JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

A Problem of Analytic Completion Related to the Jost-Lehmann-Dyson Formula

J. BROS, A. MESSIAH, AND R. STORA
Centre d'Etudes Nucléaires de Saclay, France
(Received April 17, 1961)

One considers functions of a complex four-vector Q which are analytic in the domain formed by: (a) The future tube: ImQ in the future light-cone. (b) The past tube: ImQ in the past light-cone. (c) A complex neighborhood of a domain R of real Q space limited by two space-like surfaces.

It is shown, by using techniques pertaining to the theory of analytic functions of several complex variables, that all such functions can be analytically continued in a larger domain which coincides with the one predicted by the Jost-Lehmann-Dyson formula.

INTRODUCTION

SINCE the time dispersion relations have been suggested to hold in quantum field theory, much work has been devoted to the study of the analyticity properties of transition amplitudes.

One already knows a great deal about the analyticity properties of such functions in terms of only one complex four-vector variable Q. The so-called retarded and advanced amplitudes are distributions, boundary values of functions which are analytic in the future tube (ImQ in the future light cone C^+) and in the past tube (ImQ in the past light cone C^-), respectively. Spectral conditions, on the other hand, imply that the corresponding retarded and advanced amplitudes coincide as distributions in some domain R of real Q space. It is then asserted by the edge-of-the-wedge theorem[1–4] that both amplitudes are boundary values of a single master function which coincides with the retarded amplitude in the future tube, with the advanced amplitude in the past tube, and is furthermore analytic in a complex neighborhood of the region R.

The two aforementioned properties (analyticity in the tubes, edge-of-the-wedge theorem) also hold true when the simultaneous dependence of amplitudes on several four-vectors is considered.

In the case where the dependence on only one four-vector Q is studied, all the others being kept real in C^+, a much stronger result has been obtained by Jost and Lehmann[5] in a special case, and more generally, by Dyson,[6] who showed that, under some restrictions on the shape of the region R, the master function which generates both the retarded and the advanced amplitudes, is analytic in a much larger domain than the one described above.[7] This result was obtained by means of techniques which belong to the theory of distributions and of partial differential equations.[8] Its generalization to the case where several four-vectors are involved, if it exists, has not been found yet. It might therefore be of interest to try and apply general analytic completion techniques, in order to enlarge the initial domain provided by application of the tube theorem for Fourier transforms, and by the edge-of-the-wedge theorem.

One may hope that solving this problem, in the case where the dependence upon only one four-vector is taken care of, might give some hint as to how one might attack it in a more complicated case, and reveal some general features of the method.

[1] H. Bremermann, R. Oehme, and J. G. Taylor, Phys. Rev. **109**, 2178 (1958).
[2] F. J. Dyson, Phys. Rev. **110**, 579 (1958).
[3] A. Beurling and L. Garding (to be published).
[4] H. Epstein, J. Math. Phys. **1**, 524 (1960).
[5] R. Jost and H. Lehmann, Nuovo cimento **5**, 1598 (1957).
[6] F. J. Dyson, Phys. Rev. **110**, 1460 (1958).
[7] J. Bros, M. Froissart, R. Omnes, and R. Stora (unpublished). An account of this work is given in: R. Omnes, *Lectures at Les Houches Summer School of Theoretical Physics* (Hermann, Paris, 1961).
[8] In so far as distribution theory refinements are concerned see, e.g., A. S. Wightman, *Lectures at Les Houches Summer School of Theoretical Physics* (Hermann, Paris, 1961).

In this paper, we just solve the simplest problem at hand, namely, we construct the holomorphy envelope of the domain $\mathcal{J}$ union of (a) the forward tube $\mathcal{T}^+$: $\mathrm{Im}Q\in C^+$. (b) the past tube $\mathcal{T}^-$: $\mathrm{Im}Q\in C^-$. (c) a complex neighborhood $\mathcal{R}$ of a domain R of real Q space, which we assume[6] to be bounded by the two space-like surfaces, R^+, R^-. $\mathcal{R}$ connects up $\mathcal{T}^+$ and $\mathcal{T}^-$.

