\frac{w_n}{r^{\sigma/(\sigma+2)}} - \frac{w_m}{r^{\sigma/(\sigma+2)}}\right|^{\sigma+2} dr\right)^{1/(\sigma+2)} \to 0.$$

An immediate consequence of Lemma 3 is

Lemma 4: The functional $G_N(w)$ is continuous with respect to weak convergence in H_N for $0 \leq \sigma < 4$.

Proof: If $w_n \to w$ weakly, by Lemma 3 $\mathcal{L}w_n \to \mathcal{L}w$ strongly in $L_{\sigma+2}[0, N]$ for $0 \leq \sigma < 4$. Hence

$$\|\mathcal{L}w_n - \mathcal{L}w\|_{L_{\sigma+2}} \to 0$$

as $n \to \infty$. Now, since $|k(r)| \leq k_2$ for all r, Hölder's inequality and (7) implies for a constant k_2'

$$|G_N(w_n) - G_N(w)| \leq k_2' \int_0^N |w_n - w|^{\sigma+2} r^{-\sigma}\, dr$$
$$= k_2'(\|\mathcal{L}w_n - \mathcal{L}w\|_{L_{\sigma+2}})^{\sigma+2}.$$

Hence, $G_N(w_n) \to G_N(w)$ as $n \to \infty$.

We now prove the existence of a countable number of distinct critical points C_N by means of the calculus of variations. First we find a "ground state."

Theorem 1:

$$C_1(N) = \sup \int_0^N r^{-\sigma} k(r)\, |w|^{\sigma+2}\, dr$$

over the admissible class

$$\partial\Sigma_R = \{w \mid \|w\|_{H_N}^2 = R, \text{const}\}$$

is a critical value of $G_N(w)$ corresponding to a critical point $w_{1,N}(r)$ with nonnegative values on $[0, N]$.

Proof: First note that, by Lemma 3, $C_1(N)$ is bounded above. Let $w_n \in \partial\Sigma_R$ be such that $G_N(w_n) \to C_1(N)$. Clearly, w_n has a weakly convergent subsequence with weak limit $\bar{w}$, and $G_N(\bar{w}) = C_1(N)$ (by virtue of Lemma 4). Since $(\|\bar{w}\|_{H_N})^2 = R$ [otherwise, $G_N(\bar{w}) \neq \sup_{\partial\Sigma_R} G_N(w)$], $\bar{w}$ is the desired critical point $w_{1,N}(r)$. $\bar{w}(r) \geq 0$ on $[0, N]$ since $G(w)$ is a positive even function of w, so that we may assume $w_n \geq 0$ a.e.

Next we establish the existence of an infinite number of distinct critical points in C_N.

Theorem 2: The functional $G_N(w)$ possesses an infinite number of distinct critical values $C_n(N)$ over the admissible class $\partial\Sigma_R$. The associated critical points $w_{n,N}(r)$ satisfy the equation

$$\ddot{w} - |\lambda|\, w + \beta_{n,N} r^{-\sigma} k(r)\, |w|^\sigma w = 0, \tag{9}$$

where the $\beta_{n,N}$ are positive constants which tend to ∞ as $n \to \infty$ for fixed N.

Proof: This result is a consequence of a topological theorem of Ljusternik (Ref. 3, Theorem 2, p. 26), since $\partial\Sigma_R$ is a sphere in the Hilbert space H_N, and, by virtue of Lemma 4, $G_N(w)$ is continuous with respect to weak convergence. The smoothness of the functions $w_{n,N}(r)$ is a consequence of the classical regularity theory for such variational problems.

3. PASSAGE TO THE LIMITS AS $N \to \infty$

Here we show that as $N \to \infty$ the countably infinite number of approximate solutions $w_{n,N}$ of (9) constructed in the last section converge to a limit $w_{n,\infty}$ as $N \to \infty$. Furthermore, the limit function has the following properties: (i) $(\|w_{n,\infty}\|_{H_\infty})^2 = R$ and (ii) $w_{n,\infty}$ is a critical point of

$$\mathcal{G}(w) = \int_0^\infty k(r)\, |w|^{\sigma+2} r^{-\sigma}\, dr,$$

subject to the constraint $(\|w\|_{H_\infty})^2 = R$.

To this end, we derive the following inequalities for solutions $u(r)$ of the equation

$$u_{rr} + (2/r)u_r - |\lambda|^2 u + k(r)\, |u|^\sigma u = 0,$$
$$u(N) = 0. \tag{10}$$

(The proofs of Lemmas 5 and 6 below are due to L. E. Fraenkel.)

Lemma 5: Any solution $u(r)$ of (10) satisfies

$$|u(r)| \leq A_1/r^2, \quad \text{for} \quad 1 \leq \sigma \leq 4,$$
$$\leq A_2/r, \quad \text{for} \quad 0 < \sigma < 1,$$

where A_1 and A_2 are constants independent of u and r.

Proof: The Green's function $g_\lambda(r, \rho)$ for the operator

$$L_\lambda u = u_{rr} + (2/r)u_r - |\lambda|^2 u$$

and the boundary condition $u(N) = 0$ can be written

$$g_\lambda(r, \rho) = \frac{\rho^2}{|\lambda| \sinh(|\lambda|\, N)} \frac{\sinh |\lambda|\, (N - r)}{r} \frac{\sinh(|\lambda|\, \rho)}{\rho}, \quad r \geq \rho,$$
$$= \frac{\rho^2}{|\lambda| \sinh(|\lambda|\, N)} \frac{\sinh(|\lambda|\, r)}{r} \frac{\sinh |\lambda|\, (N - \rho)}{\rho}, \quad r \leq \rho.$$

Hence, setting

$$W(r) = \int_0^r k(r)\, |u|^{\sigma+1} \rho^2\, d\rho,$$

Eq. (10) can be rewritten

$$|u(r)| \leq \frac{1}{\lambda \sinh |\lambda|\, N} \left(\frac{\sinh |\lambda|\, (N - r)}{r} \int_0^r \frac{\sinh |\lambda|\, \rho}{\rho} W'(\rho)\, d\rho + \frac{\sinh |\lambda|\, r}{r} \int_r^N \frac{\sinh |\lambda|\, (N - \rho)}{\rho} W'(\rho)\, d\rho \right).$$

Since $(\sinh \rho)/\rho$ increases with ρ, while

$$[\sinh |\lambda|\, (N - \rho)]/\rho$$

decreases, we find after a simple computation

$$|u(r)| \le \frac{1}{\lambda \sinh(|\lambda| N)}\left(\frac{\sinh \lambda r}{r}\frac{\sinh |\lambda|(N-r)}{r} W(N)\right) \le \frac{W(N)}{\lambda \sinh(|\lambda| N)}\frac{e^{\lambda N}}{4r^2}.$$

Now we note that by Ref. 2 for $1 \le \sigma \le 4$, there is a number A independent of N and u such that

$$\int_0^N k(\rho)\,|u|^{\sigma+1}\,\rho^2\,d\rho \le A.$$

Hence, $|u(r)| \le A/r^2$.

For $0 < \sigma < 1$, we obtain a weaker result by considering the equation satisfied by u^2, namely

$$\tfrac{1}{2}\Delta u^2 - |\lambda|^2 u^2 + k(r)\,|u|^{\sigma+2} - |\nabla u|^2 = 0.$$

Setting $u^2 = w$, we find

$$w_{rr} + (2/r)w_r - 2\,|\lambda|^2 w + 2k(r)\,|u|^{\sigma+1} - 2\,|\nabla u|^2 = 0.$$

Hence, defining

$$W_*(r) = \int_0^r k\,|u|^{\sigma+2}\,\rho^2\,d\rho$$

and noting again by Ref. 2 that $W_*(r) \le A$ independent of r and u, we find as above

$$\tfrac{1}{2}u^2 \le \int_0^N g_{2\lambda}(r,\rho)(k\,|u|^{\sigma+2} - |\nabla u|^2)\,d\rho \le \int_0^N g_{2\lambda}(r,\rho)k\,|u|^{\sigma+2}\,d\rho \le \frac{A_2^2}{2r^2}.$$

Hence, $|u| \le A_2/r$ where A_2 is independent of u and N.

Lemma 5 leads to the following result concerning exponential decay of solutions of (10).

Lemma 6: For N sufficiently large, any solution $u(r)$ of (10) satisfies the following inequality:

$$|u(r)| \le A_3 \frac{\exp(-|\lambda|^{\frac{1}{2}} r)}{r} \quad \text{on } r \ge A_2\left[\max_r\left(\frac{k(r)}{|\lambda|}\right)\right]^{1/\sigma},$$

where A_2 is the constant of the previous lemma and A_3 is a constant depending only on λ, $k(r)$, and σ.

Proof: Setting $k(r) = [q(r)]^{-\sigma}$ and $ru = v$, we see that (10) becomes

$$v_{rr} - v[|\lambda| - (|v|/rq)^\sigma] = 0, \quad v(0) = v(N) = 0. \tag{11}$$

If v satisfies (11), so does $-v$, and so we may assume that $\theta = -v_r(N) > 0$ and is positive on some interval (b, N). Furthermore, where $0 < v < rq_*\,|\lambda|^{1/\sigma}$, $q_* = \min_r q(r)$, Eq. (11) implies $v_{rr} > 0$. Hence, if r_1 is the greatest r such that $r = rq_*\,|\lambda|^{1/\sigma}$, $r_1 \in (b, N)$. By Lemma 5 above, $v(r_1) \equiv r_1 q_*\,|\lambda|^{1/\sigma} \le A_2$. Hence, r_1 is bounded independently of N.

Now we bound θ as follows. Let $N - r \equiv s$ (and $N - r_1 \equiv s_1$). Then multiplying (11) by v_s and integrating (11) from $s = 0$ to a point $s \le s_1$, we obtain

$$v_s^2 = \theta^2 + |\lambda|\,v^2 - I(s), \tag{12}$$

where

$$I(s) = 2\int_0^s [(N-t)q]^{-\sigma} v^{\sigma+1} v_t\,dt \le \frac{2}{q_*^\sigma}\left((N-s)^{-\sigma}\frac{v^{\sigma+2}}{\sigma+2} - \frac{\sigma}{\sigma+2}\int_0^s (N-t)^{1-\sigma}v^{\sigma+2}\,dt\right) \le \frac{2}{\sigma+2}\frac{v^{\sigma+2}}{(rq_*)^\sigma}, \quad 0 \le s \le s_1. \tag{13}$$

Hence, writing $v(r_1) = r_1 q_*\,|\lambda|^{1/\sigma} = v_1$, we find, using (12),

$$N - r_1 = s_1 = \int_0^{v_1}\frac{ds}{dv}\,dv = \int_0^{v_1}\frac{dv}{(\theta^2 + |\lambda|\,v^2 - I)^{\frac{1}{2}}} = \frac{1}{|\lambda|^{\frac{1}{2}}}\sinh^{-1}|\lambda|^{\frac{1}{2}}\left(\frac{v_1}{\theta}\right) + \int_0^{v_1}\left(\frac{1}{(\theta^2 + |\lambda|\,v^2 - I)^{\frac{1}{2}}} - \frac{1}{(\theta^2 + |\lambda|\,v^2)^{\frac{1}{2}}}\right)dv \le \frac{1}{|\lambda|^{\frac{1}{2}}}\sinh^{-1}|\lambda|^{\frac{1}{2}}\left(\frac{v_1}{\theta}\right) + \int_0^{v_1}\frac{I\,dv}{2(\theta^2 + |\lambda|\,v^2 - I)^{\frac{3}{2}}}. \tag{14}$$

To bound the last term in the above inequality independently of N, we find from (13) that both

$$I(s) \le \frac{2}{\sigma+2}\frac{v^{\sigma+2}}{(r_1 q_*)^\sigma}, \quad \text{since } r \ge r_1,$$

and

$$I(s) \le \frac{2}{\sigma+2}\,|\lambda|\,v^2, \qquad \text{since } v \le r_1 q_*\,|\lambda|^{1/\sigma}.$$

So

$$\int_0^{v_1}\frac{I\,dv}{2(\theta^2 + |\lambda|\,v^2 - I)^{\frac{3}{2}}} \le \frac{1}{(\sigma+2)(r_1 q_*)^\sigma}\int_0^{v_1}\frac{v^{\sigma+2}\,dv}{\{|\lambda|\,v^2[2\,|\lambda|/(\sigma+2)]v^2\}^{\frac{3}{2}}} \le \left(\frac{\sigma+2}{\sigma^3\,|\lambda|^3}\right)^{\frac{1}{2}} = c_0$$

(say). Thus, (14) yields

$$\theta \le \frac{A_2 |\lambda|^{\frac{1}{2}}}{\sinh |\lambda|^{\frac{1}{2}} (N - r_1 - c_0)},$$

with $r_1 \le A_2/q_* |\lambda|^{1/\sigma}$; hence,

$$\theta \le A_3 \exp(-|\lambda|^{\frac{1}{2}} N), \quad \text{for } N \text{ sufficiently large}, \quad (15)$$

where A_3 is independent of $u(r)$ and N.

Finally, we estimate $u(r)$ in terms of the above estimate for θ. Indeed, for v on (r_1, N),

$$v_{ss} - |\lambda|\, c \le 0 \quad \text{with} \quad v\big|_{s=0} = 0 \quad \text{and} \quad v_s\big|_{s=0} = \theta.$$

Hence

$$v \le \theta \sinh s\, |\lambda|^{\frac{1}{2}} \le \tilde{A}_3 \exp(-|\lambda|^{\frac{1}{2}} N)\{\tfrac{1}{2}\exp[|\lambda|^{\frac{1}{2}}(N - r)]\}$$

by (15) and so

$$u(r) = \frac{v(r)}{r} \le \tfrac{1}{2}\tilde{A}_3 \frac{\exp(-|\lambda|^{\frac{1}{2}} r)}{r} \quad \text{on } (r_1, N).$$

Furthermore, for r_1 we have the estimate

$$r_1 \le A_2 |\lambda|^{-1/\sigma} q_*^{-1} \le A_2 \max_r \left(\frac{k(r)}{|\lambda|}\right)^{1/\sigma}.$$

We now discuss the limit of the sequence of critical points $w_{n,N}$ as $N \to \infty$.

Theorem 3: Provided that $0 < \sigma < 4$, the sequence $\{w_{n,N}(r), (N \ge N_0)\}$ has a convergent subsequence with limit $\bar{w}_{n,\infty}(r)$. The function $\bar{w}_{n,\infty}(r)$ is an element of $\partial\Sigma_R$ [i.e., $\int_0^\infty (\dot{w}^2 + |\lambda|\, w^2)\, dr = R$] and satisfies the equation

$$\ddot{w} - |\lambda|\, w + \beta_n r^{-\sigma} k(r)\, |w|^\sigma = 0, \qquad (16)$$

where

$$\beta_n = \lim_{n\to\infty} \beta_{n,N}.$$

Proof: For fixed n, we extend the functions $w_{n,N}(r)$ from $[0, N]$ to $[0, \infty)$ by defining $w_{n,N}(r) = 0$ for $r \in [N, \infty)$. Furthermore, as N increases,

$$\beta_{n,N} = \frac{\int_0^N (\dot{w}_{n,N})^2 + |\lambda|\, (w_{n,N})^2}{\int_0^N r^{-\sigma} k(r)\, |w_{n,N}|^{\sigma-2}} = \frac{R}{c_n(N)}$$

decreases, but by (7) it remains greater than or equal to $R/(K_{\sigma,\lambda}R)^{(\sigma+2)/2} > 0$. Hence $\beta_{n,N}$ tends, as $N \to \infty$, to a unique nonzero limit β_n. Now we show that the functions $\{w_{n,N}(r)\}$ have a strongly convergent subsequence in H_∞. To this end, we show that the functions $u_{n,N}(r) = w_{n,N}(r)/r$ extended to $[0, \infty)$ are conditionally compact when regarded as elements of $L_{\sigma+2}(\mathbb{R}^3)$. Since

$$\int_{\mathbb{R}^3} [u_{n,N}(|\mathbf{x}|)]^{\sigma+2}\, dx = \int_0^\infty \frac{[w_{n,N}(r)]^{\sigma+2}}{r^\sigma}\, dr \le \frac{1}{k_*}\int_0^\infty k(r) r^{-\sigma}[w_{n,N}(r)]^{\sigma+2}\, dr \le \frac{1}{k_*}(K_{\sigma,\lambda}R)^{(\sigma+2)/2},$$

by (7), where $k_* = \min_r k(r)$, it suffices by the classical theorem of compactness of Tamarkin in $L_p(\mathbb{R}^N)$ (see Smirnov[4]) to show that the following limits are uniform (independent of n):

(a) $$\lim_{h\to 0} \int_{\mathbb{R}^3} |u_{n,N}(|\mathbf{x} + \mathbf{h}|) - u_{n,N}(|\mathbf{x}|)|^{\sigma+2}\, d\mathbf{x} = 0,$$

(b) $$\lim_{K\to\infty} \int_{\mathbb{R}^3 - \Sigma_K} |u_{n,N}(|\mathbf{x}|)|^{\sigma+2} = 0,$$

where Σ_K is the sphere with radius N and the origin as center. Since $\int_D |\nabla u_{n,N}|^2$ and $\int_D |u_{n,N}|^{\sigma+2}$ are uniformly bounded for any bounded set $D \subset \mathbb{R}^3$, Rellich's lemma implies that (a) holds for $x \in D$. On the other hand, given $\epsilon > 0$, D can be chosen so large that by Lemma 6,

$$\sup_{|h|\le\epsilon} \int_{\mathbb{R}^3 - D} |u_{n,N}(|\mathbf{x} + \mathbf{h}|) - u_{n,N}(|\mathbf{x}|)|^{\sigma+2} \le K\epsilon,$$

where K is independent of n. Hence, (a) holds for $\mathbf{x} \in \mathbb{R}^3$. Furthermore, to demonstrate (b), we note that for K sufficiently large

$$\begin{aligned}\int_{\mathbb{R}^3 - \Sigma_K} |u_{n,N}(\mathbf{x})|^{\sigma+2} &\le \int_K^\infty \frac{[w_{n,N}(r)]^{\sigma+2}}{r^\sigma}\, dr \\ &\le A_3^{\sigma+2} \int_K^\infty \frac{\exp[-(\sigma+2)\, |\lambda|^{\frac{1}{2}} r]}{r^{2\sigma+2}}\, dr \quad \text{(by Lemma 8)} \\ &\le A_3^{\sigma+2} K^{-2\sigma-2} \int_K^\infty \exp[-(\sigma+2)\, |\lambda|^{\frac{1}{2}} r]\, dr \\ &\le cK^{-2\sigma-2},\end{aligned}$$

where c is a constant independent of n. Thus, the sequence $\{u_{n,N}(\mathbf{x}) = w_{n,N}(r)/r,\ (N = 0, 1, 2, \cdots)\}$ is conditionally compact in $L_{\sigma+2}(\mathbb{R}^3)$ and so has a strongly convergent subsequence with

$$\lim \frac{\bar{w}_{n,\infty}(r)}{r} = u_{n,\infty}.$$

Now, since

$$\int_0^\infty [(\dot{w}_{n,N})^2 + |\lambda|\, (w_{n,N})^2]\, dr = R,$$

we see that w_{n,N_j} has a weakly convergent subsequence in H_∞ with weak limit $\tilde{w}_{n,\infty}$. Inequality (7) and the uniqueness of the weak limit imply $\tilde{w}_{n,\infty} = \bar{w}_{n,\infty}(r)$. Furthermore, since

$$\int_0^\infty \frac{\bar{w}_{n,\infty}^{\sigma+2}(r)}{r^\sigma}\,dr > 0,$$

we have $\bar{w}_{n,\infty}(r) \not\equiv 0$. Now let $\phi \in C_0^\infty(0,\infty)$; then, for sufficiently large N_j, by virtue of (7), w_{n,N_j} satisfies the integral identity

$$\int_0^\infty (\dot{w}_{n,N_j}\dot{\phi} + |\lambda|\, w_{n,N_j}\phi) = \beta_{n,N_j}\int_0^\infty \frac{k(r)}{r^\sigma}\,|w_{n,N_j}|^\sigma\, w_{n,N_j}\phi.$$

Letting $N_j \to \infty$, we obtain

$$\int_0^\infty (\dot{\bar{w}}_{n,\infty}\dot{\phi} + |\lambda|\, \bar{w}_{n,\infty}\phi) = \beta_n\int_0^\infty \frac{k(r)}{r^\sigma}\,|\bar{w}_{n,\infty}|^\sigma\, \bar{w}_{n,\infty}\phi. \quad (17)$$

Thus, since (17) holds for arbitrary $\phi \in C_0^\infty(0,\infty)$, $\bar{w}_{n,\infty}$ is the desired solution of (16). Furthermore, $w_{n,N_j} \to \bar{w}_{n,\infty}$ strongly, since (17) holds for $\phi = \bar{w}_{n,\infty}$. Hence, $w_{n,N_j} \to \bar{w}_{n,\infty}$ weakly, while $\|w_{n,N_j}\|^2_{H_\infty} \to \|\bar{w}_{n,\infty}\|_{H_\infty}{}^2$, and so $w_{n,N_j} \to \bar{w}_{n,\infty}$ strongly. The smoothness of the solutions $\bar{w}_{n,\infty}$ follow from the regularity theory for weak solutions of elliptic partial differential equations since $\bar{w}_{n,\infty}/r$ is a weak solution of the Eq. (5).

Corollary 1: The number $\beta_n \to \infty$ as $n \to \infty$, so that an infinite number of the solutions $\bar{w}_{n,\infty}$ of (16) are distinct.

Proof: For fixed R, we consider the sequence $(\beta_{n,N})^{-1}$ defined by Eq. (9). By Theorem 2, for fixed N,

$$\lim_{n\to\infty} (\beta_{n,N}^{-1}) = 0.$$

Hence it suffices to show

$$\lim_{N\to\infty}\lim_{n\to\infty} \frac{1}{\beta_{n,N}} = \lim_{n\to\infty}\lim_{N\to\infty} \frac{1}{\beta_{n,N}}.$$

This equality follows immediately from standard results[5] since each of the single limits

$$\lim_{n\to\infty} \beta_{n,N}^{-1} \quad\text{and}\quad \lim_{N\to\infty} \beta_{n,N}^{-1}$$

exist, and $(\beta_{n,N})^{-1}$ tends monotonically to β_n^{-1}.

4. NONEXISTENCE OF STATIONARY STATES

Here, we investigate the stationary states for the equation

$$i\frac{\partial u}{\partial t} = \Delta u + |u|^\sigma u, \quad \sigma \geq 4. \quad (18)$$

Theorem 4: For $\sigma \geq 4$, Eq. (18) has no nontrivial stationary states for any $\lambda \not\equiv 0$.

Proof: Any stationary solution $u(\mathbf{x}) = e^{i\lambda x}v(\mathbf{x})$ satisfies the equation

$$\Delta v + \lambda v + |v|^\sigma v = 0, \quad (19)$$

and $v(\mathbf{x}) \to 0$ exponentially as $|x| \to \infty$. Again, by Kato,[1] only the case $\lambda \leq 0$ remains to be proven. To prove the nonexistence of such solutions for (19) with $\lambda < 0$, we first take the following result (which is proven below):

Lemma 7: Any solution $v(\mathbf{x})$ vanishing exponentially at infinity and satisfying an equation of the form $\Delta v + f(v) = 0$ also satisfies the integral identity

$$6\int_{\mathbb{R}^3} F(v) - \int_{\mathbb{R}^3} f(v)v = 0,$$

where

$$F(v) = \int_0^v f(s)\,ds. \quad (20)$$

In the present case, $f(v) = +\lambda v + |v|^\sigma v$ and $F(v) = +\frac{1}{2}\lambda v^2 + (\sigma+2)^{-1}|v|^{\sigma+2}$, so that the identity (20) becomes

$$2\lambda\int_{\mathbb{R}^3} v^2 + \left(\frac{6}{\sigma+2} - 1\right)\int_{\mathbb{R}^3} |v|^{\sigma+2} = 0.$$

Hence, for $\lambda < 0$ and $\sigma \geq 4$, this identity becomes impossible unless $v \equiv 0$, so that $u(x,t) \equiv 0$. Hence, the theorem is demonstrated with the following proof.

Proof: Set

$$W(\mathbf{x}) = (\mathbf{x}\cdot\nabla v)\nabla v,$$

where $\mathbf{x} = (x_1, x_2, x_3)$. Then, by virtue of the exponential decay of $v(\mathbf{x})$, the divergence theorem yields $\int_{\mathbb{R}^3} \operatorname{div} W(\mathbf{x})\,d\mathbf{x} = 0$. By standard results of vector analysis, we obtain

$$\begin{aligned}\operatorname{div} W(\mathbf{x}) &= \operatorname{div}\,[(\mathbf{x}\cdot\nabla v)\,\nabla v] \\ &= (\mathbf{x}\cdot\nabla v)\Delta v + \nabla v\cdot\nabla(\mathbf{x}\cdot\nabla v) \\ &= -(\mathbf{x}\cdot\nabla v)f(v) + |\nabla v|^2 + \sum_i \left(x_i\frac{\partial}{\partial x_i}\nabla v\right)\cdot\nabla v.\end{aligned}$$

Thus, we obtain the following integral identity:

$$-\int_{\mathbb{R}^3}(\mathbf{x}\cdot\nabla v)f(v)+\int_{\mathbb{R}^3}|\nabla v|^2+\sum_{i=1}^{3}\int_{\mathbb{R}^3}\nabla v\cdot\left(x_i\frac{\partial}{\partial x_i}\nabla v\right)=0.$$

Evaluating each term separately via integration by parts, using the facts that $\Delta v+f(v)=0$ and that $v(\mathbf{x})\to 0$ exponentially at ∞, we obtain

$$-\int_{\mathbb{R}^3}(\mathbf{x}\cdot\nabla v)f(v)=-\sum_{i=1}^{3}\int_{\mathbb{R}^3}x_i\frac{\partial v}{\partial x_i}f(v)=-\sum_{i=1}^{3}\int_{\mathbb{R}^3}x_i\frac{\partial F(v)}{\partial x_i}=3\int_{\mathbb{R}^3}F(v),$$

$$\int_{\mathbb{R}^3}|\nabla v|^2=-\int_{\mathbb{R}^3}v\Delta v=\int_{\mathbb{R}^3}vf(v),$$

$$\begin{aligned}\sum_{i=1}^{3}\int_{\mathbb{R}^3}\nabla v\cdot\left(x_i\frac{\partial}{\partial x_i}\nabla v\right)&=-\sum_{i=1}^{3}\int_{\mathbb{R}^3}(x_i\nabla v)_{x_i}\cdot\nabla v\\&=-3\int_{\mathbb{R}^3}|\nabla v|^2-\sum_{i=1}^{3}\int x_i(\nabla v)_{x_i}\cdot\nabla v\\&=-3\int_{\mathbb{R}^3}vf(v)-\tfrac{1}{2}\sum_{i=1}^{3}\int x_i\frac{\partial}{\partial x_i}|\nabla v|^2\\&=-3\int_{\mathbb{R}^3}vf(v)+\tfrac{3}{2}\int vf(v).\end{aligned}$$

Collecting these results, we find equality (20):

$$6\int_{\mathbb{R}^3}F(v)-\int_{\mathbb{R}^3}f(v)v=0.$$

REMARKS

If $\lambda=0$ and $\sigma=4$, then (19) has a family of solutions $v_c(\mathbf{x})=(|\mathbf{x}|^2/3c+c)^{-\frac{1}{2}}$ (where c is a constant). However these solutions do not decay exponentially at ∞. A slight refinement of the proof shows that Theorem 4 holds if $\lambda=0$ and $\sigma=4$.

Theorem 4 has also been obtained by Rosen[6] by means of his "pseudovirial theorem." A virtue of the present proof is that it can be modified to yield analogous nonexistence results for solutions of semilinear Dirichlet problems on *bounded domains* in $\mathbb{R}^N$.

ACKNOWLEDGMENTS

The author is grateful to L. E. Fraenkel for many helpful suggestions and to the referee for informing him of Rosen's paper (Ref. 6).

APPENDIX: RESULTS FOR THE NONLINEAR KLEIN-GORDAN EQUATION

As mentioned in the Introduction, analogous results hold for the equation

$$\frac{\partial^2 u}{\partial t^2}=\Delta u-m^2u+f(|\mathbf{x}|,|u|^2)u,\quad m>0.\quad (21)$$

Indeed, setting $u=\exp(i\tilde{\lambda}t)v(x)$ in (21), one obtains the following equation for v:

$$\Delta v+f(|\mathbf{x}|,v^2)v+(\tilde{\lambda}^2-m^2)v=0.\quad (22)$$

Thus, setting $\lambda=\tilde{\lambda}^2-m^2$, we see that (22) becomes identical with (3). Indeed, the results of Theorem 2, Corollary 1, and Theorem 4 yield the following result.

Theorem 5: Provided that $|\tilde{\lambda}|<m$ and

$$f(|\mathbf{x}|,|u|^2)=k(|\mathbf{x}|)\,|u|^\sigma$$

with $0<\sigma<4$, Eq. (21) has a countably infinite number of distinct stationary solutions for fixed $\tilde{\lambda}$. If either $|\tilde{\lambda}|>m$ or $\sigma\geq 4$, Eq. (21) has no stationary solutions whatsoever.

Furthermore, decay estimates analogous to those of Lemma 6 also hold for any stationary solution. In a future paper, these estimates will be used to prove the Liapunov stability of these stationary states, provided that $k(|\mathbf{x}|)$ and σ are sufficiently small for fixed m.

* Research Partially supported by an NSF grant.

[1] T. Kato, Commun. Pure Appl. Math. 12, 403 (1959).
[2] L. Nirenberg, Ann. Scuola Norm. Supl. (Pisa) 13, 115 (1959).
[3] L. Ljusternik, *The Topology of the Calculus of Variations in the Large* (American Mathematical Society, Providence, R.I., 1966), Trans. Math. Monographs.
[4] V. I. Smirnov, *Course of Higher Mathematics, Vol. 5: Integration and Functional Analysis* (Addison-Wesley, Reading, Mass., 1964).
[5] R. Bartle, *The Elements of Real Analysis* (Wiley, New York, 1966).
[6] G. Rosen, J. Math. Phys. 7, 2066 (1966).

Positivity Constraints on Crossing-Symmetric Partial-Wave Expansions

RALPH ROSKIES*
Gibbs Laboratory, Yale University, New Haven, Connecticut 06520
(Received 25 February 1970)

Based on positivity of the absorptive parts, constraints on the parameters appearing in Roskies' crossing-symmetric parametrization of the $\pi^0\pi^0$ partial waves below threshold are derived. These are then expressed as constraints involving integrals over the $\pi^0\pi^0$ s wave below threshold.

1. INTRODUCTION

It has recently been emphasized that rigorous constraints on the partial waves of the $\pi\pi$ scattering amplitudes below threshold can be of practical value in finding suitable parametrizations for the low-energy $\pi\pi$ phase shifts[1] and in evaluating models for low-energy $\pi\pi$ scattering.[2] A number of such constraints, based on crossing symmetry, analyticity, and the positivity of the absorptive parts of the amplitudes, have been derived by Martin and his collaborators.[3]

A different approach to the problem of deriving such constraints has been developed by Roskies.[2,4] Based on the work of Balachandran and Nuyts,[5] he has developed a general parametrization of the partial-wave amplitudes consistent with crossing symmetry. The parametrization, valid in the region $0 < s < 4m_\pi^2$, gave rise to relations, each involving only a finite number of partial waves, which were both necessary and sufficient to ensure the crossing symmetry of the amplitude. However, while expressing the full content of crossing symmetry, this method ignored the implications of unitarity. In this paper, we examine some restrictions on the parametrization which are imposed by the positivity of the absorptive part of the amplitude. These restrictions can then be translated into inequalities involving integrals of the s-wave amplitude below threshold. For simplicity, we restrict our attention to $\pi^0\pi^0$ scattering. (Comments on the case for $\pi\pi$ scattering with isospin will be found in the conclusions).

According to Ref. 4, the $\pi^0\pi^0$ partial waves can be parametrized as

$$f_l(s) = \sum_{\sigma=l} \sum_{p=\{\frac{1}{3}\sigma\}}^{[\frac{1}{2}\sigma]} 2(\sigma+1)(c_p^\sigma)(b_p^\sigma)_l \times (1-s)^l P_{\sigma-l}^{(2l+1,0)}(2s-1), \quad (1.1)$$

where the c_p^σ are arbitrary constants, the $P_{\sigma-l}^{(2l+1,0)}$ are Jacobi polynomials,[6] the coefficients $(b_p^\sigma)_l$ are completely known

$$(b_p^\sigma)_l = \frac{(\sigma-l)!\,(\sigma+l+1)!}{(2\sigma+1)!} \times \int_{-1}^{1} P_l(z)(z^2+3)^{3p-\sigma}(1-z^2)^{\sigma-2p}\,dz, \quad (1.2)$$

with $P_l(z)$ being Legendre polynomials; we have used the notation

$$\{\tfrac{1}{3}\sigma\} = (\text{smallest integer} \geq \tfrac{1}{3}\sigma),$$
$$[\tfrac{1}{2}\sigma] = (\text{largest integer} \leq \tfrac{1}{2}\sigma) \quad (1.3)$$

and have chosen units so that

$$4m_\pi^2 = 1. \quad (1.4)$$

Whereas (1.1) with arbitrary c_p^σ is consistent with crossing symmetry, we know[7] that, in the region $0 < s < 1$, we can also express $f_l(s)$ (for $l \geq 2$) by a Froissart–Gribov formula,

$$f_l(s) = \frac{4}{\pi(1-s)} \int_1^\infty dt A_t(s,t) Q_l\left(\frac{2t}{1-s} - 1\right), \quad l \geq 2, \quad (1.5)$$

where

$$A_t(s,t) = \sum_{l'} (2l'+1)\,\mathrm{Im} f_{l'}(t) P_{l'}\left(1 + \frac{2s}{t-1}\right) \quad (1.6)$$

with

$$\mathrm{Im} f_{l'}(t) \geq 0. \quad (1.7)$$

Expansion (1.6) converges[8] in the region $0 < s < 1$ for all $t \geq 1$. Equations (1.5)–(1.7) impose restrictions on $f_l(s)$ which must be reflected as restrictions on the arbitrary coefficients c_p^σ in (1.1). Ideally, one should find the necessary and sufficient constraints on c_p^σ to assure the validity of Eqs. (1.5)–(1.7), but this appears very difficult. Ynduráin[9] and Common[9] have found necessary and sufficient constraints on $f_l(s)$ (for $l \geq 2$ and fixed s) that allow $f_l(s)$ to be written, as in (1.5), with positive $A_t(s, t)$. By rewriting $f_l(s)$ as in (1.1), one obtains conditions on the c_p^σ, but these are very difficult to analyze.

The goal of this paper is much more modest. We shall find constraints on the c_p^σ, for $\sigma \leq 4$, which follow from (1.5)–(1.7). These coefficients are important because they control the gross features of the s and d waves of $\pi^0\pi^0$ scattering. [Since the functions $P_{\sigma-l}^{(2l+1,0)}(2s-1)$ have $\sigma - l$ zeros in the interval $0 < s < 1$, higher σ corresponds to more oscillations in s and thus to the more detailed characteristics of of the partial-wave amplitudes.] The aim of much of the analysis of the partial-wave amplitudes below threshold is to discover their general characteristics which one can either extrapolate above threshold to get a feeling for the physical phase shifts[1] or which one can use as criteria for evaluating proposed models of the low partial waves of $\pi\pi$ scattering.[2] For both these purposes, the coefficients c_p^σ with small σ are most important. Since one can express these c_p^σ as suitable integrals involving only the s wave, any restrictions on the c_p^σ can easily be translated into restrictions on such integrals.

The paper is organized as follows: In Sec. 2, we develop a general technique for deriving inequalities on the c_p^σ. This technique is applied in Sec. 3 to derive appropriate inequalities on c_1^2, c_1^3, and c_2^4. In Sec. 4, the results are presented as restrictions on integrals involving the s wave, and we comment on similar work by Piguet and Wanders.[10] Appendix A contains useful properties of the Legendre functions Q_l, while Appendix B contains expressions for the relevant Jacobi polynomials.

2. GENERAL TECHNIQUE

One can invert[6] Eq. (1.1) to solve for c_p^σ in terms of $f_l(s)$,

$$\sum_{p=\{\frac{1}{2}\sigma\}}^{[\frac{1}{2}\sigma]} (c_p^\sigma)(b_p^\sigma)_l = \int_0^1 (1-s)^{l+1} P_{\sigma-l}^{(2l+1,0)}(2s-1) f_l(s)\, ds, \tag{2.1}$$

and, inserting the representation (1.5)–(1.7), one obtains, for $l \geq 2$,

$$\sum_{p=\{\frac{1}{2}\sigma\}}^{[\frac{1}{2}\sigma]} (c_p^\sigma)(b_p^\sigma)_l = \frac{4}{\pi} \sum_{l'} \int_1^\infty dt (2l'+1)\, \mathrm{Im} f_{l'}(t) B_{l'}^{\sigma l}(t), \tag{2.2}$$

with

$$B_{l'}^{\sigma l}(t) = \int_0^1 ds (1-s)^l P_{\sigma-l}^{(2l+1,0)}(2s-1) \times Q_l\left(\frac{2t}{1-s} - 1\right) P_{l'}\left(1 + \frac{2s}{t-1}\right). \tag{2.3}$$

If we can find coefficients $\eta_{\sigma l}$ such that

$$\sum_{\sigma, l} \eta_{\sigma l} B_{l'}^{\sigma l}(t) \geq 0, \quad \text{all } l', \quad t \geq 1, \tag{2.4}$$

then

$$\sum_{\sigma, l} \eta_{\sigma l} \sum_{p=\{\frac{1}{2}\sigma\}}^{[\frac{1}{2}\sigma]} (c_p^\sigma)(b_p^\sigma)_l \geq 0, \tag{2.5}$$

because of the positivity of $\mathrm{Im} f_{l'}(t)$. Equations (2.4) and (2.5) express the full content of the positivity of $\mathrm{Im} f_{l'}(t)$ and will give the conditions on c_p^σ. However, we will not be able to find the most general $\eta_{\sigma l}$ satisfying (2.4), so that our results will be necessary but not sufficient.

To proceed with the analysis, we use the following observations:

(a) If $f(s,t) \geq 0$, $0 \leq s \leq 1$, $t \geq 1$, then

$$\int_0^1 P_{l'}\left(1 + \frac{2s}{t-1}\right) f(s,t)\, ds \geq 0, \quad \text{for all } l', \quad t \geq 1.$$

This follows from the positivity $P_{l'}(x)$ for $x \geq 1$.

(b) If

$$\int_0^1 P_{l'}\left(1 + \frac{2s}{t-1}\right) g(s,t)\, ds \geq 0, \quad \text{for all } l', \quad t \geq 1,$$

then

$$g(s,t) \geq 0 \quad \text{near} \quad s = 1.$$

This follows from the estimate[11]

$$c_0 (2l+1)^{-\frac{1}{2}} [x + (x^2-1)^{\frac{1}{2}}]^l \leq P_l(x) \leq [x + (x^2-1)^{\frac{1}{2}}]^l, \quad x \geq 1,\ c_0 \text{ a constant},$$

which establishes that, for large l',

$$P_{l'}(1 + 2s/(t-1)) / P_{l'}(1 + 2/(t-1))$$

has its support concentrated near $s = 1$. [It is assumed that $g(s,t)$ is continuous in s.]

(c) If $h(s,t)$ satisfies

(1) $h(s,t) > 0$, near $s = 1$,

(2) $h(s,t)$ has at most one zero in s in the interval $0 < s < 1$, and

(3) $\int_0^1 h(s,t)\, ds \geq 0$,

then

$$\int_0^1 h(s,t) P_{l'}\left(1 + \frac{2s}{t-1}\right) ds \geq 0, \quad \text{for all } l', \quad t \geq 1.$$

This follows because $P_l(x)$ is an increasing function of x for $x \geq 1$. Thus, the Legendre function in the expression

$$\int_0^1 h(s,t) P_{l'}\left(1 + \frac{2s}{t-1}\right) ds$$

tends to emphasize the region of larger s, where $h(s,t)$ is positive by virtue of (1) and (2).

(d) If $k(s)$ satisfies

(1) $k(s) > 0$, near $s = 1$,

(2) $k(s)$ has at most one zero in the interval $0 < s < 1$, and

(3) $\int_0^1 k(s) Q_l\left(\frac{2}{1-s} - 1\right) ds \geq 0$,

then

$$\int_0^1 k(s) Q_l\left(\frac{2t}{1-s} - 1\right) ds \geq 0, \quad \text{for all} \quad t > 1.$$

The proof is exactly analogous to the proof of (c) using the result[12] that

$$Q_l(2t/(1-s) - 1) / Q_l(2/(1-s) - 1)$$

is an increasing function of s for $t > 1$.

Our object is now to find $\eta_{\sigma l}$ such that (2.4) is valid. From (2.3), this means finding $\eta_{\sigma l}$ such that

$$\int_0^1 ds P_{l'}\left(1+\frac{2s}{t-1}\right)\sum_{\sigma,l}\eta'_{\sigma l}(1-s)^l P_{\sigma-l}^{(2l+1,0)}(2s-1) \times Q_l\left(\frac{2t}{1-s}-1\right)\geq 0, \quad \text{all } l', \quad t\geq 1. \tag{2.6}$$

By observation (b), we must have

$$F(s,t,\eta)\equiv\sum_{\sigma,l}\eta_{\sigma l}(1-s)^l P_{\sigma-l}^{(2l+1,0)}(2s-1) \times Q_l\left(\frac{2t}{1-s}-1\right)\geq 0, \quad \text{near} \quad s=1. \tag{2.7}$$

Moreover, we shall try to choose $\eta_{\sigma l}$ so that $F(s, t, \eta)$ has at most one zero in s, for fixed t. Then, by (a) and (c), it will be sufficient to verify that

$$\int_0^1 ds F(s,t,\eta)\geq 0. \tag{2.8}$$

3. ANALYSIS

As mentioned in the Introduction, we will restrict our attention to those terms with $\sigma \leq 4$. Since $\sigma \geq l$, we must have $l \leq 4$ and, since (2.2) is valid only for $l \geq 2$, we have $l = 2$ or 4. Thus, the indices σl take on the values 22, 32, 42, and 44. The corresponding expressions for $P_{\sigma-l}^{(2l+1)}(2s-1)$ are listed in Appendix B.