The basic tool to be used is the so-called Continuity Theorem[8–11] which can be applied as follows:

Intersect the initial domain $\mathcal{J}$ with some continuously varying analytic manifold $\mathfrak{M}$; if one can find on $\mathfrak{M}$ a disk $\mathfrak{D}$, which, initially and for a while, lies inside $\mathcal{J}$, then $\mathfrak{D}$ belongs to the holomorphy envelope $\mathcal{J}^\dagger$ of $\mathcal{J}$ as long as its boundary $\partial\mathfrak{D}$ stays inside $\mathcal{J}$. This procedure of completion is basically simple; the main difficulty stems from the fact that one usually does not visualize well both the domain $\mathcal{J}$ and the analytic manifolds $\mathfrak{M}$. In the present case, however, it has proved possible to reduce the problem down to elementary geometry.

In Sec. I we introduce a real representation of n-dimensional complex space $\mathfrak{C}_n$ in n-dimensional real space $\mathfrak{R}_n$. We then describe in detail "real" linear and quadratic analytic manifolds which, in view of the structure of the initial domain, will provide us with a sufficiently broad class of disks.

In Sec. II, we give several simple characterizations of the domain $\mathcal{J}$, in terms of linear and quadratic manifolds, and the real representations thereof.

In Sec. III, the continuity theorem is applied to a simplified problem in which only two complex variables are involved.

The holomorphy envelope $\mathcal{J}^\dagger$ of $\mathcal{J}$ is then constructed in Sec. IV by repeated use of the results of Sec. III.

I. COMPLEX GEOMETRY

We shall use the following geometrical representation of n-dimensional complex space $\mathfrak{C}_n$: to each point

$$\mathbf{M}=\mathbf{M}_1+i\mathbf{M}_2 \quad \text{of} \quad \mathfrak{C}_n, \quad (M_1, M_2\in\mathfrak{R}_n),$$

is associated a real vector $V(M)$ in $\mathfrak{R}_n$, with origin at M_1 and equipollent to $\mathbf{M}_2$; real points of $\mathfrak{C}_n$ ($\mathbf{M}_2=0$), are represented by zero-vectors, i.e., points of $\mathfrak{R}_n$.

This representation helps a great deal visualizing some simple analytic manifolds whose equations have real coefficients; in the following, *unless otherwise stated, we shall always deal with such manifolds*, which we call "real" manifolds.

Consider, for instance, a one-dimensional[12] real, linear manifold $\mathfrak{L}\subset\mathfrak{C}_n$ defined by

$$\mathbf{M}=\mathbf{O}+t\mathbf{u}, \tag{I.1}$$

where $\mathbf{O}$ and $\mathbf{u}$ are real vectors and $t\in\mathfrak{C}_1$. $\mathfrak{L}$ is represented in $\mathfrak{R}_n$ by the set of all vectors V supported by the real straight line Re$\mathfrak{L}$ defined from (I.1) by restricting t to real values. A "complex segment" $[AB]$ will be defined by excluding from $\mathfrak{L}$ all real points outside the open segment $AB\subset$Re$\mathfrak{L}$, i.e., by having t vary over a suitably defined cut plane. Re$\mathfrak{L}$ is called the support of the complex segment.

We shall also make extensive use of quadratic analytic manifolds Γ, some properties of which we now state.

Consider first a one-dimensional[12] quadratic manifold Γ in a 2-dimensional linear manifold $\mathfrak{C}_2\subset\mathfrak{C}_n$. The real points of Γ are represented by the points of a second