To begin, we consider only those terms with $l = 2$. As outlined in Sec. 2 we want to find coefficients η_{22}, η_{32}, and η_{42} such that

$$F_1(s,t)\equiv[\eta_{22}+\eta_{32}(7s-1)+\eta_{42}(36s^2-16s+1)] \times(1-s)^2Q_2(2t/(1-s)-1) \tag{3.1}$$

satisfies

(1) $\quad F_1(s,t)\geq 0, \quad \text{near} \quad s=1, \qquad$ (3.2a)

i.e.,

$$\eta_{22}+6\eta_{32}+21\eta_{42}\geq 0. \tag{3.2b}$$

(2) $F_1(s, t)$ has at most one zero in $0 < s < 1$ for fixed t; i.e., either

$$\eta_{42}\leq 0 \tag{3.3a}$$

or

$$\eta_{42}\geq 0 \quad \text{and} \quad 7\eta_{32}\geq 16\eta_{42} \tag{3.3b}$$

or

$$\eta_{42}\geq 0 \quad \text{and} \quad -\eta_{32}\geq 8\eta_{42} \tag{3.3c}$$

or

$$\eta_{42}\geq 0 \quad \text{and} \quad \tfrac{16}{7}\eta_{42}\geq\eta_{32}\geq -8\eta_{42}$$

and

$$\eta_4(\eta_2+6\eta_3+21\eta_4)\geq[\tfrac{7}{12}(\eta_3+8\eta_4)]^2. \tag{3.3d}$$

(3) $\int_0^1 F_1(s, t)\, ds \geq 0$, all $t \geq 1$.

By observation (d) this is equivalent to

$$\int_0^1 F_1(s,1)\,ds\geq 0;$$

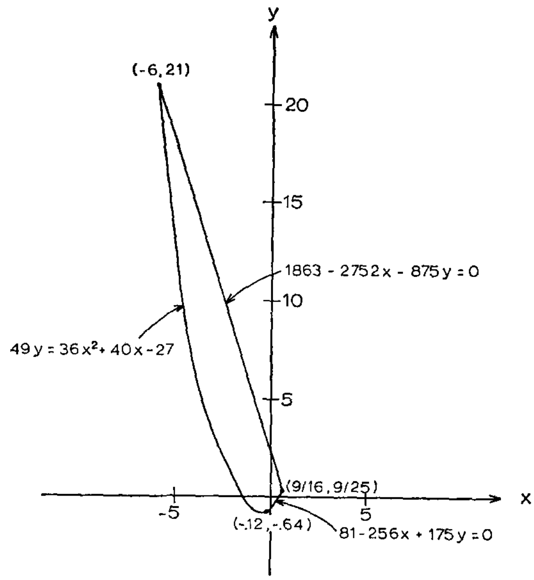


FIG. 1. Allowable (x, y) values from $l = 2$ constraints.

i.e.,

$$\tfrac{1}{18}\eta_{22}-\tfrac{1}{32}\eta_{32}+\tfrac{1}{50}\eta_{42}\geq 0. \tag{3.4}$$

Using (2.5) and (1.2), we find that, if (3.2)–(3.4) are satisfied, then

$$\tfrac{4}{15}\eta_{22}c_1^2-\tfrac{4}{105}\eta_{32}c_1^3+\tfrac{16}{315}\eta_{42}c_2^4\geq 0; \tag{3.5}$$

i.e.,

$$\eta_{22}c_1^2-\tfrac{1}{7}\eta_{32}c_1^3+\tfrac{4}{21}\eta_{42}c_2^4\geq 0. \tag{3.6}$$

Clearly, $\eta_{22} \geq 0$ and $\eta_{32} = \eta_{42} = 0$ is consistent with (3.2)–(3.4) and, therefore,

$$c_1^2\geq 0. \tag{3.7}$$

Let us define

$$x=c_1^3/7c_1^2, \tag{3.8}$$

$$y=4c_2^4/21c_1^2. \tag{3.9}$$

Then we must have

$$\eta_{22}-\eta_{32}x+\eta_{42}y\geq 0 \tag{3.10}$$

if (3.2)–(3.4) are satisfied. Omitting (3.3d), one finds that (x, y) are contained in the triangle formed by the points $(\frac{9}{16}, \frac{9}{25})$, $(-6, 21)$, $(-\frac{129}{46}, -\frac{105}{23})$, while (3.3d) implies that

$$36x^2+40x-49y-27\leq 0. \tag{3.11}$$

The region of x and y is shown in Fig. 1.

We turn now to incorporating the term with $\sigma = l = 4$. Taking $\eta_{44} > 0$, $\eta_{22} = \eta_{32} = \eta_{42} = 0$, we see trivially that $F(s, t, \eta)$ defined in (2.7) is positive and, therefore, by (2.5)

$$\eta_{44}c_2^4(b_2^4)_4\geq 0, \tag{3.12}$$

i.e.,

$$c_2^4\geq 0, \quad \text{using (1.2)},$$

i.e.

$$y \geq 0, \quad \text{using (3.7), (3.9).} \tag{3.13}$$

Now consider

$$F_2(s,t) = \eta_{22}(1-s)^2 Q_2\left(\frac{2t}{1-s}-1\right) + \eta_{44}(1-s)^4 Q_4\left(\frac{2t}{1-s}-1\right) \tag{3.14}$$

$$= \left[\eta_{22} + \eta_{44}(1-s)^2 \frac{Q_4\left(\frac{2t}{1-s}-1\right)}{Q_2\left(\frac{2t}{1-s}-1\right)}\right] \times (1-\ )^2 Q_2\left(\frac{2t}{1-s}-1\right). \tag{3.15}$$

Because we require $F_2(s,t) \geq 0$ near $s=1$, we need

$$\eta_{22} \geq 0. \tag{3.16}$$

If $\eta_{44} > 0$, then $F_2(s,t)$ is trivially positive, and no new results obtain. If η_{44} is negative, the first bracket in (3.15) is an increasing function[12] of s, and thus $F_2(s,t)$ has at most one zero. It therefore suffices to verify that

$$\int_0^1 ds F_2(s,t) \geq 0, \quad t \geq 1. \tag{3.17}$$

At $t=1$, this implies that

$$-\eta_{44}/\eta_{22} \leq 25/9. \tag{3.18}$$

One can evaluate (3.17) for arbitrary t by using the expansion[12]

$$Q_l(z-1) = \tfrac{1}{2}\sum_n \frac{(n!)^2}{(n-l)!\,(n+l+1)!}\left(\frac{2}{z}\right)^{n+1}, \quad z > 2. \tag{3.19}$$

With the value of η_{44} given by the equality in (3.18), (3.17) becomes

$$\int_0^1 ds F_2(s,t) = \tfrac{4}{9}\eta_{22} \sum_{n=2}^{\infty} \frac{1}{t^{n+1}} \frac{(n!)^2}{(n+6)!\,(n-2)!} \times (-2n^2 + 28n + 15). \tag{3.20}$$

The summand on the right is positive for $2 \leq n \leq 14$ and negative for $n \geq 15$. The expression will be most negative if one can emphasize the higher values of n, and this means taking t as small as possible. But, by construction, the sum vanishes at $t=1$ and is therefore positive for $t>1$. Consequently, (3.18) suffices for all t and, by (2.5), this implies

$$y \leq \tfrac{9}{25}. \tag{3.21}$$

There is no relation if one takes only η_{32} and η_{44} different from zero, because it is impossible to make the corresponding $F(s,t,\eta)$ satisfy (2.7) and (2.8) for large t. There is clearly no new result involving only η_{42} and η_{44} since such a relation could only involve the sign of c_2^4.

Consider now the expression

$$F_3(s,t) = (1-s)^2 Q_2(2t/(1-s)-1) \times [\eta_{22} + \eta_{32}(7s-1)] + \eta_{44}(1-s)^4 Q_4(2t/(1-s)-1) \tag{3.22}$$

$$= f_3(s,t)(1-s)^2 Q_2(2t/(1-s)-1), \tag{3.23}$$

with

$$f_3(s,t) = \eta_{22} + \eta_{32}(7s-1) + \eta_{44}(1-s)^2 \times Q_4(2t/(1-s)-1)/Q_2(2t/(1-s)-1). \tag{3.24}$$

$F_3(s,t)$ has the same zeros in $0<s<1$ as $f_3(s,t)$. Noting that

$$\frac{\partial}{\partial s} f_3(s,t) = 7\eta_{32} + \eta_{44}\frac{\partial}{\partial s}(1-s)^2 \frac{Q_4(2t/(1-s)-1)}{Q\ \ 2t/(1-s,-1)} \tag{3.25}$$

and that[12]

$$\frac{\partial}{\partial s}(1-s)^2 \frac{Q_4(2t/(1-s)-1)}{Q_2(2t/(1-s)-1)} \tag{3.26}$$

is negative and monotonic increasing in s, we see that $\partial f_3(s,t)/\partial s$ has at most one zero; thus, f_3 has at most one extremum (maximum if $\eta_{44}<0$, minimum if $\eta_{44}>0$). Since $f_3(s,t)$ must be positive near $s=1$, it has at most one zero if

$$\eta_{44} \leq 0 \tag{3.27}$$

or if

$$\eta_{44} \geq 0 \quad \text{and} \quad f_3(0,t) \leq 0 \tag{3.28}$$

or if

$$\eta_{44} \geq 0 \quad \text{and} \quad f_3(s,t) \geq 0, \quad 0 \leq s \leq 1. \tag{3.29}$$

Considering first (3.27), we impose

$$f_3(1,t) \geq 0, \quad \text{i.e.,} \quad \eta_{22} + 6\eta_{32} \geq 0 \tag{3.30}$$

and

$$\int_0^1 F_3(s,1)\,ds \geq 0, \quad \text{i.e.,} \quad \tfrac{1}{18}\eta_{22} - \tfrac{1}{32}\eta_{32} + \tfrac{1}{50}\eta_{44} \geq 0. \tag{3.31}$$

These suffice for us to establish that

$$\int_0^1 F_3(s,t)\,ds \geq 0, \quad \text{all} \quad t \geq 1, \tag{3.32}$$

using the techniques of Eqs. (3.19) and (3.20). Thus we obtain the constraint

$$6 + x - \tfrac{875}{48}y \geq 0. \tag{3.33}$$

No new results are obtained from conditions (3.28) or (3.29). Similarly, no new results are obtained by considering an expression with $\eta_{32}=0$. Moreover, it

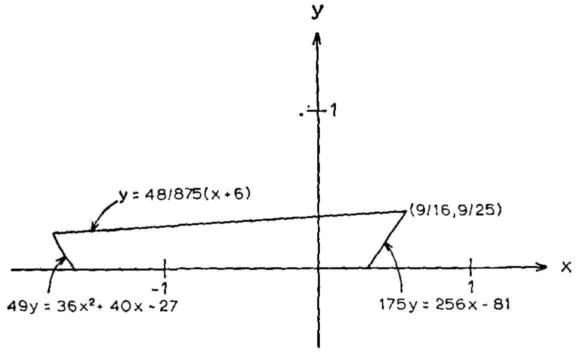


FIG. 2. Allowable (x, y) values from $l = 2$ and $l = 4$ constraints. (Note the change in scale from Fig. 1.)

is not possible to take $\eta_{22} = 0$ and to insist that the corresponding $F(s, t, \eta)$ has only one zero in s and satisfies (2.8) for large t. I have not investigated the case in which all four $\eta_{\sigma l}$ differ from zero.

Gathering all the results, we find that x and y are restricted to lie in the region indicated in Fig. 2.

4. CONCLUSION

The restrictions on x and y which we have found can now easily be expressed as restrictions on the $\pi^0\pi^0$ s wave below threshold. Writing (2.1) for $l = 0$, we find from (3.8) and (3.9)

$$x = \frac{\frac{5}{2}\int_0^1 (1-s)P_3^{(1,0)}(2s-1)f_0(s)\,ds}{\int_0^1 (1-s)P_2^{(1,0)}(2s-1)f_0(s)\,ds}, \quad (4.1)$$

$$y = \frac{\frac{5}{7}\int_0^1 (1-s)P_4^{(1,0)}(2s-1)f_0(s)\,ds}{\int_0^1 (1-s)P_2^{(1,0)}(2s-1)f_0(s)\,ds}, \quad (4.2)$$

and thus the restrictions on x and y indicated in Fig. 2 lead to inequalities on the s wave, which follow from crossing and positivity. The expressions for the Jacobi polynomials are listed in Appendix B.

After completing much of this work, I received a preprint[10] from Professor G. Wanders in which a similar analysis was applied to $\pi\pi$ scattering with isospin. Using the results of Ref. 2 and the techniques of the present paper, we can straightforwardly reproduce those results. However, there are differences between the techniques of Piguet and Wanders and those of the present paper. Instead of adopting the crossing symmetric parametrization and finding constraints on the c_p^σ as in (2.2), these authors determine positivity constraints on the functions

$$d_{ln} = \sum_{l'} \int_0^1 dt(2l'+1)\,\mathrm{Im}\,f_{l'}(t)R_{l'}^{nl}(t), \quad (4.3)$$

where

$$R_{l'}^{nl}(t) = \frac{4}{\pi}\int_0^1 ds(1-s)^l Q_l\left(\frac{2t}{1-s} - 1\right) \times P_{l'}\left(1 + \frac{2s}{t-1}\right)s^n. \quad (4.4)$$

They do not impose the constraints of crossing on the d_{ln} until they express their positivity constraints as constraints on the s-wave amplitudes. It seems to me more cumbersome to work with redundant variables which are eliminated at the end of the calculation, when they can be eliminated from the beginning. Secondly, they have found necessary and *sufficient* conditions on the d_{ln}, for $l + n \leq 3$ ($\sigma \leq 3$ in the notation of the present paper), to assure the positivity of $\mathrm{Im}\,f_{l'}(t)$, but ignoring all constraints on $\mathrm{Im}\,f_{l'}(t)$ which arise from crossing. It should be stressed that, if one takes such crossing constraints into account, their results lead only to necessary, but not sufficient conditions on the c_p^σ. This can most easily be seen by comparing their result

$$-6 < x < \tfrac{9}{16} \quad (4.5)$$

with our stronger result

$$-1.73 < x < \tfrac{9}{16}. \quad (4.6)$$

APPENDIX A

We wish to establish several properties of the functions $Q_l(x)$.

(1) $$Q_l(z-1) = \tfrac{1}{2}\sum_n \frac{(n!)^2}{(n-l)!\,(n+l+1)!}\left(\frac{2}{z}\right)^{n+1}, \quad \text{for } z > 2. \quad (A1)$$

This follows from the representation[13]

$$Q_l(z) = \frac{1}{2}\int_{-1}^1 \frac{P_l(x)}{z-x}\,dx, \quad (A2)$$

so that

$$Q_l(z-1) = \frac{1}{2}\int_{-1}^1 \frac{P_l(x)}{z-(x+1)}. \quad (A3)$$

Expanding the denominator into inverse powers of z leads to the desired result:

(2) $$\frac{2l+1}{2l+2}\frac{1}{z+(z^2-1)^{\frac{1}{2}}} \leq \frac{Q_{l+1}(z)}{Q_l(z)} \leq \frac{1}{z+(z^2-1)^{\frac{1}{2}}}, \quad z \geq 1. \quad (A4)$$

Proof: We use the integral representation[14]

$$Q_{l+1}(z) = \int_0^\infty \left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+2} dt, \tag{A5}$$

$$Q_{l+1}(z) \le \int_0^\infty \frac{1}{z + (z^2-1)^{\frac{1}{2}}} \left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+1} dt \tag{A6}$$

$$\le \frac{1}{z + (z^2-1)^{\frac{1}{2}}} Q_l(z), \tag{A7}$$

which establishes the second inequality in (A4). To establish the first inequality, we write (A5) as

$$Q_{l+1}(z) = \frac{1}{z + (z^2-1)^{\frac{1}{2}}} \times \left[\int_0^\infty dt \left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+1} + (z^2-1)^{\frac{1}{2}} \int_0^\infty dt (1 - \cosh t)\left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+2}\right] \tag{A8}$$

$$= \frac{1}{z + (z^2-1)^{\frac{1}{2}}}\left[Q_l(z) + \int_0^\infty dt(\cosh t - 1)\frac{1}{(l+1)\sinh t}\frac{d}{dt}\left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+1}\right] \tag{A9}$$

Integrating the second term by parts gives

$$Q_{l+1}(z) = \frac{1}{z + (z^2-1)^{\frac{1}{2}}}\left[Q_l(z) - \frac{1}{(l+1)}\int_0^\infty \frac{1}{(1+\cosh t)}\left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+1} dt\right] \tag{A10}$$

$$\ge \frac{1}{z + (z^2-1)^{\frac{1}{2}}}\left[Q_l(z) - \frac{1}{2(l+1)}\int_0^\infty \left(\frac{1}{z + (z^2-1)^{\frac{1}{2}} \cosh t}\right)^{l+1} dt\right] \tag{A11}$$

$$\ge \frac{1}{z + (z^2-1)^{\frac{1}{2}}}\frac{2l+1}{2l+2} Q_l(z). \qquad \text{QED} \tag{A12}$$

(3)

$$\frac{d}{dz}\frac{Q_{l+1}(z)}{Q_l(z)} \le 0, \quad z \ge 1, \tag{A13}$$

$$\frac{d^2}{dz^2}\frac{Q_{l+1}(z)}{Q_l(z)} \ge 0, \quad z \ge 1. \tag{A14}$$

Proof: (A13) has been proved by Martin.[15] Using the relations[16]

$$\frac{d}{dz} Q_{l+1}(z) = \frac{l+1}{z^2-1}[zQ_{l+1}(z) - Q_l(z)], \tag{A15}$$

$$\frac{d}{dz} Q_l(z) = \frac{l+1}{z^2-1}[Q_{l+1}(z) - zQ_l(z)], \tag{A16}$$

we see that (A14) is equivalent to

$$(y - z)[(l+1)y^2 - yz(2l+1) + l] \ge 0, \tag{A17}$$

where we have defined

$$y(z) = Q_{l+1}(z)/Q_l(z). \tag{A18}$$

Using (A4), one sees that

$$y - z \le 0 \tag{A19}$$

and, consequently, we must establish that

$$(l+1)y^2 - yz(2l+1) + l \le 0. \tag{A20}$$

But this follows immediately from (A4). QED

(4)

$$\frac{d}{ds}(1-s)\frac{Q_{l+1}(2t/(1-s)-1)}{Q_l(2t/(1-s)-1)} \le 0, \tag{A21}$$

$$\frac{d^2}{ds^2}(1-s)\frac{Q_{l+1}(2t/(1-s)-1)}{Q_l(2t/(1-s)-1)} \ge 0, \quad 0 \le s \le 1, \quad t \ge 1. \tag{A22}$$

Proof: Defining $z = 2t/(1-s) - 1$, we see that both statements follow immediately from (A13) and (A14) and the positivity of $Q_l(z)$.

(5)

$$\frac{d}{ds}(1-s)^2\frac{Q_{l+2}(2t/(1-s)-1)}{Q_l(2t/(1-s)-1)} \le 0, \tag{A23}$$

$$\frac{d^2}{ds^2}(1-s)^2\frac{Q_{l+2}(2t/(1-s)-1)}{Q_l(2t/(1-s)-1)} \ge 0, \quad 0 \le s \le 1, \quad t \ge 1. \tag{A24}$$

Proof: Both statements follow immediately from (A21) and (A22).

(6) $$\frac{d}{ds}\frac{Q_l(2t/(1-s)-1)}{Q_l(2/(1-s)-1)} \geq 0, \quad 0 \leq s \leq 1, \quad t \geq 1. \qquad \text{(A25)}$$

Proof: Defining
$$z = 2/(1-s), \qquad \text{(A26)}$$
and using (A16), we must establish that
$$\frac{y(tz-1)-y(z-1)}{t-1} + \frac{z[1-y(z-1)]}{z-2} \geq 0, \quad z \geq 2, \qquad \text{(A27)}$$
where we have defined $y(z)$ as in (A18). Using the mean-value theorem, we can write the first term of (A27) as
$$\frac{d}{dt}\, y(tz-1)|_{t=t_0}, \quad 1 \leq t_0 \leq t; \qquad \text{(A28)}$$
this expression is negative and monotonic increasing in t_0 by virtue of (A13) and (A14). Consequently, it suffices to establish (A27) at $t = 1$, i.e., to show that
$$(x-1)y'(x) + 1 - y(x) \geq 0, \quad x \geq 1. \qquad \text{(A29)}$$
Using (A15) and (A16), we see that (A29) is equivalent to
$$[(l+1)y^2 - yx(2l+1) + l] + (y-x) \leq 0, \qquad \text{(A30)}$$
which follows from (A19) and (A20).

APPENDIX B

Tabulation of some Jacobi polynomials:

$$P_0^{(5,0)}(2s-1) = P_0^{(9,0)}(2s-1) = 1,$$
$$P_1^{(5,0)}(2s-1) = 7s - 1,$$
$$P_2^{(5,0)}(2s-1) = 36s^2 - 16s + 1,$$
$$P_2^{(1,0)}(2s-1) = 10s^2 - 8s + 1,$$
$$P_3^{(1,0)}(2s-1) = 35s^3 - 45s^2 + 15s - 1,$$
$$P_4^{(1,0)}(2s-1) = 126s^4 - 224s^3 + 126s^2 - 24s + 1.$$

* Research (Report Yale 2726-562) supported by the U.S. Atomic Energy Commission under contract AT(30-1)2726.

[1] G. Auberson, O. Piguet, and G. Wanders, Phys. Letters **28B**, 41 (1968).
[2] R. Roskies, Nuovo Cimento **65A**, 467 (1970).
[3] A. Martin, Nuovo Cimento **58A**, 303 (1968); **63A**, 167 (1969). G. Auberson *et al.*, CERN Preprint TH 1032, 1969, presented at the Argonne-Purdue Conference on $\pi\pi$ interactions.
[4] R. Roskies, J. Math. Phys. **11**, 482 (1970).
[5] A. P. Balachandran and J. Nuyts, Phys. Rev. **172**, 1821 (1968).
[6] See, e.g., *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill, New York, 1953), Vol. II, Chap. X.
[7] Y. S. Jin and A. Martin, Phys. Rev. **135**, B1375 (1969).
[8] A. Martin, Nuovo Cimento **42A**, 930 (1966).
[9] F. J. Yndurâin, Nuovo Cimento **64A**, 225 (1969); A. K. Common, Nuovo Cimento **63A**, 863 (1969).
[10] O. Piguet and G. Wanders, Phys. Letters **30B**, 418 (1969).
[11] See, e.g., lectures by A. Martin in *Strong Interactions and High Energy Physics*, R. G. Moorhouse, Ed. (Oliver and Boyd, London, 1964), p. 106.
[12] See Appendix A.
[13] Reference 6, Vol. I, p. 154.
[14] Reference 6, Vol. I, p. 157.
[15] A. Martin, Nuovo Cimento **61A**, 56 (1969), Appendix A.
[16] Reference 6, Vol. I, p. 161.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Extremum Principles for Irreversible Processes*

JOSEPH B. KELLER
Department of Mathematics, University Heights and Courant Institute of Mathematical Sciences New York University, New York, N.Y. 10012

(Received 16 March 1970)

A general problem is considered for the determination of the generalized fluxes and the conjugate generalized forces describing irreversible processes in nonequilibrium thermodynamics. It is shown that for linear systems the solution of this problem is also the solution of a minimum problem and of a maximum problem. In certain cases the functional which is minimized is the rate of entropy production, while in other cases the functional which is maximized is minus this rate. Thus, in these cases Prigogine's principle of the minimum rate of entropy production is valid. For certain dynamical systems it is also shown that the functional in the minimum problem is a decreasing function of time, while for other systems the functional in the maximum problem is an increasing function of time. The results are applied to a system of chemical components in which various chemical reactions, diffusion, and heat conduction are occurring. Similar results are obtained for special nonlinear systems of this kind and for certain linear systems in which convection is also occurring. The results are based, in part, upon the theory of reciprocal variational problems, which is shown to follow from Fenchel's duality theorem involving convex functions.

1. INTRODUCTION

Nonequilibrium thermodynamics is the macroscopic theory of irreversible processes. It concerns the generalized fluxes J which describe the rates at which the processes occur, and the conjugate generalized forces X which determine the fluxes through the equation $J = LX$. The operator L is characteristic of the system under consideration. It is positive definite because the rate of entropy production, $\sigma = (X, J) = (X, LX) = (L^{-1}J, J)$, must be positive if $X \neq 0$. For linear systems in the absence of magnetic fields and rotation, it is also symmetric, as a consequence of the Onsager

Proof: Both statements follow immediately from (A21) and (A22).

(6) $$\frac{d}{ds}\frac{Q_l(2t/(1-s)-1)}{Q_l(2/(1-s)-1)} \geq 0, \quad 0 \leq s \leq 1, \quad t \geq 1. \tag{A25}$$

Proof: Defining
$$z = 2/(1-s), \tag{A26}$$
and using (A16), we must establish that
$$\frac{y(tz-1)-y(z-1)}{t-1} + \frac{z[1-y(z-1)]}{z-2} \geq 0, \quad z \geq 2, \tag{A27}$$
where we have defined $y(z)$ as in (A18). Using the mean-value theorem, we can write the first term of (A27) as
$$\frac{d}{dt}y(tz-1)|_{t=t_0}, \quad 1 \leq t_0 \leq t; \tag{A28}$$
this expression is negative and monotonic increasing in t_0 by virtue of (A13) and (A14). Consequently, it suffices to establish (A27) at $t = 1$, i.e., to show that
$$(x-1)y'(x) + 1 - y(x) \geq 0, \quad x \geq 1. \tag{A29}$$
Using (A15) and (A16), we see that (A29) is equivalent to
$$[(l+1)y^2 - yx(2l+1) + l] + (y-x) \leq 0, \tag{A30}$$
which follows from (A19) and (A20).

APPENDIX B

Tabulation of some Jacobi polynomials:

$$P_0^{(5,0)}(2s-1) = P_0^{(9,0)}(2s-1) = 1,$$
$$P_1^{(5,0)}(2s-1) = 7s - 1,$$
$$P_2^{(5,0)}(2s-1) = 36s^2 - 16s + 1,$$
$$P_2^{(1,0)}(2s-1) = 10s^2 - 8s + 1,$$
$$P_3^{(1,0)}(2s-1) = 35s^3 - 45s^2 + 15s - 1,$$
$$P_4^{(1,0)}(2s-1) = 126s^4 - 224s^3 + 126s^2 - 24s + 1.$$

* Research (Report Yale 2726-562) supported by the U.S. Atomic Energy Commission under contract AT(30-1)2726.

[1] G. Auberson, O. Piguet, and G. Wanders, Phys. Letters **28B**, 41 (1968).
[2] R. Roskies, Nuovo Cimento **65A**, 467 (1970).
[3] A. Martin, Nuovo Cimento **58A**, 303 (1968); **63A**, 167 (1969). G. Auberson *et al.*, CERN Preprint TH 1032, 1969, presented at the Argonne-Purdue Conference on $\pi\pi$ interactions.
[4] R. Roskies, J. Math. Phys. **11**, 482 (1970).
[5] A. P. Balachandran and J. Nuyts, Phys. Rev. **172**, 1821 (1968).
[6] See, e.g., *Higher Transcendental Functions*, A. Erdelyi, Ed. (McGraw-Hill, New York, 1953), Vol. II, Chap. X.
[7] Y. S. Jin and A. Martin, Phys. Rev. **135**, B1375 (1969).
[8] A. Martin, Nuovo Cimento **42A**, 930 (1966).
[9] F. J. Ynduráin, Nuovo Cimento **64A**, 225 (1969); A. K. Common, Nuovo Cimento **63A**, 863 (1969).
[10] O. Piguet and G. Wanders, Phys. Letters **30B**, 418 (1969).
[11] See, e.g., lectures by A. Martin in *Strong Interactions and High Energy Physics*, R. G. Moorhouse, Ed. (Oliver and Boyd, London, 1964), p. 106.
[12] See Appendix A.
[13] Reference 6, Vol. I, p. 154.
[14] Reference 6, Vol. I, p. 157.
[15] A. Martin, Nuovo Cimento **61A**, 56 (1969), Appendix A.
[16] Reference 6, Vol. I, p. 161.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Extremum Principles for Irreversible Processes*

JOSEPH B. KELLER
Department of Mathematics, University Heights and Courant Institute of Mathematical Sciences New York University, New York, N.Y. 10012

(Received 16 March 1970)

A general problem is considered for the determination of the generalized fluxes and the conjugate generalized forces describing irreversible processes in nonequilibrium thermodynamics. It is shown that for linear systems the solution of this problem is also the solution of a minimum problem and of a maximum problem. In certain cases the functional which is minimized is the rate of entropy production, while in other cases the functional which is maximized is minus this rate. Thus, in these cases Prigogine's principle of the minimum rate of entropy production is valid. For certain dynamical systems it is also shown that the functional in the minimum problem is a decreasing function of time, while for other systems the functional in the maximum problem is an increasing function of time. The results are applied to a system of chemical components in which various chemical reactions, diffusion, and heat conduction are occurring. Similar results are obtained for special nonlinear systems of this kind and for certain linear systems in which convection is also occurring. The results are based, in part, upon the theory of reciprocal variational problems, which is shown to follow from Fenchel's duality theorem involving convex functions.

1. INTRODUCTION

Nonequilibrium thermodynamics is the macroscopic theory of irreversible processes. It concerns the generalized fluxes J which describe the rates at which the processes occur, and the conjugate generalized forces X which determine the fluxes through the equation $J = LX$. The operator L is characteristic of the system under consideration. It is positive definite because the rate of entropy production, $\sigma = (X, J) = (X, LX) = (L^{-1}J, J)$, must be positive if $X \neq 0$. For linear systems in the absence of magnetic fields and rotation, it is also symmetric, as a consequence of the Onsager reciprocal relations. The fluxes must also satisfy the "conservation" equations of mass, momentum, energy, etc. In addition, the forces must satisfy certain equations, such as equations expressing them in terms of potentials. Thus, J and X are determined by a set of equations which usually include algebraic equations, partial differential equations, boundary conditions, etc. The problem we consider is that of finding an alternative characterization of J and X by means of variational principles.

One reason for considering this problem is to clarify "the principle of the minimum rate of entropy production" introduced by Prigogine[1] in 1945. Another is to find variational principles which apply when this one does not. Despite its significance, this principle has not been stated precisely and generally nor has its validity been demonstrated in general. As a consequence, some confusion surrounds it, such as the doubt about whether it is a new law of nature or a consequence of known laws.

We shall show that, for linear systems, J and X are the solutions of two reciprocal variational problems. (See Ref. 2, pp. 252–257.) One of these is a minimum problem for X and the other is a maximum problem for J. Furthermore, the minimum value of the functional $g(X)$ to be minimized in the first problem is equal to the maximum value of the functional $f(J)$ to be maximized in the second problem. In certain special cases,

$$f(J) = -(J, L^{-1}J) + \text{const} = -\sigma + \text{const}$$

is minus the rate of entropy production plus a constant for any J, while in other cases $g(X) = (X, LX) = \sigma$ is the rate of entropy production for any X. Thus, in these cases one of our variational principles yields the principle of the minimum rate of entropy production—in one case in terms of J and in the other in terms of X. This derivation shows that this principle, when it applies, is a consequence of known physical laws.

There are still other cases in which the common extreme value of $f(J)$ and $g(X)$ is the rate of entropy production σ for the extremizing functions, or differs from it by a known constant, but in which $f(J)$ does not equal $-\sigma$ plus a constant for arbitrary J, and $g(X)$ does not equal σ for arbitrary X. Thus, in these cases, as well as when the principle of the minimum rate of entropy production holds, we can find upper and lower bounds on σ from our two variational principles. We shall also show that for certain dynamical problems $g(X)$ decreases in time, while for other problems $f(J)$ increases with time. When $g(X) = \sigma$ or when $f(J) = -\sigma + \text{const}$, then σ decreases in time. The fact that σ decreases in time was shown for some particular problems by Prigogine,[3] and was shown for slow incompressible viscous flows by Keller, Rubenfeld, and Molyneux.[4]

As an application of the results just described, we obtain variational principles for a system of chemical components in which chemical reactions, diffusion, and heat conduction are occurring. By interpreting the fluxes and forces in two different ways, we obtain two different forms of these principles. We also show that in the time-dependent case $g(X)$ decreases in time; thus, when $g(X) = \sigma$, σ decreases in time.

As another example, we consider the convection as well as the diffusion and production of a single component, either matter or heat. We obtain similar results provided that the convection velocity divided by the diffusivity is irrotational.

Finally, we consider the nonlinear case of chemical reaction, diffusion, and heat conduction. Under rather special conditions we again obtain reciprocal variational principles and show that in the time-dependent case $g(X)$ decreases in time.

A special case of the principle of the minimum rate of entropy production was first introduced by Helmholtz.[5] This was the principle of the minimum dissipation rate for slow incompressible viscous flows. He proved that the dissipation rate was stationary, and then Korteweg[6] proved that it was a minimum when the velocity on the boundary was prescribed. Recently Keller, Rubenfeld, and Molyneux[4] have proved it for rather general boundary conditions and have extended it to cover flows containing moving solid particles, drops of liquid, and gas bubbles.

Other variational principles and their applications are described in Donnelly, Herman, and Prigogine.[7] In particular, Gage, Schiffer, Kline, and Reynolds [Ref. 6, pp. 283–286] proved that there is no variational principle of a certain kind for general nonlinear, inhomogeneous problems. Because they consider only a very special type of principle, their result does not exclude the variational principle for the nonlinear case which we present in Sec. 5.

In the course of this study, we have found that the theory of reciprocal variational principles[2] is derivable, in part, from Fenchel's[8] duality theorem involving convex functions. This is shown for quadratic variational problems in Appendix B and is discussed more generally by Brezis and Keller.[9]

2. VARIATIONAL PRINCIPLES FOR LINEAR SYSTEMS

Let us consider a dissipative thermodynamic system in which there are various generalized forces and conjugate fluxes, represented respectively by the vectors

X and J in a real Hilbert space H. The flux J is related to the force X by

$$J = LX. \tag{2.1}$$

We assume that L is a bounded, symmetric, positive-definite linear operator independent of J and X. Therefore, we consider only linear systems in the absence of magnetic fields and rotation. The rate of entropy production σ is given by

$$\sigma = (J, X). \tag{2.2}$$

A simple problem involving (2.1) is that of prescribing $X = X_0$ and finding J. Another is that of prescribing $J = J_0$ and finding X. A more general one is that of prescribing some components of X and the complementary components of J, and finding both J and X. We can state this problem as follows: Find X having certain components the same as those of a given vector X_0, and J having the rest of its components the same as those of a given vector J_0, and satisfying (2.1).

To formulate this problem precisely, we introduce a subspace Σ of H and its orthogonal complement Ω, so that $H = \Sigma + \Omega$. Then, for any two given vectors X_0 and J_0 in H, we define the sets Σ_0 and Ω_0 by $\Sigma_0 = \Sigma + X_0$ and $\Omega_0 = \Omega + J_0$. This means that any vector in Σ_0 is the sum of X_0 and any vector in Σ, while any element of Ω_0 is the sum of J_0 and any vector in Ω. Thus, $X \in \Sigma_0$ if $X - X_0 \in \Sigma$ and $J \in \Omega_0$ if $J - J_0 \in \Omega$. Now we can formulate this problem, which we shall call:

Problem P: Find an $X \in \Sigma_0$ and a $J \in \Omega_0$ such that $J = LX$.

When $\Sigma = 0$, then $\Sigma_0 = X_0$ and so $X = X_0$; this is the first problem above. When $\Sigma = H$, then $\Omega = 0$ and $\Omega_0 = J_0$, so $J = J_0$. Then this is the second problem above.

Let us denote by J^* and X^* a solution of Problem P. We shall now show that X^* is also the solution of a minimum problem, and that J^* is the solution of a maximum problem. These problems are:

Minimum Problem I: Among all $X \in \Sigma_0$ find one which minimizes

$$g(X) = (X, LX) - 2(J_0, X - X_0). \tag{2.3}$$

Maximum Problem I: Among all $J \in \Omega_0$ find one which maximizes

$$f(J) = -(J, L^{-1}J) + 2(J, X_0). \tag{2.4}$$

In Appendix A, by using the method of Courant–Hilbert[2] we prove the following theorem, in which we write $\sigma(J) \equiv (J, L^{-1}J)$ and $\sigma(X) \equiv (X, LX)$:

Theorem 1:

(i) Minimum Problem I has a unique solution X^*.

(ii) Maximum Problem I has a unique solution J^*.

(iii) $J^* = LX^*$, so J^* and X^* constitute the unique solution of Problem P.

(iv)

$$f(J^*) = \max_{J \in \Omega_0} f(J) = \min_{X \in \Sigma_0} g(X) = g(X^*).$$

(v) J^* and X^* are, respectively, the unique stationary points of $f(J)$ and $g(X)$.

(vi) When $X_0 \in \Sigma$, then $f(J) = -\sigma(J) + 2(J_0, X_0)$; and when $J_0 \in \Omega$, then $g(X) = \sigma(X)$. In these two cases, and only in these cases, the principle of the minimum rate of entropy production is valid.

(vii) When $\Sigma = 0$, i.e., when $X^* = X_0$, then $f(J^*) = g(X^*) = \sigma(X^*)$. When $\Omega = 0$, i.e., when $J^* = J_0$, then $f(J^*) = g(X^*) = -\sigma(X^*) + 2(J_0, X_0)$.

(viii) If $J \in \Omega_0$ and $X \in \Sigma_0$, then J^* lies on the sphere with center $\frac{1}{2}(J + LX)$ and radius one-half the distance from J to LX in the L^{-1} metric, i.e.,

$$\left(J^* - \frac{J + LX}{2}, L^{-1}\left[J^* - \frac{J + LX}{2}\right]\right) = \left(\frac{J - LX}{2}, L^{-1}\left[\frac{J - LX}{2}\right]\right).$$

Part (vi) of this theorem shows that the principle of the minimum rate of entropy production is valid if and only if either $X_0 \in \Sigma$ or $J_0 \in \Omega$. If $X_0 \in \Sigma$, then $\Sigma_0 = \Sigma$; while if $J_0 \in \Omega$, then $\Omega_0 = \Omega$. Thus, the principle applies if and only if either the condition on X is homogeneous or the condition on J is homogeneous. When the condition on X is homogeneous, the rate of entropy production $\sigma(X)$ expressed in terms of X is minimized; while when the condition on J is homogeneous, then the entropy production rate $\sigma(J)$ expressed in terms of J is minimized.

The two extremum problems can be made more symmetric by adding to $f(J)$ and $g(X)$ the constant $-(J_0, X_0)$. Furthermore, by replacing g by $-g$ and f by $-f$, we can obtain a maximum problem for X^* and a minimum problem for J^*.

Let us now consider a dynamical problem in which $J = J(t)$ satisfies the equation

$$\frac{dJ}{dt} = -P_\Omega[L^{-1}J - X_0]. \tag{2.5}$$

Here, P_Ω is the projection operator on Ω. We suppose that, at some time t_0,

$$J(t_0) \in \Omega_0. \tag{2.6}$$

Then we can prove the following theorem:

Theorem 2: If $J(t)$ satisfies (2.5) and (2.6), then $J(t) \in \Omega_0$, $df(J)/dt > 0$ if $J(t) \neq J^*$, and $df(J)/dt = 0$ if $J = J^*$.

To prove that $J(t) \in \Omega_0$, we see from (2.6) that this is true at $t = t_0$, and from (2.5) $dJ/dt \in \Omega$. By considering the smallest value of $|t - t_0|$ for which $J(t) \notin \Omega_0$, we arrive at a contradiction, which proves the assertion.

To prove the second part of the theorem we differentiate $f(J)$ and use (2.5) to obtain

$$\begin{aligned} \frac{df(J)}{dt} &= -2\left(\frac{dJ}{dt}, L^{-1}J - X_0\right) \\ &= 2(P_\Omega[L^{-1}J - X_0], L^{-1}J - X_0) \\ &= 2(P_\Omega[L^{-1}J - X_0], P_\Omega[L^{-1}J - X_0]) \geq 0. \end{aligned} \quad (2.7)$$

The third equality follows from the fact that the component of $L^{-1}J - X_0 \in \Sigma$ is orthogonal to any vector in Ω. The final inequality yields the theorem. This inequality shows that $f(J)$ increases until $L^{-1}J - X_0 \in \Sigma$. In the Appendix, in proving Theorem 1, we have shown that $L^{-1}J^* - X_0 \in \Sigma$ and $J^* - J_0 \in \Omega$. Therefore, when $df/dt = 0$, we have

$$L^{-1}(J^* - J) \in \Sigma \quad \text{and} \quad J^* - J \in \Omega,$$

so $(J^* - J, L^{-1}[J^* - J]) = 0$ because Σ and Ω are orthogonal. Thus, $J^* = J$ because L^{-1} is positive definite, so $f(J)$ increases unless $J = J^*$, proving the theorem.

In the same way we can prove this theorem:

Theorem 3: Let $X(t)$ satisfy

$$\frac{dX}{dt} = -P_\Sigma(LX - J_0), \quad (2.8)$$

$$X(t_0) \in \Sigma_0. \quad (2.9)$$

Then $X(t) \in \Sigma_0$, $dg(X)/dt < 0$, if $X(t) \neq X^*$ and $dg(X)/dt = 0$, if $X(t) = X^*$.