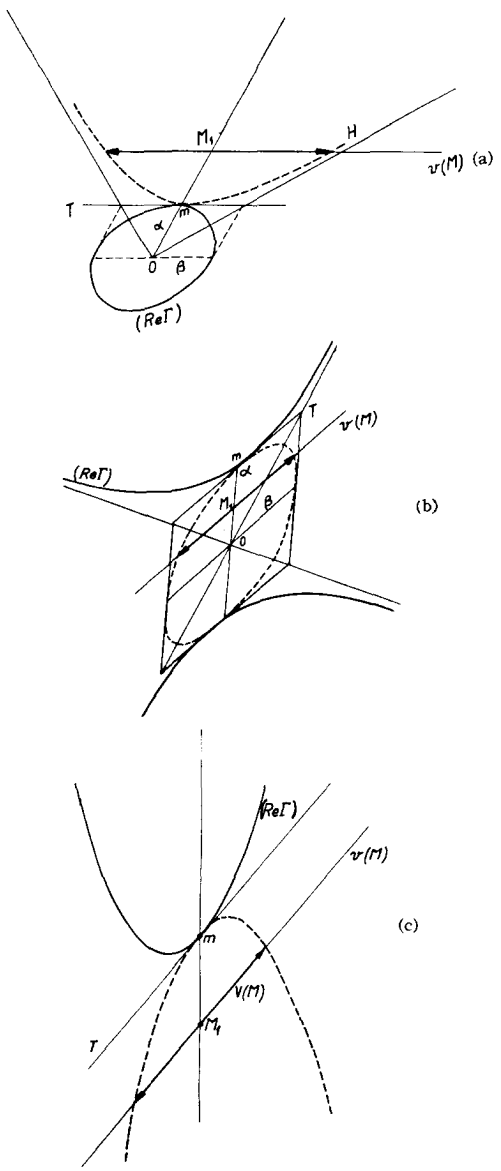

FIG. 1. Vector representation of a "real" quadratic manifold. (a) ReΓ is an ellipse. (b) ReΓ is a hyperbola. (c) ReΓ is a parabola. ReΓ is represented in full. The dashed lines represent the trajectory of the tip of $V(M)$ when M_1 describes the diameter of ReΓ conjugate to $v(M)$.

[9] Or "Theorem of the disc"; See, e.g., references 10 and 11.

[10] P. Lelong, Mimeographed notes on the theory of analytic functions of several complex variables (Saclay, 1960).

[11] H. Bremermann, Trans. Am. Math. Soc. 82, 17 (1956).

[12] Unless otherwise stated, we shall always characterize analytic manifolds by their number of *complex* dimensions.

degree curve $\mathrm{Re}\Gamma \subset \mathfrak{R}_2$, which we assume to be non-degenerate. Each complex point $M \in \Gamma$ is represented by a vector $V(M)$ whose support $v(M)$ does not intersect $\mathrm{Re}\Gamma$, because $v(M)$ is the real part of a linear manifold which has, in common with Γ, only two points: $\mathbf{M}=\mathbf{M}_1+i\mathbf{M}_2$ and its complex conjugate $\mathbf{M}^*=\mathbf{M}_1-i\mathbf{M}_2$. Conversely, on any straight line v which does not intersect $\mathrm{Re}\Gamma$, one finds two opposite vectors with common origin, $V(M)$ and $V(M^*)$, which represent two complex conjugate points of Γ. Their common origin is the intersection of v with the diameter which is conjugate to the direction of v with respect to $\mathrm{Re}\Gamma$ [Fig. 1 (a)–(c)].

These properties are best seen from an analytic parametrization of Γ.

For instance, if $\mathrm{Re}\Gamma$ is an ellipse referred to a couple of conjugate half diameters $\mathbf{a}$, $\mathbf{b}$, there corresponds to each point $M \in \Gamma$ a complex angle θ such that

$$\mathbf{M}(\theta)=\mathbf{O}+\mathbf{a}\cos\theta+\mathbf{b}\sin\theta, \tag{I.2}$$

where $\theta=\varphi+i\psi$, $0 \leqslant \varphi < 2\pi$, $-\infty < \psi < +\infty$.

For fixed φ, the origin M_1 of $V(M)$ describes that part of the half axis Om, $(\mathbf{m}=\mathbf{M}(\varphi) \in \mathrm{Re}\Gamma)$, which is outside $\mathrm{Re}\Gamma$ [Fig. 1(a)], while $v(M)$ stays parallel to the tangent mT to $\mathrm{Re}\Gamma$ (i.e., the direction conjugate to Om). Note that $R=OM_1$ and $V=|V(M)|$ are related to the half-diameter $Om=\alpha$ and the conjugate half-diameter β of $\mathrm{Re}\Gamma$, according to

$$R^2/\alpha^2 - V^2/\beta^2 = 1, \tag{I.3}$$

so that the tip of $V(M)$ describes a branch of a hyperbola, H tangent to $\mathrm{Re}\Gamma$ at m, and with the same center O [Fig. 1(a)].

For fixed ψ, on the other hand, M_1 describes an ellipse homotetic to $\mathrm{Re}\Gamma$, with center O and conjugate half-diameters $\mathbf{a}\cosh\psi$, $\mathbf{b}\sinh\psi$; $v(M)$ is tangent at M_1 to this ellipse.

An elliptic "complex segment" $[A \frown B]$ on Γ will be defined as the set of points: $\{M\colon \varphi_A < \varphi < \varphi_B, -\infty < \psi < +\infty\}$, A and B being two points of $\mathrm{Re}\Gamma$. The origins M_1 of the corresponding vectors are restricted within the angle $(\mathbf{OA},\mathbf{OB})$ [Fig. 2(b)]. It is then easy to visualize a disk belonging to $[A \frown B]$ as the image of a disk belonging to the strip $\varphi_A < \varphi < \varphi_B$ of the θ complex plane: Consider in this strip a closed curve δ which, for simplicity, we choose symmetrical with respect to the φ axis; δ encloses a disk Δ [Fig. 2(a)]. When θ describes δ, the origin M_1 of $V(M)$ describes a portion of a curve d limited at two points U, $(\theta=\varphi_U)$, V, $(\theta=\varphi_V)$, from U to V and back. If one starts from a point $M_1 \in d$ with a certain orientation of $V(M)$, one shall go through M_1 again, but $V(M)$ will then take up the opposite orientation $(\psi \to -\psi)$, after having vanished when M_1 is at U or V. Thus, the image D of Δ is the set of all vectors $V(M)$ $(M \in \Gamma)$, with origins inside the region limited by d and the arc $U \frown V$ of $\mathrm{Re}\Gamma$ [Fig. 2(b)].

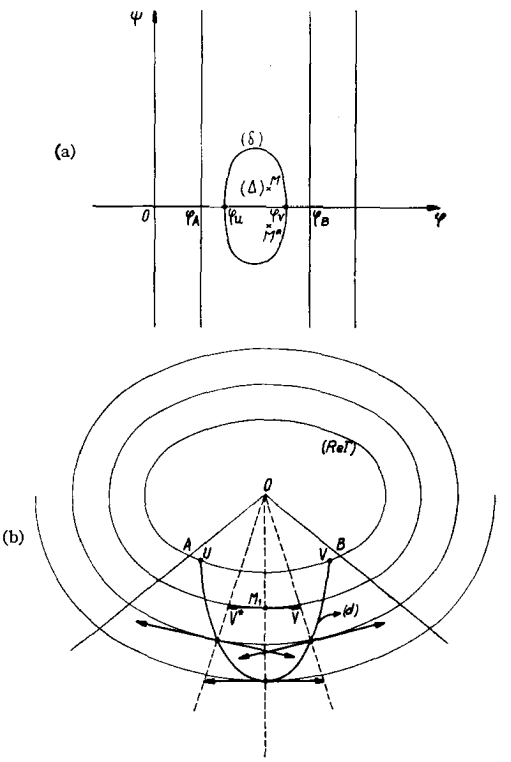

FIG. 2. A disk on an elliptic segment. (a) A disk Δ in θ-complex plane. (b) Its vector image; the dashed and light curves are level curves φ=constant (diameters), ψ=constant, (ellipses), for the origin M_1 of $V(M)$. M_1V and M_1V^* are the images of M and M^*.

Of particular interest to us is the observation that, in the degenerate case where $|\mathbf{b}| \to 0$, the elliptic complex segment $[A \frown B]$ goes to the linear complex segment defined by the limits of A and B.

The above discussion is easily extended to the cases where $\mathrm{Re}\Gamma$ is a hyperbola or a parabola [cf. Fig. 1 (b) and (c)]. We shall not go into this, since only elliptic disks will be used in subsequent arguments.

One can generalize some of the above results to quadratic manifolds of higher dimensions (e.g., the manifold $Z=Q^2$ used in Sec. II). If $\mathrm{Re}\Gamma$ is a $(n-1)$-dimensional quadratic surface in $\mathfrak{R}_n$, then a support $v(M)$, $M \in \Gamma$, cannot intersect $\mathrm{Re}\Gamma$ unless it belongs to $\mathrm{Re}\Gamma$. The latter situation can only occur if $\mathrm{Re}\Gamma$ contains, in its tangent planes, $(n-2)$-dimensional quadratic cones.

Before we leave these geometrical preliminaries we shall finally describe in terms of representative vectors the projection m of a point M, from a real pole into a real plane.

If the projection is a cylindrical one, the representative vector $v(m)$ of the projection m of M is nothing else than the cylindrical projection of $V(M)$.