Theorem 2 shows that, when J satisfies the dynamical equation (2.5), $f(J)$ increases toward its maximum value $f(J^*)$. Theorem 3 shows that, when X satisfies (2.8), $g(X)$ decreases toward its minimum $g(X^*)$. In case $X_0 \in \Sigma$, we have $f(J) = -\sigma(J) + 2(J_0, X_0)$ and in case $J_0 \in \Omega$, we have $g(X) = \sigma(X)$. In these cases, Theorems 2 and 3, respectively, show that the rate of entropy production steadily decreases toward its minimum value. The fact that σ decreases has been shown to be so for some particular irreversible processes by Prigogine[3] and has been shown for slow flows of a viscous incompressible fluid by Keller, Rubenfeld, and Molyneux.[4]

3. APPLICATION TO CHEMICAL REACTIONS, DIFFUSION, AND HEAT CONDUCTION

Let us now treat a system of N chemical components in a domain D bounded by a surface B, in which k chemical reactions, diffusion, and heat conduction are occurring. At each point $\mathbf{r}$ of D there are N flux vectors $J_1(\mathbf{r}), \cdots, J_N(r)$ and k scalar fluxes $J_{N+1}(\mathbf{r}), \cdots, J_{N+k}(\mathbf{r})$, as well as N force vectors $X_1(\mathbf{r}), \cdots, X_N(\mathbf{r})$ and k scalar forces $X_{N+1}(\mathbf{r}), \cdots, X_{N+k}(\mathbf{r})$. The heat flux is J_1, while J_i is the flux of the ith component, $i = 2, \cdots, N$, and J_{N+j} is the rate of the jth chemical reaction, $j = 1, \cdots, k$. The X_i are the conjugate forces. The flux of the first component is determined by the fluxes of the other components, so it need not be considered explicitly. We suppose that each flux vector is related to all the force vectors at the same point, and that each scalar flux is related to the scalar forces at the same point by the equations

$$J_i(\mathbf{r}) = \sum_{j=1}^{N} L_{ij}(\mathbf{r})X_j(\mathbf{r}), \quad i = 1, \cdots, N, \quad (3.1)$$

$$J_{N+j}(\mathbf{r}) = \sum_{m=1}^{k} L_{N+j,m}(\mathbf{r})X_{N+m}(\mathbf{r}), \quad j = 1, \cdots, k. \quad (3.2)$$

In (3.1), the L_{ij} are matrices, while in (3.2) they are scalars. The Onsager reciprocal relations assert that the *st* element of L_{ij} in (3.1) is equal to the *ts* element of L_{ji}, while $L_{N+n,m} = L_{N+m,n}$ in (3.2).

When the jth chemical reaction is proceeding at unit rate, component i is assumed to be produced by it at the rate ν_{ij} per unit volume and heat at the rate ν_{1j}, $j = 1, \cdots, k$. We include the possibility that heat and the various components are also produced by some other source at the rate $\rho_i(\mathbf{r})$ per unit volume, $i = 1, \cdots, N$. Then conservation of energy and of the masses of the various components requires that the following equations hold:

$$\nabla \cdot J_i - \sum_{j=1}^{k} \nu_{ij} J_{N+j} = \rho_i, \quad i = 1, \cdots, N. \quad (3.3)$$

All the forces are assumed to be expressible in terms of N scalar potentials $\varphi_j(\mathbf{r})$, $j = 1, \cdots, N$ by the relations

$$X_j = \nabla\varphi_j, \quad j = 1, \cdots, N, \quad (3.4)$$

$$X_{N+m} = \sum_{t=1}^{N} \nu_{tm}\varphi_t, \quad m = 1, \cdots, k. \quad (3.5)$$

The potentials are related to the temperature $T(\mathbf{r})$ and the chemical potential $\mu_i(\mathbf{r})$ of component i by $\varphi_1 = T^{-1}$, $\varphi_i = T^{-1}(\mu_i - \mu_1)$, $i = 2, \cdots, N$.

To complete the description of the system, we must prescribe boundary conditions on B. We suppose that, for each i, B is decomposed into two parts B_{i1} and B_{i2}, $i = 1, \cdots, N$. Then we take as boundary conditions

$$\mathbf{n} \cdot J_i(\mathbf{r}) = \alpha_i(\mathbf{r}), \quad \mathbf{r} \text{ on } B_{i1}, \quad i = 1, \cdots, N, \quad (3.6)$$

$$\varphi_i(\mathbf{r}) = \beta_i(\mathbf{r}), \quad \mathbf{r} \text{ on } B_{i2}, \quad i = 1, \cdots, N. \quad (3.7)$$

In these conditions the functions α_i and β_i are supposed to be given, and $\mathbf{n}$ denotes the outward normal to B.

To determine the forces, fluxes, and potentials we must find X_i, J_i, $i = 1, \cdots, N + k$, and φ_j, $j = 1, \cdots, N$ satisfying (3.1)–(3.7). The L_{ij}, ρ_i, ν_{ij}, α_i, and β_i are assumed to be given. This can be reduced to the problem of finding the potentials alone, since (3.4) and (3.5) give the X_i in terms of the φ_i, while (3.1) and (3.2) give the J_i in terms of the X_i and therefore in terms of the φ_i. Upon eliminating the J_i in this way, we can write (3.3) and (3.6) as follows:

$$\nabla \cdot \sum_{j=1}^{N} L_{ij} \nabla \varphi_j - \sum_{j=1}^{N} s_{ij} \varphi_j = \rho_i(\mathbf{r}), \quad i = 1, \cdots, N, \quad (3.8)$$

$$\mathbf{n} \cdot \sum_{j=1}^{N} L_{ij} \nabla \varphi_j = \alpha_i(\mathbf{r}), \quad \mathbf{r} \text{ on } B_{i1}, \quad i = 1, \cdots, N. \quad (3.9)$$

In (3.8) we have introduced $s_{ij}(\mathbf{r})$, defined by

$$s_{ij} = \sum_{n=1}^{k} \sum_{m=1}^{k} \nu_{in} L_{N+n,m} \nu_{jm}, \quad i, j = 1, \cdots, N. \quad (3.10)$$

The problem for the potentials is thus that of solving (3.8) subject to the boundary conditions (3.7) and (3.9).

We shall now show that the problem (3.1)–(3.7) for the X_i, J_i, and φ_j is a special case of Problem P formulated in Sec. 2. Then Theorem 1 will hold for the present problem, and will yield a minimum and a maximum problem equivalent to it. First, we must define the Hilbert space H in which J and X lie. The elements of H are real vector functions of $\mathbf{r}$, defined for $\mathbf{r}$ in D and having $N + k$ components. The first N components are 3-component vectors, while the last k components are scalars. The inner product of two elements of J and X is defined by

$$(J, X) = \int_D J(\mathbf{r}) \cdot X(\mathbf{r})\, d\mathbf{r} = \int_D \sum_{j=1}^{N+k} J_j(\mathbf{r}) \cdot X_j(\mathbf{r})\, d\mathbf{r}. \quad (3.11)$$

All continuously differentiable vectors are contained in H, which is assumed to be complete with respect to the inner product (3.11).

Next we define the subspace Σ to consist of all X in H that can be written in the form (3.4) and (3.5) in terms of N potential functions φ_i satisfying $\varphi_i(\mathbf{r}) = 0$, for $\mathbf{r}$ on B_{i2}, $i = 1, \cdots, N$. The vector X_0 is expressed in terms of some particular set of potentials $\varphi_i^{(0)}$ which satisfy (3.7). Thus Σ_0 consists of all X of the form (3.4) and (3.5) for which the φ_i satisfy (3.7). The subspace Ω is defined to consist of all J in H which satisfy (3.3) with $\rho_i = 0$ and (3.6) with $\alpha_i = 0$. The vector J_0 is some particular vector satisfying (3.3) and (3.6). Thus the set Ω_0 consists of all J in H satisfying (3.3) and (3.6).

To show that $X \in \Sigma$ and $J \in \Omega$ are orthogonal, we write (3.11) in the form

$$(J, X) = \int_D \left(\sum_{i=1}^{N} J_i \cdot X_i + \sum_{j=1}^{k} J_{N+j} X_{N+j} \right) d\mathbf{r}. \quad (3.12)$$

Since $X \in \Sigma$, X is given by (3.4) and (3.5), so (3.12) becomes

$$(J, X) = \int_D \left(\sum_{i=1}^{N} J_i \cdot \nabla \varphi_i + \sum_{j=1}^{k} J_{N+j} \sum_{t=1}^{k} \nu_{tj} \varphi_t \right) d\mathbf{r}. \quad (3.13)$$

Now $J_i \cdot \nabla \varphi_i = \nabla \cdot (\varphi_i J_i) - \varphi_i \nabla \cdot J_i$. Because $J \in \Omega$, J_i satisfies (3.3) with $\rho_i = 0$ and we have

$$J_i \cdot \nabla \varphi_i = \nabla \cdot (\varphi_i J_i) - \varphi_i \sum_{j=1}^{k} \nu_{ij} J_{N+j}. \quad (3.14)$$

By using (3.14),we can write (3.13) as follows and then use Gauss's theorem to get

$$(J, X) = \int_D \sum_{i=1}^{N} \nabla \cdot (\varphi_i J_i)\, d\mathbf{r} = \sum_{i=1}^{N} \int_B \varphi_i \mathbf{n} \cdot J_i\, dS. \quad (3.15)$$

Each of the surface integrals in (3.15) vanishes since $\mathbf{n} \cdot J_i = 0$ on B_{i1} and $\varphi_i = 0$ on B_{i2}, according to the definitions of Ω and Σ. Thus, we obtain from (3.15) the desired orthogonality:

$$(J, X) = 0. \quad (3.16)$$

To prove that $H = \Sigma + \Omega$, we must show that every vector in H has a unique representation as the sum of some $J \in \Omega$ and some $X \in \Sigma$. The uniqueness follows from the orthogonality of Σ and Ω. To demonstrate the existence of the representation, we let K be any element of H, with components $K_i(\mathbf{r})$, $i = 1, \cdots, N + k$, and we seek J_i and φ_i such that

$$K_i = J_i + \nabla \varphi_i, \quad i = 1, \cdots, N, \quad (3.17)$$

$$K_{N+j} = J_{N+j} + \sum_{t=1}^{N} \nu_{tj} \varphi_t, \quad j = 1, \cdots, k. \quad (3.18)$$

We now take the divergence of (3.17) and use (3.3) with $\rho_i = 0$ to eliminate $\nabla \cdot J_i$, with the result

$$\nabla \cdot K_i = \sum_{j=1}^{k} \nu_{ij} J_{N+j} + \Delta \varphi_i, \quad i = 1, \cdots, N. \quad (3.19)$$

Now we eliminate J_{N+j} from (3.19) by using (3.18), which yields

$$\Delta \varphi_i - \sum_{j=1}^{k} \nu_{ij} \sum_{t=1}^{N} \nu_{tj} \varphi_t = \nabla \cdot K_i - \sum_{j=1}^{k} \nu_{ij} K_{N+j}, \quad i = 1, \cdots, N. \quad (3.20)$$

Equations (3.20) are a linear elliptic system of N equations for the N functions φ_i. In order that X be in Σ, φ_i must vanish on B_{i2}, while, in order for J to be in Ω, it follows from (3.6) with $\alpha_i = 0$ and from (3.17) that $n \cdot \nabla\varphi_i = n \cdot K_i$ on B_{i1}. This boundary-value problem always has a solution, provided the coefficients and the boundary satisfy appropriate regularity conditions, as we assume. Then, if X is given by (3.4) and (3.5), and if J is given by (3.17) and (3.18), it follows that $X \in \Sigma$, $J \in \Omega$, and $K = J + X$, as we wished to show.

With the above definitions, (3.1)–(3.7) is an instance of Problem P of Sec. 2, for J and L are required to be related by (3.1) and (3.2), which is of the form (2.2). For each $\mathbf{r}$ the coefficient matrix L in (3.1) and (3.2) is assumed to be symmetric and positive definite, and therefore invertible. The requirement that J satisfy (3.3) and (3.6) is just the same as the condition that $J \in \Omega_0$. The requirement that X satisfy (3.4) and (3.5) and that φ_1 satisfy (3.7) is exactly the same as the condition that $X \in \Sigma_0$. Thus, the problem is an instance of Problem P. As a consequence, Theorem 1 applies to the present problem, showing its equivalence to the maximum and minimum problems of Sec. 2.

We shall now write out for the present case the variational expressions $f(J)$ and $g(x)$ which occur in those problems. From (2.3) and (3.11) we have

$$\begin{aligned} f(J) = -\int_D \Bigg[&\sum_{i=1}^{N} J_i(\mathbf{r}) \cdot \sum_{j=1}^{N} (L^{-1})_{ij}(\mathbf{r}) J_j(\mathbf{r}) \\ &+ \sum_{j=1}^{k} J_{n+j}(\mathbf{r}) \sum_{m=1}^{k} (L^{-1})_{N+j,m}(\mathbf{r}) J_{N+m}(\mathbf{r}) \Bigg] d\mathbf{r} \\ &+ 2\int_D \Bigg[\sum_{i=1}^{N} J_i(\mathbf{r}) \cdot \nabla\varphi_i^{(0)}(\mathbf{r}) \\ &+ \sum_{j=1}^{k} J_{N+j}(\mathbf{r}) \sum_{t=1}^{N} \nu_{tj}\varphi_t^{(0)}(\mathbf{r}) \Bigg] d\mathbf{r}. \end{aligned} \tag{3.21}$$

In (3.21) we have written X_0 by means of (3.4) and (3.5) in terms of some particular potentials $\varphi_i^{(0)}$ which satisfy (3.7). Next, from (2.4) and (3.11) we obtain

$$\begin{aligned} g(X) = -\int_D \Bigg[&\sum_{i=1}^{N} X_i(\mathbf{r}) \cdot \sum_{j=1}^{N} L_{ij}(\mathbf{r}) X_j(\mathbf{r}) \\ &+ \sum_{j=1}^{k} X_{N+j}(\mathbf{r}) \sum_{m=1}^{k} L_{N+j,m}(\mathbf{r}) X_{N+m}(\mathbf{r}) \Bigg] d\mathbf{r} \\ &- 2\int_D \Bigg[\sum_{i=1}^{N+k} J_i^{(0)}(\mathbf{r}) \cdot X_i(\mathbf{r}) \\ &- 2\sum_{i=1}^{N} J_i^{(0)}(\mathbf{r}) \cdot \nabla\varphi_i^{(0)}(\mathbf{r}) \\ &- 2\sum_{j=1}^{k} J_{N+j}^{(0)}(\mathbf{r}) \sum_{t=1}^{N} \nu_{tj}\varphi_t^{(0)}(\mathbf{r}) \Bigg] dr. \end{aligned} \tag{3.22}$$

Here $J^{(0)}$ is a particular vector satisfying (3.3) and (3.6).

The extremum problems are now the following:

Minimum Problem II: Among all vectors $X_i(\mathbf{r})$, $i = 1, \cdots, N + k$, in H of the form (3.4) and (3.5) in which the φ_i satisfy (3.7), find one which minimizes $g(X)$.

Maximum Problem II: Among all vectors $J_i(\mathbf{r})$, $i = 1, \cdots, N + k$, in H satisfying (3.3) and (3.6), find one which maximizes $f(J)$.

Theorem 1 applies to these two problems.

It is of interest to note that we can choose $\varphi_i^{(0)} \equiv 0$ if $\beta_i(\mathbf{r}) = 0$ on B_{i2}, $i = 1, \cdots, N$. In this case $X_0 = 0$ and part (vi) of Theorem 1 shows that $f(J) = -\sigma$. Similarly, we can choose $J_0 = 0$ if $\rho_i = 0$ and $\alpha_i = 0$, $i = 1, \cdots, N$. Then part (vi) shows that $g(X) = \sigma$. Thus in both of these cases we obtain the principle of the minimum rate of entropy production.

We shall now show that the functions $\varphi_i^{(0)}$ and J_0, which are largely arbitrary, can be essentially eliminated from (3.21) and (3.22). We begin with the last integral in (3.21) and write

$$\begin{aligned} J_i \cdot \nabla\varphi_i^{(0} &= \nabla \cdot (\varphi_i^{(0)} J_i) - \varphi_i^{(0)} \nabla \cdot J_i \\ &= \nabla \cdot (\varphi_i^{(0)} J_i) - \varphi_i^{(0)}\rho_i - \varphi_i^{(0)} \sum_{j=1}^{k} \nu_{ij} J_{N+j}. \end{aligned} \tag{3.23}$$

The last equality follows from (3.3). When (3.23) is used in (3.21), the final sums cancel. Then use of the divergence theorem and the boundary conditions (3.6) and (3.7) yields

$$\begin{aligned} f(J) = -\int_D \Bigg[&\sum_{i=1}^{N} J_i \cdot \sum_{j=1}^{N} (L^{-1})_{ij} J_j \\ &+ \sum_{j=1}^{k} J_{N+j} \sum_{m=1}^{k} (L^{-1})_{N+j,m} J_{N+m} + 2\sum_{i=1}^{N} \varphi_i^{(0)}\rho_i \Bigg] dr \\ &+ 2\sum_{i=1}^{N} \left[\int_{B_{i1}} \varphi_i^{(0)}\alpha_i \, dS + \int_{B_{i2}} \beta_i \mathbf{n} \cdot J_i \, dS \right]. \end{aligned} \tag{3.24}$$

We see that, although $\varphi_i^{(0)}$ still occurs in (3.24), it occurs only in constant terms, i.e., in terms independent of J.

Similarly, to eliminate J_0 from (3.22), we use (3.4) and (3.5) to express X in terms of the φ_i. We then write the ith component of J_0 as $J_i^{(0)}$ and we have

$$\begin{aligned} J_i^{(0)} \cdot \nabla(\varphi_i - \varphi_i^{(0)}) &= \nabla \cdot [(\varphi_i - \varphi_i^{(0)}) J_i^{(0)}] \\ &- (\varphi_i - \varphi_i^{(0)}) \sum_{j=1}^{k} \nu_{ij} J_{N+j}^{(0)} - (\varphi_i - \varphi_i^{(0)})\rho_i. \end{aligned} \tag{3.25}$$

Now we use (3.25) in the last integral in (3.22), use the divergence theorem and boundary conditions, and eliminate X from the first integral to obtain

$$g(\varphi) = \int_D \sum_{i,j=1}^{N} [\nabla\varphi_i \cdot L_{ij}\nabla\varphi_j + s_{ij}\varphi_i\varphi_j]\, d\mathbf{r} - 2\sum_{i=1}^{N}\left[\int_{B_{i1}} (\varphi_i - \varphi_i^{(0)})\alpha_i\, dS - \int_D (\varphi_i - \varphi_i^{(0)})\rho_i\, d\mathbf{r}\right]. \quad (3.26)$$

We have written $g(\varphi)$ instead of $g(X)$ because only the φ_i occur in (3.26). Again the $\varphi_i^{(0)}$ occur only in constant terms, i.e., in terms independent of the φ_i.

We shall conclude this section by stating the consequences of Theorem 1 for the present case, assuming that all the prescribed functions are sufficiently regular. First we reformulate the minimum problem:

Minimum Problem III: Among all $\varphi = (\varphi_1, \cdots, \varphi_N)$ satisfying (3.7), such that X given by (3.4) and (3.5) is in H, find one which minimizes $g(\varphi)$ given by (3.26). Now we can state the following theorem:

Theorem 4: (i) Minimum Problem III has a unique solution $\varphi^* = (\varphi_i^*, \cdots, \varphi_N^*)$ which is the solution of (3.7)–(3.9).

(ii) Maximum Problem II has a unique solution J^*.

(iii) The solution J^* is given by (3.1) and (3.2) in terms of X^*, which in turn is given by (3.4) and (3.5) in terms of φ^*. Thus J^* is related to φ^* by (3.1), (3.2), (3.4), and (3.5).

(iv) $f(J^*) = g(\varphi^*)$.

(v) J^* and φ^* are the unique stationary points of $f(J)$ and $g(\varphi)$, respectively.

(vi) If $\beta_i(\mathbf{r}) = 0$ on B_{i2}, $i = 1, \cdots, N$, then we can choose $\varphi_i^{(0)} = 0$ and $f(J)$ is minus the rate of entropy production, $f(J) = -\sigma$. If $\rho_i = 0$ and $\alpha_i = 0$, $i = 1, \cdots, N$, then we can choose $J_i^{(0)} = 0$ and $g(\varphi)$ is the rate of entropy production, $g(\varphi) = \sigma$. In both these cases the principle of the minimum rate of entropy production applies.

Part (vii) is not applicable in the present case since neither Σ nor Ω is zero. Part (viii) is applicable, but we shall not write out the resulting equation. The fact that the rate of entropy production is stationary in this case was shown by de Groot and Mazur[10] under slightly more restrictive conditions.