If the projection is a conical one, of pole O, this property does not hold any more, in general: m is the intersection of the linear manifold OM with the projection plane. Now OM is not real unless $v(M)$ goes through

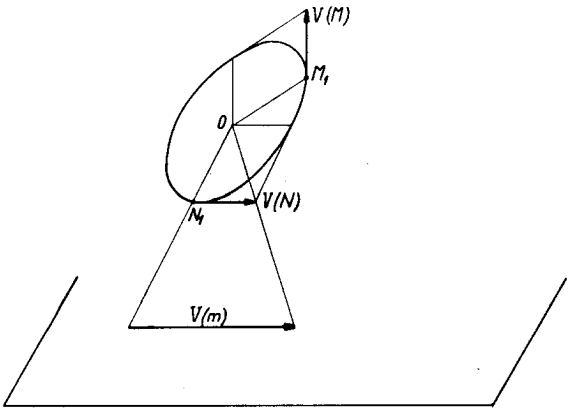

FIG. 3. Vector representation of a conical projection. m, represented by $V(m)$, is the projection of M, represented by $V(M)$.

O. The vector representation of the complex straight line is however simple [cf. Fig. 3]. It consists of all points N such that: $\mathbf{ON}=\rho e^{i\varphi}\mathbf{OM}$, i.e.:

$$\mathbf{ON}_1=\rho[\cos\varphi\mathbf{OM}_1-\sin\varphi\mathbf{OM}_2],$$
$$\mathbf{ON}_2=\rho[\sin\varphi\mathbf{OM}_2+\cos\varphi\mathbf{OM}_2].$$

It therefore follows that $v(m)$ is in the projecting 2-plane $O-V(M)$ (Fig. 3), which will be sufficient for further applications.

II. THE INITIAL DOMAIN

1. The real domain R is assumed to be bounded by two space-like surfaces. For convenience, this structural property of R will be called henceforth *space-boundedness*; it plays an essential role throughout this work. Clearly, the requirement that the domain R be space-bounded is equivalent to the requirement that any time-like segment SS' whose end points belong to R has all its points in R. Better, any time-like path $S\frown S'$, i.e., any segment of curve linking S to S' and whose tangents are all time-like, has all its points in R; accordingly, the whole double light-cone $\langle SS'\rangle$ around SS' belongs to R.

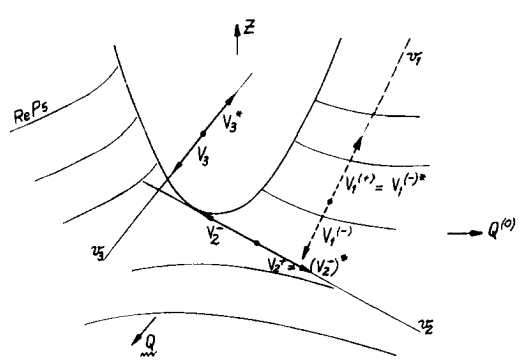

FIG. 4. A three-dimensional view of P_5 and its complex points. v_1 is below $\mathrm{Re}P_5$; $V_1^{\pm}$ represent points of the tubes $\mathcal{T}_5^{\pm}$. v_2 belongs to $\mathrm{Re}P_5$; $V_2^{\pm}$ represent points on the boundaries of $\mathcal{T}_5^{\pm}$. v_3 is above $\mathrm{Re}P_5$.

Thus, the initial domain $\mathcal{J}$, as defined in the Introduction, is the union of the two tubes $\mathcal{T}^+$, $\mathcal{T}^-$ with a complex neighborhood of a space-bounded real domain R.

2. We then recall a convenient representation $\mathcal{J}_5$ of $\mathcal{J}$ which already proved useful in an analysis of the Jost-Lehmann-Dyson formula.[7]

Consider the analytic manifold P_5,

$$Z=Q^2=Q^{(0)2}-\mathbf{Q}^2, \tag{II.1}$$

in 5-dimensional (Z,Q) space. To each point $M(Z,Q)\in P_5$

$$Z=Z_1+iZ_2, \quad Q=Q_1+iQ_2,$$
$$Z_1=Q_1^2-Q_2^2, \quad Z_2=2Q_1\cdot Q_2, \tag{II.2}$$

we associate a representative vector $V(M)$, with support $v(M)$ as defined in Sec. I. Then (cf. Fig. 4), if

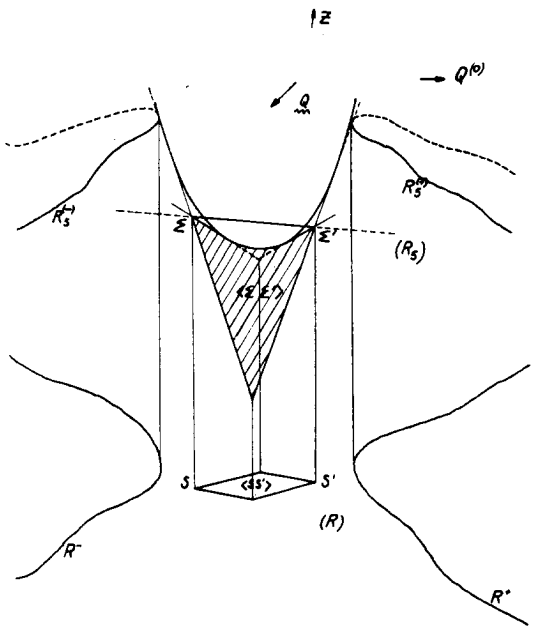

FIG. 5. The double-cone structure of R and R_5.

$Q_2^2>0$ (respectively $Q_2^2<0$), $Z_1<Q_1^2$ (respectively $Z_1>Q_1^2$) and $v(M)$ is entirely below[13] (respectively above) $\mathrm{Re}P_5$; if $Q_2^2=0$, $Z_1=Q_1^2$ and $v(M)$ belongs to $\mathrm{Re}P_5$.

Concerning the image of real points ($Q_2=0$), we have the following geometrical properties. If $\Sigma,\Sigma'\in\mathrm{Re}P_5$ are the images of two points separated by a time-like (respectively space-like) interval, the straight segment $\Sigma\Sigma'$ is above (respectively below) $\mathrm{Re}P_5$. Further, the image of a time-like path $S\frown S'$ is a segment of curve $\Sigma\frown\Sigma'$ drawn on $\mathrm{Re}P_5$, all tangents of which are entirely below $\mathrm{Re}P_5$; the set of all "*time-like*" *paths* linking Σ to Σ' describe a "*double light cone*" $\langle\Sigma\Sigma'\rangle$ with vertices Σ,Σ', which is the image on $\mathrm{Re}P_5$ of the double light cone $\langle SS'\rangle$.

[13] The ascendant vertical is chosen along the positive Z direction.

Now the image $\mathcal{G}_5$ of $\mathcal{G}$, on P_5, consists of (a) the union of the two tubes $\mathcal{T}_5=\mathcal{T}_5^+\bigcup\mathcal{T}_5^-$, i.e., the set of all vectors with supports below $\mathrm{Re}P_5$; and (b) a complex neighborhood $\mathcal{R}_5\subset P_5$ of a real domain $R_5\subset\mathrm{Re}P_5$, represented by a set of small vectors with origins close to R_5, and supports above, below, or on $\mathrm{Re}P_5$. R_5 is defined by its projection R into Q space.

In view of the spaceboundedness of R, if $\Sigma,\Sigma'\in R_5$, the segment $\Sigma\Sigma'$ being above $\mathrm{Re}P_5$, the whole "double light-cone" $\langle\Sigma\Sigma'\rangle$ with vertices Σ,Σ' belongs to R_5 (cf. Fig. 5).

Some comments about the intersections of P_5 by planes are in order. First, consider a 4-dimensional plane π_4, whose equation may be conveniently written $Z-2Q\cdot u+u^2=\zeta$ (u, a real 4-vector; ζ, a real number). (Planes parallel to the Z axis are not considered here.) Its intersection with P_5 projects in Q space along the hyperboloid, $(Q-u)^2=\zeta$. It is a one-sheeted or a two-sheeted hyperboloid according to whether π_4 is below ($\zeta<0$) or above ($\zeta>0$) the plane which is parallel to it and tangent to $\mathrm{Re}P_5$. If $\zeta=0$, the intersection reduces to a cone.

Two-sheeted hyperboloids are of particular interest. Straight segments with end points on the same sheet are below $\mathrm{Re}P_5$ (space-like interval); those with end points on different sheets are above $\mathrm{Re}P_5$ (time-like interval). In keeping with Dyson's terminology, such a two-sheeted hyperboloid is called *admissible* with respect to R_5 if it has no point in R_5; it is called *doubly inadmissible* with respect to R_5 if both sheets have points in R_5.