We shall now obtain some different extremum problems governing diffusion, chemical reactions, and heat conduction. To do so we introduce the $2N$ component vector $J = [J_1(\mathbf{r}), \cdots, J_{2N}(\mathbf{r})]$. The first N components of J are 3-component vectors and the last N components are scalars. We interpret $J_1(\mathbf{r})$ as the heat-flux vector at $\mathbf{r}$, $J_i(\mathbf{r})$ as the flux vector of component i for $i = 2, \cdots, N$, $J_{N+1}(\mathbf{r})$ as the total rate of heat production at $\mathbf{r}$ by chemical reactions and $J_{N+i}(\mathbf{r})$ as the total rate of production of component i at $\mathbf{r}$ by chemical reactions $i = 2, \cdots, N$. Similarly, X is a $2N$-component vector, the first N components of which are 3-component vectors and the last N components of which are scalars. We write

$$X = \{X_1(\mathbf{r}), \cdots, X_N(\mathbf{r}), \varphi_1(\mathbf{r}), \cdots, \varphi_N(x)]$$

to indicate that the X_i are the forces introduced previously and the φ_i are the potentials, which are now viewed as forces conjugate to the rates of production J_{N+i}, $i = 1, \cdots, N$.

The linear relation between J and X is still given by (3.1) for the vector components $i = 1, \cdots, N$, and the scalar components are related by

$$J_{N+i}(\mathbf{r}) = \sum_{j=1}^{N} s_{ij}(\mathbf{r})\varphi_j(\mathbf{r}), \quad i = 1, \cdots, N. \quad (3.27)$$

Here the $s_{ij}(\mathbf{r})$ are the elements of a positive-definite invertible symmetric matrix S. They are related to the previously defined quantities ν_{in} and $L_{N+n,m}(\mathbf{r})$ by (3.10). The space H is the completion of the set of $2N$-component vector functions of finite norm, with the inner product defined by

$$(J, X) = \int_D J(\mathbf{r}) \cdot X(\mathbf{r})\, d\mathbf{r} = \int_D \sum_{j=1}^{2N} J_j(\mathbf{r}) \cdot X_j(\mathbf{r})\, d\mathbf{r}. \quad (3.28)$$

The subspace Σ consists of all X in H for which (3.4) holds with $\varphi_i(\mathbf{r}) = 0$ for $\mathbf{r}$ on B_{i2}, $i = 1, \cdots, N$. The subspace Ω consists of all J in H such that $n \cdot J_i(\mathbf{r}) = 0$ for $\mathbf{r}$ on B_{i1} and

$$\nabla \cdot J_i = J_{N+i}, \quad i = 1, \cdots, N.$$

The operator L is a matrix defined by (3.1) and (3.27). Essentially the same proof as that given before shows that Ω and Σ are orthogonal and that H is their direct sum.

We choose for X_0 any vector in H such that $\varphi_i^{(0)}(\mathbf{r}) = \beta_i(\mathbf{r})$ for $\mathbf{r}$ on B_{i2}, $= 1, \cdots, N$. For J_0 we choose any vector in H such that $\mathbf{n} \cdot J_i^{(0)}(\mathbf{r}) = \alpha_i(\mathbf{r})$ for $\mathbf{r}$ on B_{i1} and

$$\nabla \cdot J_i^{(0)} = J_{N+i}^{(0)} + \rho_i, \quad i = 1, \cdots, N. \quad (3.29)$$

Then Problem P of Sec. 2 is that of finding J and X such that $J - J^{(0)} \in \Omega$, $X - X^{(0)} \in \Sigma$ and $J = LX$. The relation $J = LX$ is given by (3.1) and (3.27) and the condition $J - J^{(0)} \in \Omega$ implies that

$$\nabla \cdot J_i = J_{N+i} + \rho_i, \quad i = 1, \cdots, N, \quad (3.30)$$

$$\mathbf{n} \cdot J_i = \alpha_i, \quad \mathbf{r} \text{ on } B_{i1}, \quad i = 1, \cdots, N. \quad (3.31)$$

The condition $X \in \Sigma_0$ implies that

$$X_i = \nabla\varphi_i, \quad i = 1, \cdots, N, \quad (3.32)$$

$$\varphi_i = \beta_i, \quad \mathbf{r} \text{ on } B_{i2}, \quad i = 1, \cdots, N. \quad (3.33)$$

Upon eliminating J and the X_i from these equations, we find that the φ_i again satisfy (3.7), (3.8), and (3.9), so they are the same as the potentials determined by the problem previously formulated. It follows that the J_i and X_i, $i = 1, \cdots, N$, are also the same as the corresponding quantities in the previous problem.

We can now formulate Minimum Problem I and Maximum Problem I of Sec. 2 for the present case. Minimum Problem I becomes identical with Minimum Problem III when $X_1, \cdots, X_N$ are eliminated. Maximum Problem I becomes, when X_0 and J_0 are essentially eliminated, the following problem:

Maximum Problem IV: Among all $J_1(\mathbf{r}), \cdots, J_{2N}(\mathbf{r})$ in H satisfying (3.30) and (3.31), find one which maximizes $f(J)$ given by

$$f(J) = \sum_{i=1}^{N} \left\{ -\int_D \left[J_i \cdot \sum_{j=1}^{N} (L^{-1})_{ij} J_j + J_{N+i} \sum_{j=1}^{N} (S^{-1})_{ij} J_{N+j} + 2\varphi_i^{(0)} \rho_i \right] d\mathbf{r} + 2\int_{B_{i2}} \varphi_i^{(0)} \alpha_i \, dS + 2\int_{B_{i2}} \beta_i \mathbf{n} \cdot J_i \, dS \right\}. \quad (3.34)$$

Theorem 1 applies and relates the solutions of Minimum Problem III and Maximum Problem IV to the solution of the problem formulated above. As in the case of Theorem 4, the principle of the minimum rate of entropy production applies when $\rho_i = 0$ and $\alpha_i = 0$, $i = 1, \cdots, N$, or when $\beta_i = 0$, $i = 1, \cdots, N$. Although this second formulation is more symmetric than the first one, the vector J has more components in it than in the first formulation if $N > k$, i.e., if the number of chemical components exceeds the number of chemical reactions. However, if $N < k$, J in the second formulation has fewer components than in the first formulation. In (3.34), $\varphi_i^{(0)}$ only enters a constant term, i.e., a term independent of J, just as in (3.26).

4. TIME-DEPENDENT DIFFUSION, HEAT CONDUCTION, CHEMICAL REACTION, AND CONVECTION IN A LINEAR SYSTEM

Let us consider the time-dependent processes of diffusion, heat conduction, and chemical reaction in a fixed domain D. We suppose that the potentials $\varphi_1, \cdots, \varphi_N$ satisfy the boundary conditions (3.7) and (3.9) and the differential equations

$$\sum_{k=1}^{N} M_{ik}(\mathbf{r}, t) \frac{\partial \varphi_k}{\partial t} = \nabla \cdot \sum_{j=1}^{N} L_{ij} \nabla \varphi_j - \sum_{j=1}^{N} s_{ij} \varphi_j - \rho_i(\mathbf{r}), \quad i = 1, \cdots, N, \quad (4.1)$$

$$\varphi_i(\mathbf{r}, 0) = \tau_i(\mathbf{r}), \quad i = 1, \cdots, N. \quad (4.2)$$

Here the M_{ik} are the elements of a positive-definite matrix $M(\mathbf{r}, t)$, and the $\tau_i(\mathbf{r})$ are given initial values.

Let $\varphi(\mathbf{r}, t) = [\varphi_1(\mathbf{r}, t), \cdots, \varphi_N(\mathbf{r}, t)]$ be a solution of (4.1), (4.2), (3.7), and (3.9). We assume that L_{ij}, s_{ij}, the boundary values $\alpha_i(\mathbf{r})$ and $\beta_i(\mathbf{r})$, and the source functions $\rho_i(\mathbf{r})$ are all independent of t. Then we denote by $\varphi^*(\mathbf{r})$ the solution of the steady-state problem (3.7)–(3.9). We use the solution φ in (3.26) to evaluate $g[\varphi(\mathbf{r}, t)]$. We shall now prove the following theorem:

Theorem 5: If $\alpha_i = 0$ and if $\varphi_i^{(0)}$ is independent of t, $i = 1, \cdots, N$, then

$$\frac{dg[\varphi(\mathbf{r}, t)]}{dt} < 0 \quad \text{if} \quad \varphi(\mathbf{r}, t) \neq \varphi^*(\mathbf{r})$$

and

$$\frac{dg[\varphi]}{dt} = 0 \quad \text{if} \quad \varphi = \varphi^*.$$

If in addition $\rho_i = 0$, $i = 1, \cdots, N$, then $g = \sigma$, so the rate of entropy production decreases toward its minimum.

To prove the first part we differentiate the expression (3.26) for $g[\varphi]$ and obtain

$$\begin{aligned}\frac{dg[\varphi]}{dt} &= 2\int_D \left\{ \sum_{i,j=1}^{N} \left[\nabla \frac{\partial \varphi_i}{\partial t} \cdot L_{ij} \nabla \varphi_j + \frac{\partial \varphi_i}{\partial t} s_{ij} \varphi_j \right] + \sum_{i=1}^{N} \frac{\partial \varphi_i}{\partial t} \rho_i \right\} d\mathbf{r} \\ &= 2\int_D \left\{ \sum_{i,j=1}^{N} \frac{\partial \varphi_i}{\partial t} [-\nabla \cdot L_{ij} \nabla \varphi_j + s_{ij} \varphi_j] + \sum_{i=1}^{N} \frac{\partial \varphi_i}{\partial t} \rho_i \right\} d\mathbf{r}. \end{aligned} \quad (4.3)$$

The second form of the right-hand side is obtained by applying Gauss's theorem to the first term of the first form on the right. The boundary terms vanish because $\partial \varphi_i / \partial t = \partial \beta_i / \partial t = 0$ on B_{i2} and

$$\mathbf{n} \cdot \sum_{j=0}^{N} L_{ij} \nabla \varphi_j = \alpha_i = 0$$

on B_{i1}. Next we use (4.1) in (4.3) to obtain

$$\frac{dg[\varphi]}{dt} = -2\int_D \sum_{i,k=1}^{N} \frac{\partial \varphi_i}{\partial t} M_{ik} \frac{\partial \varphi_k}{\partial t} d\mathbf{r}. \quad (4.4)$$

The right-hand side of (4.4) is negative if $\partial \varphi_i / \partial t \neq 0$, because M_{ik} is positive definite, and it vanishes only if $\partial \varphi_i / \partial t = 0$, $i = 1, \cdots, N$. From (4.1), this occurs only when the right side of (4.1) vanishes, and that yields (3.8). Thus the right side of (4.4) vanishes if and only if φ satisfies (3.7)–(3.9), which implies that $\varphi = \varphi^*$. This completes the proof of the first part.

When $\rho_i = 0$, $i = 1, \cdots, N$, since we have already assumed that $\alpha_i = 0$, $i = 1, \cdots, N$, we see from (3.26) or from part (vi) of Theorem 4 that $g[\varphi] = \sigma$. Thus in this case the rate of entropy production σ decreases toward its minimum, which it attains in the steady state. This completes the proof of the theorem. The steady state is maintained by the prescribed values $\beta_i(\mathbf{r})$ of φ_i on B_{i1}, $i = 1, \cdots, N$. This result, which is analogous to Theorem 3, shows that, in a time-dependent process, $g[\varphi]$ decreases toward its minimum value, which it attains in the steady state.

We shall conclude this section by including the effects of convection as well as diffusion and production. For simplicity we shall consider a system consisting of a single component, which may be either matter or heat. Let φ denote the concentration of this component if it is matter, or the temperature if the component is heat. Let $k(\mathbf{r})$ denote the diffusivity, $\mathbf{u}(\mathbf{r})$ the convection velocity, and $-s(\mathbf{r})\varphi + \rho(\mathbf{r})$ the rate of production. Then φ satisfies the equation

$$\varphi_t + \nabla \cdot (-k\nabla\varphi + \mathbf{u}\varphi) + s\varphi = \rho, \quad \mathbf{r} \text{ in } D. \tag{4.5}$$

We note that the signs in (4.5) differ from those in (4.1) because φ now represents T rather than T^{-1}, etc. As initial and boundary conditions we take

$$\varphi(\mathbf{r}, 0) = \tau(\mathbf{r}), \tag{4.6}$$

$$\varphi(\mathbf{r}, t) = \beta(\mathbf{r}), \quad \mathbf{r} \text{ on } B_2, \tag{4.7}$$

$$-k\frac{\partial\varphi}{\partial n}(\mathbf{r}, t) = \alpha(\mathbf{r}), \quad \mathbf{r} \text{ on } B_1. \tag{4.8}$$

The steady-state equation obtained by setting $\varphi_t = 0$ in (4.5) can be made self-adjoint by multiplying it by a factor $\psi(\mathbf{r})$ if and only if there exists a $\psi(\mathbf{r})$, such that

$$k^{-1}\mathbf{u} = -\nabla \log \psi. \tag{4.9}$$

When (4.9) holds, $k^{-1}\mathbf{u}$ is irrotational. Only self-adjoint linear equations are derivable from variational principles. When (4.9) holds, we can find variational principles equivalent to the steady-state problem (4.5) with $\varphi_t = 0$, (4.7) and (4.8). To do so we multiply (4.5) with $\varphi_t = 0$ by ψ and obtain

$$\nabla \cdot (-\psi k\nabla\varphi) = -\psi(s + \nabla \cdot \mathbf{u})\varphi + \psi\rho. \tag{4.10}$$

Let us introduce the vector flux J_1 and scalar flux J_2, and the vector and scalar forces X_1 and X_2 defined by

$$J_1 = -\psi k\nabla\varphi, \quad J_2 = -\psi(s + \nabla \cdot \mathbf{u})\varphi, \tag{4.11}$$

$$X_1 = -\nabla\varphi, \quad X_2 = -\varphi. \tag{4.12}$$

Then $J_1 = L_{11}X_1$ and $J_2 = s_{11}X_2$ with $L_{11} = \psi k$ and $s_{11} = \psi(\nabla \cdot \mathbf{u} + s)$. With these definitions the present problem is an instance of the second formulation in the previous section with $N = 1$, $\rho_1 = \psi\rho$, $\alpha_1 = \psi\alpha$, and $\beta_1 = \beta$, provided that $\psi > 0$ and $\nabla \cdot \mathbf{u} + s > 0$. Therefore, the solutions of Minimum Problem III and Maximum Problem IV are related to the solution of the present problem by Theorem 1.

The functional $g(\varphi)$ given by (3.26) now becomes, with φ replaced by $-\varphi$,

$$g(\varphi) = \int_D \psi[k(\nabla\varphi)^2 + (s + \nabla \cdot u)\varphi^2 - 2(\varphi - \varphi^{(0)})\rho]\, d\mathbf{r} + 2\int_{B_1} \psi(\varphi - \varphi^{(0)})\alpha\, ds. \tag{4.13}$$

When $\varphi^{(0)}$ is independent of t, and when $\varphi(\mathbf{r}, t)$ satisfies (4.5), differentiation of (4.13) with respect to t and use of (4.5), (4.7), and (4.8) shows that $dg(\varphi)/dt < 0$ if $\varphi_t \neq 0$.

5. EXTREMUM PRINCIPLES FOR A NONLINEAR CASE OF DIFFUSION, HEAT CONDUCTION, AND CHEMICAL REACTION

So far we have considered only linear systems. We shall now reconsider the processes considered in Secs. 3 and 4 for nonlinear systems. To do so we permit L_{ij} and s_{ij} to be functions of the potentials $\varphi = (\varphi_1, \cdots, \varphi_N)$, but L_{ij} may not depend upon the position vector $\mathbf{r}$. Thus, $L_{ij} = L_{ij}(\varphi)$ and $s_{ij} = s_{ij}(\varphi, \mathbf{r})$. We also require L_{ij} to be a scalar rather than a matrix. Then the φ_i satisfy the nonlinear system of equations, (3.8) and (3.9).

To treat this system, we impose the restrictive conditions

$$\frac{\partial L_{ij}}{\partial \varphi_k} = \frac{\partial L_{ik}}{\partial \varphi_j}, \quad i, j, k = 1, \cdots, N. \tag{5.1}$$

These conditions imply that for each i there exists a scalar function $F_i(\varphi)$ such that

$$L_{ij}(\varphi) = \frac{\partial F_i(\varphi)}{\partial \varphi_j}, \quad i, j = 1, \cdots, N. \tag{5.2}$$

Since the matrix L_{ij} is invertible, the determinant of the L_{ij} is not zero. By (5.2), this is the Jacobian of the transformation from φ to $F = (F_1, \cdots, F_N)$, so this transformation can be inverted to give

$$\varphi = \varphi(F) \quad \text{or} \quad \varphi_i = \varphi_i(F). \tag{5.3}$$

Then (3.7)–(3.9) can be written in the form

$$\Delta F_i - \sum_{j=1}^{N} s_{ij}[\varphi(F), \mathbf{r}]\varphi_j(F) = \rho_i(\mathbf{r}), \quad i = 1, \cdots, N, \tag{5.4}$$

$$\frac{\partial F_i}{\partial n} = \alpha_i(\mathbf{r}), \quad \mathbf{r} \text{ on } B_{i1}, \quad i = 1, \cdots, N, \tag{5.5}$$

$$F_i = F_i[\beta_1(\mathbf{r}), \cdots, \beta_N(\mathbf{r})], \quad \mathbf{r} \text{ on } B_{i2}, \quad i = 1, \cdots, N. \tag{5.6}$$

We can view (5.4)–(5.6) as a nonlinear boundary-value problem for the functions $F_i(\mathbf{r})$. To obtain variational principles equivalent to it, we impose the additional conditions

$$\frac{\partial}{\partial F_k}\sum_{j=1}^{N} s_{ij}[\varphi(F),\mathbf{r}]\varphi_j(F) = \frac{\partial}{\partial F_i}\sum_{j=1}^{N} s_{kj}[\varphi(F),\mathbf{r}]\varphi_j(F),\quad i,k=1,\cdots,N. \tag{5.7}$$

These conditions imply the existence of a function $P(F,\mathbf{r})$ such that

$$\sum_{j=1}^{N} s_{ij}[\varphi(F),\mathbf{r}]\varphi_j(F) = \frac{\partial P(F,\mathbf{r})}{\partial F_i},\quad i=1,\cdots,N. \tag{5.8}$$

To guarantee that the function $P(F,\mathbf{r})$ is convex in F, we require that the matrix of second derivatives of P with respect to F be positive definite. The elements of this matrix are given by either side of (5.7). By using (5.2) in (5.8), we find that these conditions can be expressed in terms of the original quantities s_{ij} and L_{ij} by requiring the following matrix to be symmetric and positive definite:

$$\sum_{t=1}^{N}[L^{-1}(\varphi)]_{tk}\frac{\partial}{\partial\varphi_t}\sum_{j=1}^{N} s_{ij}(\varphi,\mathbf{r})\varphi_j \text{(symm. pos. def.)}. \tag{5.9}$$

When this condition holds, (5.4) can be written

$$\Delta F_i - \frac{\partial P}{\partial F_i} = \rho_i(\mathbf{r}),\quad i=1,\cdots,N. \tag{5.10}$$

We shall now formulate the following problem:

Minimum Problem V: Among all piecewise continuously differentiable functions $F=[F_1(\mathbf{r}),\cdots,F_N(\mathbf{r})]$ satisfying (5.6), find one which minimizes $G(F)$ defined by

$$G(F) = \int_D\left\{\sum_{i=1}^{N}[(\nabla F_i)^2 + 2\rho_i F_i] + 2P(F,\mathbf{r})\right\}d\mathbf{r} - 2\sum_{i=1}^{N}\int_{B_{i1}} F_i\alpha_i\,dS. \tag{5.11}$$

The Euler equations of this problem are just (5.10), and the natural boundary conditions are just (5.5). Furthermore, the minimum problem has a unique solution because the functional $G(F)$ is convex and bounded below. Therefore we have proved the following theorem:

Theorem 6: Minimum Problem V has a unique solution F^* which is the solution of (5.10), (5.5), and (5.6) and is also the unique stationary point of $G[F]$.

The potentials φ are determined uniquely in terms of the solution F^* by (5.3).

The maximum problem reciprocal to Minimum Problem V can be found by using the Friedrichs transformation (Ref. 2, pp. 233–238). To formulate it we consider the class Ω_0 of $2N$ component vectors $J(\mathbf{r})$ satisfying (3.30) and (3.31). Then we introduce the function

$$H(J) = \int_D\left[\sum_{i=1}^{N}\left(-J_i^2 - 2F_i(J_{N+j},\mathbf{r})\frac{\partial P[F(J_{N+j},\mathbf{r}),\mathbf{r}]}{\partial F_i}\right) + 2P[F(J_{N+j},\mathbf{r})]\right]d\mathbf{r} + 2\sum_{i=1}^{N}\int_{B_{i2}}\beta_i J_i\cdot\mathbf{n}\,dS. \tag{5.12}$$

Here, $F_i(J_{N+1},\cdots,J_{2N},\mathbf{r})$ is the solution of the system of equations

$$J_{N+i} = \frac{\partial P(F,\mathbf{r})}{\partial F_i},\quad i=1,\cdots,N. \tag{5.13}$$

That this system is solvable for F follows from the fact that P is a convex function of F, so its matrix of second derivatives with respect to F is positive definite. Therefore, the Jacobian of the right-hand side of (5.13) is not zero, so (5.13) can be solved for F. Now we have the reciprocal problem:

Maximum Problem V: Among all J in Ω_0, i.e., among all J satisfying (3.30) and (3.31), find one which maximizes $H(J)$.