All these considerations, including the notion of admissibility, apply with obvious changes to the intersection of P_5 by 3- and 2-dimensional planes.[14]

3. The consequences of Lorentz invariance are easily formulated. Since the notions of time-like path and tube are invariant with respect to inhomogeneous Lorentz transformations, our problem can be set up in any Lorentz frame; a change of Lorentz frame amounts to a linear change of variables in Q space. Each Lorentz transformation induces in (Z,Q) space a linear transformation which leaves P_5 invariant. For instance, a typical Lorentz translation (χ) reads in (Z,Q) space:

$$Q'=\chi+Q,\quad Z'=\chi^2+2\chi\cdot Q+Z. \tag{II.3a}$$

A typical homogeneous Lorentz transformation (Λ) reads

$$Q'=\Lambda Q,\quad Z'=Z. \tag{II.3b}$$

All these transformations conserve not only P_5 as a whole, but also $\mathcal{T}_5$ and the notations of "time-like" path and "double cone."

4. The Continuity Theorem will not be applied directly in Q space, but after a change of variables, which can be described geometrically as a conical projection of P_5 onto a 4-dimensional plane. We proceed with the definition of such projections and of the related changes of variables.

We consider only the case where the center of projection O is above $\mathrm{Re}P_5$. The plane of projection is the plane π conjugate to O with respect to $\mathrm{Re}P_5$.

First, we bring O on the positive Z axis by a suitable Lorentz translation [cf. Eq. (II.3a)]. In the new frame, the coordinates of O are (ζ,O) with ζ real >0, and the equation of π is $Z=-\zeta$.

To each point $M(Z,Q)$ of P_5 there corresponds its projection $m(-\zeta, q)$ in π. A straightforward calculation gives

$$q=-2\zeta Q/(Q^2-\zeta). \tag{II.4}$$

Equation (II.4) defines the two-to-one change of variables associated with the conical projection. The one-to-two inverse transformation reads:

$$Q_\pm=q/(1\pm[1+(q^2/\zeta)]^{\frac{1}{2}}). \tag{II.5}$$

Note that

$$Q_\pm=-\zeta Q_\mp/Q_\mp^2. \tag{II.6}$$

Thus, the two points in Q space, which lead to the same point in q space, are transforms of each other in a par-

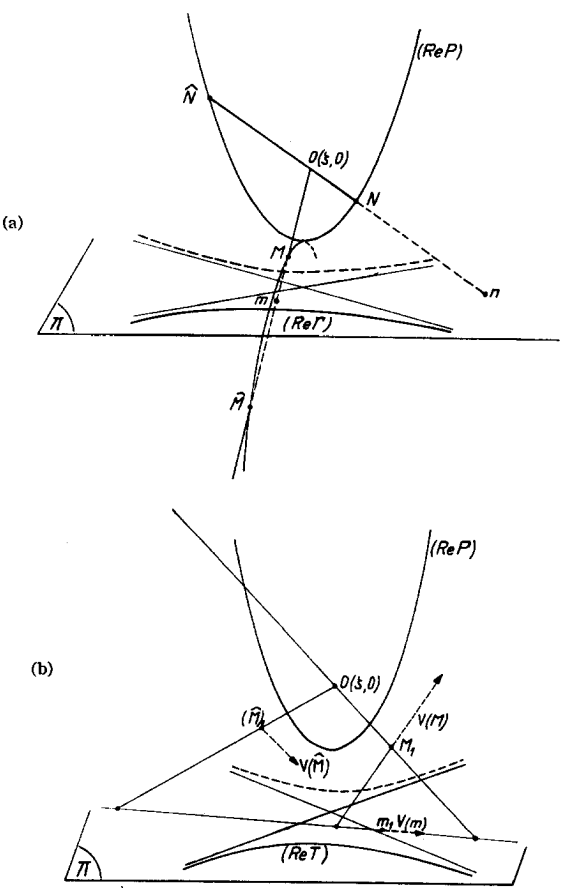

FIG. 6. Conical projection of P_5 and related conformal transformation: (a) real points, (b) complex points.

[14] In the case of 2-dimensional planes, hyperboloids reduce to hyperbolas. The property of a two-sheeted hyperboloid mentioned above applies to a hyperbola, whose plane is above the parallel plane tangent to $\mathrm{Re}P_5$ deduced from it by a translation along the Z direction.

ticular conformal transformation. The latter can be defined as the following one-to-one correspondence between points of P_5

$$M(Z,Q) \leftrightarrow M^c(Z^c,Q^c), \qquad Z^c=\zeta^2/Z, \quad Q^c=-\zeta Q/Q^2. \tag{II.7}$$

Now, let us call Γ the intersection of the analytic manifolds P_5 and π. Γ is defined by the equation $q^2+\zeta=0$, and ReΓ is a one-sheeted hyperboloid (since $\zeta>0$). The set of real points for which $q^2+\zeta>0$ will be called the *inside* of ReΓ (i.e., the region bounded by ReΓ which contains its asymptotic cone).