The Euler equations and natural boundary conditions of this problem can be obtained by using the method of Lagrange multipliers to take into account the conditions (3.30) and (3.31). After some simplification, they show that the solution J^* of Maximum Problem V is given by

$$J_i^* = \nabla F_i^*,\quad i=1,\cdots,N, \tag{5.14}$$

$$J_{N+i}^* = \sum_{j=1}^{N} s_{ij}[\varphi(F^*),\mathbf{r}]\varphi_j(F^*),\quad i=1,\cdots,N. \tag{5.15}$$

Here F^* is the solution of (5.4)–(5.6), which is also the solution of Minimum Problem V. The use of this solution in (5.12), or use of the theory of the Friedrichs transformation, shows that $H(J^*) = G(F^*)$. We can summarize these results as follows:

Theorem 7: Maximum Problem V has a unique solution J^* given by (5.14) and (5.15), which is also the unique stationary point of $H(J)$. Furthermore, $H(J^*) = G(F^*)$.

Let us finally consider the time-dependent case in which the potentials $\varphi_i(\mathbf{r},t)$ satisfy (4.1) with $L_{ij}(\varphi)$ a scalar satisfying (5.1) and with $s_{ij} = s_{ij}(\varphi,\mathbf{r})$ such that (5.9) holds. The coefficient matrix $M_{ij}(\varphi,\mathbf{r},t)$

may depend upon φ, $\mathbf{r}$, and $\mathbf{t}$, and we require that the matrix LM^{-1} be positive definite. Then, by setting $\varphi = \varphi(F)$ in (4.1) and using (5.2), we obtain

$$\sum_{j,k=1}^{N} M_{ij}(L^{-1})_{jk}\frac{\partial F_k}{\partial t} = \Delta F_i - \sum_{j=1}^{N} s_{ij}[\varphi(F), \mathbf{r}]\varphi_j(F) - \rho_i(\mathbf{r}), \quad i = 1, \cdots, N. \tag{5.16}$$

The boundary conditions (3.7) and (3.9), which we assume to hold, become (5.5) and (5.6).

We now evaluate $G[F(\mathbf{r}, t)]$, where $F(\mathbf{r}, t)$ satisfies (5.16), (5.5), and (5.6) and G is given by (5.11). Then differentiation of G yields

$$\begin{aligned}\frac{dG(F)}{dt} &= 2\int_D \sum_{i=1}^{N}\left(\nabla F_i\cdot\nabla\frac{\partial F_i}{\partial t} + \rho_i\frac{\partial F_i}{\partial t} + \frac{\partial P}{\partial F_i}\frac{\partial F_i}{\partial t}\right)d\mathbf{r} - 2\sum_{i=1}^{N}\int_{B_{i1}}\alpha_i\frac{\partial F_i}{\partial t}\,dS\\ &= 2\int_D\sum_{i=1}^{N}\left(-\Delta F_i + \rho_i + \frac{\partial P}{\partial F_i}\right)\frac{\partial F_i}{\partial t}\,d\mathbf{r} + 2\sum_{i=1}^{N}\left[\int_{B_{i1}}\left(\frac{\partial F_i}{\partial n} - \alpha_i\right)\frac{\partial F_i}{\partial t}\,dS + \int_{B_{i2}}\frac{\partial F_i}{\partial n}\frac{\partial F_i}{\partial t}\,dS\right].\end{aligned} \tag{5.17}$$

The second form of the right-hand side results from applying Gauss's theorem to the first term in the first integrand. From (5.6) we find that $\partial F_i/\partial t = 0$ on B_{i2}, so the last integral in (5.17) vanishes; from (5.5) we see that the second integral also vanishes. We then calculate $\partial F_i/\partial t$ from (5.16) and use it in (5.17) to obtain

$$\frac{dG(F)}{dt} = -2\int_D\sum_{i,j,k=1}^{N}\left(\Delta F_i - \rho_i - \frac{\partial P}{\partial F_i}\right) \times L_{ij}(M^{-1})_{jk}\left(\Delta F_k - \rho_k - \frac{\partial P}{\partial F_k}\right)d\mathbf{r} \le 0. \tag{5.18}$$

Here we have used (5.8) to simplify the integrand. The inequality follows from the assumed positive definiteness of LM^{-1}. The equality holds only if F satisfies (5.4), in which case $F = F^*$.

It is to be noted that, by setting $F = F(\varphi)$ in G, we obtain a functional $G[F(\varphi)]$ for which $dG/dt \le 0$. Furthermore, Minimum Problem V could be formulated for this functional, in terms of φ. Thus we have proved the following theorem:

Theorem 8: Let $\varphi_i(\mathbf{r}, t)$ satisfy (4.1), (3.7), and (3.9) with $L_{ij}(\varphi)$ a scalar satisfying (5.1). Let $s_{ij}(\varphi, \mathbf{r})$ and $M_{ij}(\varphi, \mathbf{r}, t)$ be such that (5.9) holds and that $L^{-1}M$ is positive definite. Then

$$\frac{dG[F(\varphi)]}{dt} < 0 \quad \text{if} \quad \varphi \ne \varphi^*$$

and

$$\frac{dG[F(\varphi)]}{dt} = 0 \quad \text{if} \quad \varphi = \varphi^*.$$

APPENDIX A: PROOF OF THEOREM 1

To prove Theorem 1 we recall that L is positive definite and bounded and therefore L^{-1} is also positive definite and bounded. The positive definiteness and boundedness of L and L^{-1} show that $g(X)$ is bounded below and that $f(J)$ is bounded above. Therefore, $g(X)$ has a greatest lower bound, and that bound is attained at some point $X = X^* \in \Sigma_0$ because $g(X)$ is continuous, the set Σ_0 in the Hilbert space is complete, and $g \to \infty$ as $|x| \to \infty$. Similarly, $f(J)$ attains its least upper bound at some point $J^* \in \Omega_0$.

In order to show that J^* is unique, we consider any $J \in \Omega_0$. Since $J - J_0 \in \Omega$ and $J^* - J_0 \in \Omega$, it follows that $J^* - J \in \Omega$. Therefore any $J \in \Omega_0$ can be written as $J = J^* + \omega$, where $\omega \in \Omega$. Then

$$\begin{aligned}f(J) &= f(J^* + \omega)\\ &= f(J^*) - 2(\omega, L^{-1}J^* - X_0) - (\omega, L^{-1}\omega).\end{aligned} \tag{A1}$$

In order that $f(J)$ attain its maximum value at J^* or that it be stationary at J^*, the term in (A1) which is linear in ω must vanish. Since ω is any element of Ω, this implies that $L^{-1}J^* - X_0 \in \Sigma$. If the maximum were also attained at J', then $L^{-1}J' - X_0 \in \Sigma$. Therefore $L^{-1}(J^* - J') \in \Sigma$, but we also have $J^* - J' \in \Omega$. In view of the orthogonality of Ω and Σ, it follows that $(J^* - J', L^{-1}[J^* - J']) = 0$. Since L^{-1} is positive definite, this shows that $J^* - J' = 0$, so J^* is unique. In the same way it follows for the minimum problem that $LX^* - J_0 \in \Omega$ and $LX' - J_0 \in \Omega$ if X' is also a minimum point. But then $L(X^* - X') \in \Omega$; and since $X^* - X' \in \Sigma$, it follows that

$$(X^* - X', L[X^* - X']) = 0.$$

Since L is positive definite, $X^* - X' = 0$, so X^* is unique. Thus we have proved parts (i), (ii), and (v) of Theorem 1.

To prove part (iii) we observe that in proving the uniqueness of J^* we showed that $L^{-1}J^* - X_0 \in \Sigma$, while admissibility of X^* requires $X^* - X_0 \in \Sigma$. Therefore $L^{-1}J^* - X^* \in \Sigma$. In proving the uniqueness of X^* we found that $LX^* - J_0 \in \Omega$, while admissibility of J^* requires $J^* - J_0 \in \Omega$. Therefore $J^* - LX^* \in \Omega$.

Since Ω and Σ are orthogonal, these two results yield $(J^* - LX^*, L^{-1}[J^* - LX^*]) = 0$. Because L^{-1} is positive definite, it follows that $J^* = LX^*$, which proves part (iii) of the theorem.

We now use the orthogonality of $J^* - J_0$, which is in Ω, and $X^* - X_0 = L^{-1}J^* - X_0$, which is in Σ, to obtain

$$0 = 2(J^* - J_0, L^{-1}J^* - X_0)$$
$$= 2(J^*, L^{-1}J^*) - 2(J^*, X_0) - 2(J_0, L^{-1}J^*) + 2(J_0, X_0). \quad \text{(A2)}$$

Next we add (A2) to $f(J^*)$, which given by (2.4) with $J = J^*$, to obtain

$$f(J^*) = (J^*, L^{-1}J^*) - 2(J_0, L^{-1}J^*) + 2(J_0, X_0). \quad \text{(A3)}$$

Upon setting $J^* = LX^*$, we see that the right-hand side of (A3) is just $g(X^*)$ given by (2.3) with $X = X^*$. Thus $f(J^*) = g(X^*)$, which proves part (iv) of the theorem.

When $X_0 \in \Sigma$, we can evaluate (J, X_0) for $J \in \Omega_0$ by writing $J = J_0 + \omega$, where $\omega \in \Omega$. Then we have, upon using the orthogonality of ω and X_0, the result that

$$(J, X_0) = (J_0 + \omega, X_0) = (J_0, X_0). \quad \text{(A4)}$$

By using (A4) in (2.4), we obtain $f(J) = -\sigma(J) + 2(J_0, X_0)$, which proves the first statement in part (vi). Similarly, when $J_0 \in \Omega$ and $X \in \Sigma_0$, we have $X - X_0 \in \Sigma$. Then, from the orthogonality of J_0 and $X - X_0$, we obtain

$$g(X) = \sigma(X) - 2(J_0, X - X_0) = \sigma(X). \quad \text{(A5)}$$

This proves the second statement in part (vi). To prove the "only if" statement we must show that (J, X_0) is constant for $J \in \Omega_0$ only if $X_0 \in \Sigma$, and that $(J_0, X - X_0)$ is constant for $X \in \Sigma_0$ only if $J_0 \in \Omega$. These conclusions follow at once from the fact that Σ and Ω are orthogonal complements of each other, which completes the proof of part (vi).

When $\Sigma = 0$, then the only X in Σ_0 is X_0, so $X^* = X_0$. From (2.4) the second term in $g(X^*)$ vanishes, and $g(X^*) = \sigma(X^*)$. When $\Omega = 0$, then $J^* = J_0$ and (2.4) yields $f(J^*) = -\sigma(J^*) + 2(J_0, X_0)$. Combining the results with part (iv) yields part (vii).

To prove part (viii) we note that if $J \in \Omega_0$ and $X \in \Sigma_0$, then $J^* - J \in \Omega$ and $L^{-1}J^* - X \in \Sigma$. Therefore, $J^* - J$ is orthogonal to $L^{-1}J^* - X$, i.e.,

$$(J^* - J, L^{-1}[J^* - LX]) = 0. \quad \text{(A6)}$$

Now we use (A6) to obtain

$$\left(J^* - \frac{J + LX}{2}, L^{-1}\left[J^* - \frac{J + LX}{2}\right]\right)$$
$$= \tfrac{1}{4}(J^* - J + J^* - LX, L^{-1}[J^* - J + J^* - LX])$$
$$= \tfrac{1}{4}(J^* - J, L^{-1}[J^* - J]) + \tfrac{1}{4}(J^* - LX, L^{-1}[J^* - LX])$$
$$= \tfrac{1}{4}(J - J^*, L^{-1}[J - J^*]) + \tfrac{1}{4}(J^* - LX, L^{-1}[J^* - LX])$$
$$= \tfrac{1}{4}(J - LX, L^{-1}[J - LX]). \quad \text{(A7)}$$

This proves part (viii) and completes the proof of the theorem.

APPENDIX B: RELATION BETWEEN COURANT-HILBERT'S RECIPROCAL QUADRATIC VARIATIONAL PROBLEMS AND FENCHEL'S DUALITY THEOREM INVOLVING CONJUGATE FUNCTIONS

In 1953 Fenchel[8] proved a general duality theorem which states that the minimum value of a function in a certain minimum problem is equal to the maximum value of a different function in a related maximum problem. A particularly clear presentation of this theorem and its proof is given by Karlin.[11] By examining that proof, we observe that the theorem is true for functions of vectors X in a Hilbert space, and not merely for functions of vectors in a finite-dimensional Euclidean space. We shall now relate to this theorem the reciprocal quadratic variational problems considered by Courant–Hilbert[2] and employed in Sec. 2.

To state Fenchel's theorem we consider a convex function $\varphi(X)$ defined for X in a convex set C in a real Hilbert space H, and a concave function $\psi(X)$ defined for X in a convex set D in H. The conjugate functions of φ and ψ are denoted by $\varphi^*(\xi)$ and $\psi^*(\xi)$, respectively, and their domains of definition are C^* and D^*, respectively. These quantities are defined as follows:

$$C^* = \left\{\xi \,\middle|\, \sup_{X \in C} [(\xi, X) - \varphi(X)] < \infty\right\}, \quad \text{(B1)}$$

$$\varphi^*(\xi) = \sup_{X \in C} [(\xi, X) - \varphi(X)], \quad \text{(B2)}$$

$$D^* = \left\{\xi \,\middle|\, \inf_{X \in D} [(\xi, X) - \psi(X)] < \infty\right\}, \quad \text{(B3)}$$

$$\psi^*(\xi) = \inf_{X \in D} [(\xi, X) - \psi(X)]. \quad \text{(B4)}$$

Now we consider the following two extremum problems, assuming that $C \cap D$ is not empty:

Maximum Problem: Among all $X \in C \cap D$, find one which maximizes $\psi(X) - \varphi(X)$.

Minimum Problem: Among all $J \in C^* \cap D^*$, find one which minimizes $\varphi^*(\xi) - \psi^*(\xi)$.

Then we have the following theorem:

Theorem 9: Each of these problems has a unique solution and

$$\min_{X\in C\cap D}[\varphi(X) - \psi(X)] = \max_{\xi\in C^*\cap D^*}[\psi^*(\xi) - \varphi^*(\xi)]. \tag{B5}$$

Let us now define φ, ψ, C, and D as follows:

$$\varphi(X) = (X, LX) - 2(X, J_0) + 2(J_0, X_0), \quad C = H, \tag{B6}$$

$$\psi(X) = 0, \quad D = \Sigma_0. \tag{B7}$$

Here L is an invertible symmetric, positive-definite operator, J_0 and X_0 are given vectors in H, and $\Sigma_0 = \Sigma + X_0$, where Σ is a linear subspace of H. By using (B6) in (B1) and (B2), we find that

$$\varphi^*(\xi) = \tfrac{1}{4}(\xi, L^{-1}\xi) + (\xi, L^{-1}J_0) + (J_0, L^{-1}J_0) - 2(J_0, X_0), \quad C^* = H. \tag{B8}$$

Then by using (B7) in (B1) and (B2) we find that

$$\psi^*(\xi) = (\xi, X_0), \quad D^* = \Omega. \tag{B9}$$

Here Ω is the orthogonal complement of Σ in H. Now the extremum problems formulated above become the following:

Minimum Problem: Among all $X \in \Sigma_0$, find one which minimizes $(X, LX) - 2(X, J_0) + 2(X_0, J_0)$.

Maximum Problem: Among all $\xi \in \Omega$, find one which maximizes

$$-\tfrac{1}{4}(\xi, L^{-1}\xi) - (\xi, L^{-1}J_0) - (J_0, L^{-1}J_0) + 2(J_0, X_0).$$

If we define $J = \xi/2 + J_0$ and $\Omega_0 = \Omega + J_0$, then the maximum problem becomes the following problem:

Maximum Problem: Among all $J \in \Omega_0$, find one which maximizes $-(J, L^{-1}J) + 2(J, X_0)$.

Comparison shows that the minimum problem above and the last form of the maximum problem are identical, respectively, with Minimum Problem I and Maximum Problem I of Sec. 2. Then the theorem above yields parts (i), (ii), and (iv) of Theorem 1 of Sec. 2, i.e., the existence of unique solutions to the two problems and the equality of the maximum and minimum values. However, it does not yield part (iii), which is the relation $J^* = LX^*$ between the solutions X^* and J^* of the two problems.

* This work was supported in part by the National Science Foundation.

[1] I. Prigogine, Bull. Acad. Belgium Class Sci. **31**, 600 (1945).
[2] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.
[3] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes* (Interscience, New York, 1967), 3rd ed.
[4] J. B. Keller, L. Rubenfeld and J. Molyneux, J. Fluid Mech. **30**, 97–125 (1967).
[5] H. Helmholtz, Verh. Naturhist.-Med. Vereins Wiss. Abh. **1**, 223 (1868).
[6] D. J. Korteweg, Phil. Mag. **16**, 1 (1883).
[7] *Nonequilibrium Thermodynamics Variational Techniques and Stability*, edited by R. J. Donally, R. Herman, and I. Prigogine (University of Chicago Press, Chicago, 1966).
[8] W. Fenchel, "Convex cones, Sets and Functions," Princeton University, Princeton (1953), lecture notes.
[9] H. Brezis and J. B. Keller (to be published).
[10] S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[11] S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics* (Addison-Wesley, Reading, 1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Suggestive Approximate Dynamical Symmetry of an Electron Moving in the Field of Many Stationary Nuclei*

SUPRIYA RAY
Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439

(Received 6 February 1970)

A group-theoretical study has been employed in an effort to uncover the suggestive operator, if it exists, of the form $L^2 + \hat{O}$ which commutes with the Hamiltonian of an electron moving in the many-nucleus Coulomb field. The Casimir operator L^2 may easily be identified as the square angular-momentum operator, while the operator $\hat{O}$ may involve the internuclear separations and operators of the Lie algebra. Finally, however, the analysis has revealed the fact that such operator or dynamical invariant does not, in general, exist for any arbitrary stationary nuclei except in few special cases which, however, have their special geometrical symmetry.

I. INTRODUCTION

In recent years considerable interest has been drawn to uncover the symmetry properties in molecules by using the group-theoretical technique.[1] But Wulfman and Takahata[2a] have formulated the problem of one electron moving in the field of many stationary nuclei in terms of the operations of the Lie algebras of E_4, R_4, R_5, and $O_{4,1}$, all noninvariance groups of quantal electrostatics. The authors strongly suggest the existence of an invariant for the Hamiltonian. In the present work, an investigation has been carried out to single out the invariant, if it exists. An exclusive

Then we have the following theorem:

Theorem 9: Each of these problems has a unique solution and

$$\min_{X\in C\cap D}[\varphi(X)-\psi(X)] = \max_{\xi\in C^*\cap D^*}[\psi^*(\xi)-\varphi^*(\xi)]. \tag{B5}$$

Let us now define φ, ψ, C, and D as follows:

$$\varphi(X) = (X, LX) - 2(X, J_0) + 2(J_0, X_0), \quad C = H, \tag{B6}$$

$$\psi(X) = 0, \quad D = \Sigma_0. \tag{B7}$$

Here L is an invertible symmetric, positive-definite operator, J_0 and X_0 are given vectors in H, and $\Sigma_0 = \Sigma + X_0$, where Σ is a linear subspace of H. By using (B6) in (B1) and (B2), we find that

$$\varphi^*(\xi) = \tfrac{1}{4}(\xi, L^{-1}\xi) + (\xi, L^{-1}J_0) + (J_0, L^{-1}J_0) - 2(J_0, X_0), \quad C^* = H. \tag{B8}$$

Then by using (B7) in (B1) and (B2) we find that

$$\psi^*(\xi) = (\xi, X_0), \quad D^* = \Omega. \tag{B9}$$

Here Ω is the orthogonal complement of Σ in H. Now the extremum problems formulated above become the following:

Minimum Problem: Among all $X \in \Sigma_0$, find one which minimizes $(X, LX) - 2(X, J_0) + 2(X_0, J_0)$.

Maximum Problem: Among all $\xi \in \Omega$, find one which maximizes

$$-\tfrac{1}{4}(\xi, L^{-1}\xi) - (\xi, L^{-1}J_0) - (J_0, L^{-1}J_0) + 2(J_0, X_0).$$

If we define $J = \xi/2 + J_0$ and $\Omega_0 = \Omega + J_0$, then the maximum problem becomes the following problem:

Maximum Problem: Among all $J \in \Omega_0$, find one which maximizes $-(J, L^{-1}J) + 2(J, X_0)$.

Comparison shows that the minimum problem above and the last form of the maximum problem are identical, respectively, with Minimum Problem I and Maximum Problem I of Sec. 2. Then the theorem above yields parts (i), (ii), and (iv) of Theorem 1 of Sec. 2, i.e., the existence of unique solutions to the two problems and the equality of the maximum and minimum values. However, it does not yield part (iii), which is the relation $J^* = LX^*$ between the solutions X^* and J^* of the two problems.

* This work was supported in part by the National Science Foundation.

[1] I. Prigogine, Bull. Acad. Belgium Class Sci. **31**, 600 (1945).
[2] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.
[3] I. Prigogine, *Introduction to Thermodynamics of Irreversible Processes* (Interscience, New York, 1967), 3rd ed.
[4] J. B. Keller, L. Rubenfeld and J. Molyneux, J. Fluid Mech. **30**, 97–125 (1967).
[5] H. Helmholtz, Verh. Naturhist.-Med. Vereins Wiss. Abh. **1**, 223 (1868).
[6] D. J. Korteweg, Phil. Mag. **16**, 1 (1883).
[7] *Nonequilibrium Thermodynamics Variational Techniques and Stability*, edited by R. J. Donally, R. Herman, and I. Prigogine (University of Chicago Press, Chicago, 1966).
[8] W. Fenchel, "Convex cones, Sets and Functions," Princeton University, Princeton (1953), lecture notes.
[9] H. Brezis and J. B. Keller (to be published).
[10] S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics* (North-Holland, Amsterdam, 1962).
[11] S. Karlin, *Mathematical Methods and Theory in Games, Programming and Economics* (Addison-Wesley, Reading, 1959).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 11, NUMBER 9 SEPTEMBER 1970

Suggestive Approximate Dynamical Symmetry of an Electron Moving in the Field of Many Stationary Nuclei*

SUPRIYA RAY
Chemistry Division, Argonne National Laboratory, Argonne, Illinois 60439

(Received 6 February 1970)

A group-theoretical study has been employed in an effort to uncover the suggestive operator, if it exists, of the form $L^2 + \hat{O}$ which commutes with the Hamiltonian of an electron moving in the many-nucleus Coulomb field. The Casimir operator L^2 may easily be identified as the square angular-momentum operator, while the operator $\hat{O}$ may involve the internuclear separations and operators of the Lie algebra. Finally, however, the analysis has revealed the fact that such operator or dynamical invariant does not, in general, exist for any arbitrary stationary nuclei except in few special cases which, however, have their special geometrical symmetry.

I. INTRODUCTION

In recent years considerable interest has been drawn to uncover the symmetry properties in molecules by using the group-theoretical technique.[1] But Wulfman and Takahata[2a] have formulated the problem of one electron moving in the field of many stationary nuclei in terms of the operations of the Lie algebras of E_4, R_4, R_5, and $O_{4,1}$, all noninvariance groups of quantal electrostatics. The authors strongly suggest the existence of an invariant for the Hamiltonian. In the present work, an investigation has been carried out to single out the invariant, if it exists. An exclusive use of the Lie algebra and commutator relations has been made. In Sec. II, a brief review of the field is given so that a systematic development of the current work becomes plausible. Section III, however, is entirely devoted to give the detailed algebra of the present analysis. Finally, a brief discussion is added to the concluding Sec. IV.

II. BRIEF REVIEW OF THE FIELD

In the study of the dynamical group of the many-nucleus 1-electron problem, it is found possible to transform[2b] the molecular Schrödinger equation into an equation involving ten operators which together form a Lie ring and generate a Lie algebra and its corresponding 10-parameter-continuous group.

The equation for the motion of an electron in the field of several stationary nuclei is given by the eigenvalue equation[2]

$$P_{0,0p}\Psi(\mathbf{p}) = p_0\Psi(\mathbf{p}), \tag{1}$$

where $p_0 = \sqrt{-E}$ is the root-mean-square (rms) momentum of the electron, and where $\Psi(\mathbf{p})$, which is the momentum space eigenfunction of the electron, depends parametrically on the nuclear co-ordinates R_j. In atomic units, the rms momentum operator $P_{0,0p}$ is given by

$$P_{0,0p} = \sum_j \zeta_j[\exp(-i\mathbf{p}\cdot\mathbf{R}_j)]\Pi[\exp(+i\mathbf{p}\cdot\mathbf{R}_j)] \tag{2}$$

and the Hamiltonian is given by

$$\mathcal{H} = -\tfrac{1}{2}P_{0,0p}^2. \tag{3}$$

In Eq. (2), ζ_j is the charge of nucleus j and Π is the hydrogen atom rms momentum operator, whose eigenvalues are $1/n$, if n is the principal quantum number.

Also, it is now well known that[3] if J_{ab} is the Hermitian rotation operator in the (a, b) plane of Fock's 4-space,[4] then

$$\Pi = \left(\sum_{a<b} J_{ab}^2 + 1\right)^{-\frac{1}{2}} \tag{4}$$

and

$$J_{ab} = -i\left(x_a\frac{\partial}{\partial x_b} - x_b\frac{\partial}{\partial x_a}\right) = -J_{ba}. \tag{5}$$

From the defining relations of Fock's stereographic projection onto hypersphere $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$, it follows that

$$\mathbf{p} = p_0[(x_1, x_2, x_3)/(1 + x_4)] \tag{6}$$

if x_4 is in the direction "perpendicular" to the 3-space of $\mathbf{p}$. Thus, in Fock's 4-space, the only operators required for the expression of $P_{0,0p}$ are the six rotation operators J_{ab} and the four translation operator x_c, which together generate the 4-dimensional Euclidean group E_4. But all Hermitian representations of E_4 are infinite dimensional. It is, therefore, not a very desirable group. However, E_4 may be obtained via a Wigner–Inönü contraction[5] from the groups R_5 (5-dimensional rotation group) and $O_{4,1}$ (de Sitter group) which have been shown to be appropriate to the Kepler problem. All of the Hermitian representations of R_5 are of finite dimension. Therefore, R_5 is the most desirable group; it is the group of linear substitutions of positive determinant that leave invariant the quadratic form

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2.$$

In the corresponding 5-space, there are ten 2-planes $a{:}b$, and in each of these planes the Hermitian operator of an infinitesimal rotation is J_{ab} given in Eq. (6). The J_{ab}'s satisfy the commutation relations

$$[J_{ab}, J_{cd}] = 0, \quad \text{if} \quad a \neq b \neq c = d, \tag{7}$$

and

$$[J_{ab}, J_{ac}] = iJ_{bc}, \quad \text{if} \quad b \neq c. \tag{8}$$

It has become customary to define the following operators:

$$\mathbf{L} = (L_x, L_y, L_z) = (J_{23}, J_{31}, J_{12}), \tag{9a}$$

$$\mathbf{A} = (A_x, A_y, A_z) = (J_{14}, J_{24}, J_{34}), \tag{9b}$$

$$\mathbf{B} = (B_x, B_y, B_z) = (J_{15}, J_{25}, J_{35}), \tag{9c}$$

$$S = J_{45}. \tag{9d}$$

$\mathbf{L}$ may be interpreted as an angular momentum operator, and within the R_3 subgroup of R_5, $\mathbf{L}$, $\mathbf{A}$, and $\mathbf{B}$ transform as vectors, while S transforms as a scalar. The Lie group R_5 has a subgroup R_4, and considerable simplifications are introduced if we make use of the fact that R_4 is locally isomeric with the Kronecker product of R_3 itself.[6]

In Sec. III, we will give complete treatment to show how the present analysis is developed.

III. PRESENT ANALYSIS

For an electron moving in a spherically symmetric field, L^2 is the invariant for the Hamiltonian. But, in a many-nucleus Coulomb field whose spherical symmetry is generally lost, the operator L^2 is obviously suspected not to be invariant for $P_{0,0p}$. On the other hand, it is agreed that the commutator

$$[P_{0,0p}, L^2] \neq 0 \tag{10}$$

must have some definite value. Here the attention has therefore been diverted to evaluate its value. In doing so one needs to work out some fundamental or key commutators like $[L^2, p_1]$, $[L^2, p_2]$, $[L^2, p_3]$, etc., because one can easily express the commutator sought in Eq. (10) in terms of those commutators. In laying out the ground work, we here sort out the following relations:

$$[J_{ab}, p_c] = 0, \quad \text{if} \quad c \neq a \text{ or } b, \tag{11a}$$

$$[J_{ab}, p_a] = ip_b, \quad \text{if} \quad a, b = 1, 2, 3, \tag{11b}$$

$$[J_{a4}, p_a] = i(p_a^2 + p_4), \tag{11c}$$

$$[J_{a4}, p_4] = i(p_a p_4 - p_a), \tag{11d}$$

$$[J_{a4}, p_b] = ip_a p_b, \quad \text{if} \quad b \neq 4 \text{ or } a, \tag{11e}$$

$$p_1^2 + p_2^2 + p_3^2 + 2p_4 = 1, \tag{11f}$$

and

$$[L^2, \Pi] = 0, \tag{11g}$$

where

$$L^2 = J_{23}^2 + J_{31}^2 + J_{12}^2, \tag{12}$$

$$A^2 = J_{14}^2 + J_{24}^2 + J_{34}^2, \tag{13}$$

and

$$\Pi = [1 + (L^2 + A^2)]^{-\frac{1}{2}}. \tag{14}$$

Next, by using the relations given in Eqs. (11a)–(12), one can, however, work out that

$$[L^2, p_a] = 2p_a + 2i(p_b J_{ab} - p_c J_{ca}), \tag{15}$$

where (a, b, c) are (1, 2, 3), (2, 3, 1), and (3, 1, 2); p_1, p_2, and p_3 are three components of $\mathbf{p}$ [given in Eq. (6)]. Multiplying p_a by R_{ja}, ath component of $\mathbf{R}_j(R_{j1}, R_{j2}, R_{j3})$, we replace Eq. (15) by

$$[L^2, (p_a R_{ja})] = 2(p_a R_{ja}) + 2i(p_b R_{ja} J_{ab} - p_c R_{ja} J_{ca}). \tag{15'}$$

Then we combine $[L^2, (p_1 R_{j1})]$, $[L^2, (p_2 R_{j2})]$, and $[L^2, (p_3 R_{j3})]$ given in Eq. (15′) to obtain the commutator

$$[L^2, (\mathbf{p} \cdot \mathbf{R}_j)] = 2(\mathbf{p} \cdot \mathbf{R}_j) - 2i[(\mathbf{p} \times \mathbf{R}_j) \cdot \mathbf{L}], \tag{16}$$

which, of course, depends upon the orientation of $\mathbf{R}_j$ with respect to $\mathbf{p}$. We substitute

$$\boldsymbol{\beta}_j = i(\mathbf{p} \cdot \mathbf{R}_j) \tag{17a}$$

and

$$\boldsymbol{\lambda}_j = i(\mathbf{p} \times \mathbf{R}_j), \tag{17b}$$

and then Eq. (16) and the commutator given in Eq. (10) turn out to be of the following form, respectively:

$$[L^2, \beta_j] = 2\beta_j - 2i(\boldsymbol{\lambda}_j \cdot \mathbf{L}) \tag{18}$$

and

$$[(e^{-\beta_j} \Pi e^{\beta_j}), L^2] \neq 0. \tag{10'}$$

Case 1: When $\mathbf{R}_j$ is either parallel or antiparallel to $\mathbf{p}$, the term involving $\boldsymbol{\lambda}_j$ in Eq. (18) is zero. Therefore we get

$$[L^2, \beta_j] = 2\beta_j. \tag{18'}$$

Extending further the use of the commutator given in Eq. (18′), one can also show that

$$[L^2, e^{\pm\beta_j}] = \pm 2\beta_j e^{\pm\beta_j}. \tag{19}$$

Once the key commutators are evaluated, we are then ready to work out the value of the commutator sought in Eq. (10′). After carrying out the usual simplification in combining with Eqs. (11g), (19), etc., we obtain the value of the commutator

$$[(e^{-\beta_j} \Pi e^{\beta_j}), L^2] = -[(e^{-\beta_j} \Pi e^{\beta_j}), (2\beta_j)], \tag{20}$$

which leads to the following commutation relations:

$$[(e^{-\beta_j} \Pi e^{\beta_j}), (L^2 + 2\beta_j)] = 0 \tag{21}$$

or

$$[(e^{-i\mathbf{p}\cdot\mathbf{R}_j} \Pi e^{+i\mathbf{p}\cdot\mathbf{R}_j}), (L^2 + 2(i\mathbf{p} \cdot \mathbf{R}_j))] = 0. \tag{22}$$

Thus, the invariant operator $\hat{Q}$ is expressed here by

$$\hat{Q} = L^2 + 2(i\mathbf{p} \cdot \mathbf{R}_j), \tag{23}$$

where

$$\hat{O}_j = 2(i\mathbf{p} \cdot \mathbf{R}_j). \tag{23'}$$

Case 2: When $\mathbf{R}_j$ is neither parallel nor antiparallel to $\mathbf{p}$, the term involving $\boldsymbol{\lambda}_j$ in Eq. (18) is nonvanishing. However, its value ought to be sufficiently small. Because the expression for the invariant $\hat{Q}$ shows that

$$\hat{Q} = L^2 + 2\beta_j, \quad \text{for } \mathbf{R}_j \text{ is either } \uparrow\uparrow \text{ or } \downarrow\uparrow \ \mathbf{p},$$

and

$$\hat{Q} = L^2 \quad \text{when } \mathbf{R}_j \text{ is } \perp \mathbf{p},$$

obviously $\hat{Q}$ fluctuates between two extreme limits of $(L^2 + 2\beta_j)$ and L^2 for any other orientation of $\mathbf{R}_j$ with respect to $\mathbf{p}$. Since

$$\lambda_j^2 = (pR_j)^2 - (\mathbf{p} \cdot \mathbf{R}_j)^2, \tag{24}$$

the fluctuating term involving $\boldsymbol{\lambda}_j$ is quite small, and hence it is assumed reasonably kosher to throw away any term involving λ_j^2 or higher powers of $\boldsymbol{\lambda}_j$. Keeping this assumption in mind, we move ahead to evaluate a general expression for $\hat{Q}$.

Therefore, the expression in Eq. (18) is used to evaluate requisite commutators such as

$$[L^2, e^{\pm\beta_j}] = \pm 2\beta_j e^{\pm\beta_j} \mp ie^{\pm\beta_j}(\boldsymbol{\lambda}_j \cdot \mathbf{L}) - e^{\pm\beta_j}\lambda_j^2 \tag{25a}$$

in which the last term, quadratic in $\boldsymbol{\lambda}_j$, is to be neglected. Thus, instead, the expression that we will use is

$$[L^2, e^{\pm\beta_j}] = \pm 2\beta_j e^{\pm\beta_j} \mp ie^{\pm\beta_j}(\boldsymbol{\lambda}_j \cdot \mathbf{L}), \tag{25b}$$

and similarly we show that

$$[e^{\pm\beta_j}, (\boldsymbol{\lambda}_j \cdot \mathbf{L})] = \pm i\lambda_j^2 e^{\pm\beta_j} \cong 0. \tag{26}$$

Now, using the commutators given in Eqs. (25b), (26), and (11g), we again work out the said commutator:

$$[(e^{-\beta_j}\Pi e^{\beta_j}), L^2] = -[(e^{-\beta_j}\Pi e^{\beta_j}), 2\beta_j] + 2i[(e^{-\beta_j}\Pi e^{\beta_j}), (\boldsymbol{\lambda}_j \cdot \mathbf{L})]. \quad (27)$$

This immediately leads to

$$[(e^{-\beta_j}\Pi e^{\beta_j}), \{L^2 + 2\beta_j - 2i(\boldsymbol{\lambda}_j \cdot \mathbf{L})\}] = 0 \quad (28)$$

or

$$[(e^{-i\mathbf{p}\cdot\mathbf{R}_j}\Pi e^{+i\mathbf{p}\cdot\mathbf{R}_j}), \{L^2 + 2i(\mathbf{p}\cdot\mathbf{R}_j) + 2[(\mathbf{p}\cdot\mathbf{R}_j)\cdot\mathbf{L}]\}] = 0. \quad (29)$$

Thus, the general form of the operator $\hat{Q}$ is given by

$$\hat{Q} = L^2 + 2(i\mathbf{p}\cdot\mathbf{R}_j) + 2[(\mathbf{p}\times\mathbf{R}_j)\cdot\mathbf{L}], \quad (30)$$

which satisfies all cases relating to the orientation of $\mathbf{R}_j$ with respect to $\mathbf{p}$. Finally, we replace $(e^{\beta_j}\Pi e^{\beta_j})$ by $(\zeta_j e^{-\beta_j}\Pi e^{\beta_j})$ in Eq. (28) and take the summation over j to obtain

$$[P_{0,0p}, L^2] + 2\sum_j [(\zeta_j e^{-\beta_j}\Pi e^{\beta_j}), \{\beta_j - i(\boldsymbol{\lambda}_j \cdot \mathbf{L})\}] = 0 \quad (31)$$

or

$$[P_{0,0p}, \{L^2 + 2\beta_k - 2i(\boldsymbol{\lambda}_k \cdot \mathbf{L})\}] + 2\sum_j [(\zeta_j e^{-\beta_j}\Pi e^{\beta_j}), \{(\beta_j - \beta_k) - i(\boldsymbol{\lambda}_j - \boldsymbol{\lambda}_k)\cdot\mathbf{L}\}] = 0, \quad (32)$$

from which one obviously notes that the second commutator of Eq. (31) does not in general reduce to the form

$$[P_{0,0p}, \hat{O}] \quad \text{and} \quad \hat{O} = f(\mathbf{R}_j, \mathbf{p}, \mathbf{L}, \text{etc.}),$$

so that one will have for $P_{0,0p}$ the desired invariant $L^2 + \hat{O}$. On the other hand, one can very easily show from Eq. (32) that, in few special cases where $\beta_k = \beta_j$ and $\boldsymbol{\lambda}_k$ and $\boldsymbol{\lambda}_j$ are perpendicular to $\mathbf{L}$ for all values of k and j,

$$[P_{0,0p}, \{L^2 + 2\beta_k - 2i(\boldsymbol{\lambda}_k \cdot \mathbf{L})\}] = 0, \quad (33)$$

that is, one finds the desired invariant $L^2 + \hat{O}$ for $P_{0,0p}$. Thus, it appears quite obvious that such a desired invariant $L^2 + \hat{O}$ does really not exist for general cases of arbitrary stationary nuclei, although one can find them (invariants) in exceptional cases which have special geometrical symmetry, satisfying all the conditions to have $L^2 + \hat{O}$.

IV. DISCUSSION

Our analysis in the previous section makes it quite clear that dynamical symmetry is not independent of geometrical symmetry. Equations (29)–(33) confirm one fact: there exists a dynamical invariant $L^2 + \hat{O}$ for two stationary nuclei. But a system of two stationary nuclei has well-known geometrical symmetry. For a system of more than two nuclei, one may have an invariant $L^2 + \hat{O}$, if all the nuclei lie in one plane, if one nucleus lies at the center and the rest of the nuclei lie on the circumference of a circle, and if $\mathbf{L}$ is perpendicular to the plane formed by $\mathbf{p}$ and $\mathbf{R}_j$. That means that the system should have well-specified geometrical symmetry. Thus, in conclusion, one may say that, for a system of many stationary nuclei which has no built-in geometrical symmetry, mother nature forbids it to have any dynamical symmetry.

Finally, we should also have a limit to the number of nuclei in order to maintain some sort of geometrical symmetry. As the number increases,so the symmetry is destroyed.

ACKNOWLEDGMENTS

The author would like to extend his cordial thanks to Professor C. E. Wulfman, Professor of Physics, University of the Pacific, Stockton, Calif., for suggesting this problem and for his many comments. This work was first initiated while the author was a research associate at University of the Pacific (1967–68) under an NSF research project.

* Based on work performed under the auspices of the U.S. Atomic Energy Commission.

[1] C. E. Wulfman and Y. Takahata, J. Chem. Phys. **47**, 488 (1967); T. D. Bouman, A. L. H. Chung, and G. L. Goodman, in *Sigma Molecular Orbital Theory* (Yale Press, New Haven, London, 1970); J. S. Alper and O. Sinanoğlu, Phys. Rev. **177**, 77 (1969).
[2] (a) C. E. Wulfman and Y. Takahata, J. Chem. Phys. **47**, 488 (1967); (b) T. Shibuya and C. E. Wulfman, Proc. Roy. Soc. (London) **A286**, 376 (1965).
[3] H. Bacry, Nuovo Cimento **41**, 223 (1966); A. Böhm, Nuovo Cimento **43**, 665 (1966); M. Bander and C. Itykson, Rev. Mod. Phys. **38**, 330, 346 (1966); C. C. Sudarshan, N. Mukunda, and L. O'Raifeartaigh, Phys. Rev. Letters **15**, 1041 (1965).
[4] V. Fock, Z. Physik **98**, 145 (1935); V. Bargmann, Z. Physik **99**, 576 (1936).
[5] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. U.S. **39**, 510 (1953); **40**, 119 (1954).
[6] L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).
[7] J. D. Louk, J. Mol. Spectry. **4**, 298 (1960).