In the aforementioned conical projection, points on ReP_5 lead to points inside or on ReΓ, straight lines entirely below ReP_5 lead to straight lines entirely inside ReΓ, straight lines entirely above ReP_5 lead to straight lines which either intersect ReΓ or are entirely outside it (cf. Fig. 6).

As a consequence, the image of the tubes in π is represented by the set of *all* vectors $V(m)$ of arbitrary lengths with origin and support inside ReΓ; this is obvious since the support $v(m)$ of the representative vector of the conical projection of M is the conical projection of the support $v(M)$ of the representative vector of M (cf. the end of Sec. I).

Also, the image of the region R in π is a region R_4 inside ReΓ, which is the conical projection of R_5. By definition, the projection $\sigma\frown\sigma'$ of a "time-like" path $\Sigma\frown\Sigma'$, drawn on ReP_5, will be called a "time-like" path; such a path exhibits infinite branches if $\Sigma\frown\Sigma'$ intersects the plane $Z=\zeta$. Clearly, a path $\sigma\frown\sigma'$ is "time-like" if, and only if, all its tangents lie entirely inside ReΓ (except possibly for their points of contact which lie on ReΓ). In view of the spaceboundedness of R, if $\sigma, \sigma'\in R_4$ can be linked by a "time-like" path $\sigma\frown\sigma'$, all the points of this path belong to R_4.

Note that the projection of the boundary ∂R_5 of R_5 does not completely coincide with the boundary ∂R_4 of R_4; in general, only part of ∂R_5 projects on ∂R_4 which also contains $R_5\bigcap$ReΓ.

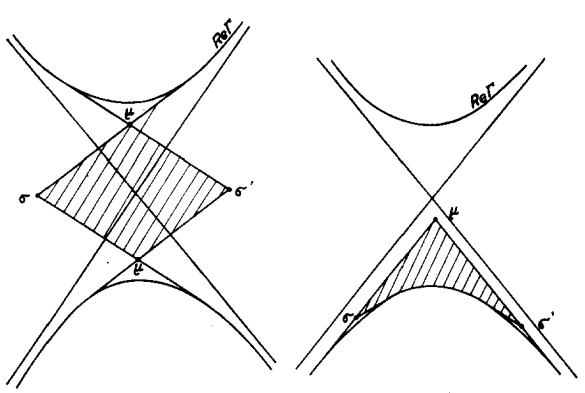

FIG. 7. Typical shapes of finite "double cones." (a) $\sigma\sigma'$ does not intersect ReΓ. (b) $\sigma\sigma'$ intersects ReΓ.

III. A TWO DIMENSIONAL PROBLEM

1. In this section we consider a problem in two complex variables, $M\equiv(u,v)\in\mathfrak{C}_2$.

Call Γ a certain quadratic manifold in $\mathfrak{C}_2$ such that ReΓ is either a hyperbola or a parabola. In the following, the "inside" of ReΓ denotes the set of real points from which two distinct real tangents to ReΓ can be drawn.

By definition, a segment of curve inside ReΓ is a (finite) "time-like" path if all the tangents to this segment are entirely inside ReΓ. The set of all the "time-like" paths linking two given points σ, σ', span the "double light cone" $\langle\sigma\sigma'\rangle$. Typical "double light cones" are shown in Fig. 7.

The following result is easily shown by inspection: Two points σ, σ' inside ReΓ can be linked by a finite "time-like" path in, and only in, the following two cases[15]: Case (a). The straight line $\sigma\sigma'$ is entirely inside ReΓ. Case (b). The straight line $\sigma\sigma'$ intersects ReΓ between σ and σ'.

Call $\mathcal{T}_2$ the set of complex points whose representative vectors have their support entirely inside ReΓ. Then the following theorem holds:

Theorem III. *Consider the domain* $\mathcal{G}_2\equiv\mathcal{T}_2\bigcup\mathcal{R}_2$, *where* $\mathcal{R}_2$ *is the complex neighborhood of a certain real region* R_2 *inside* ReΓ. *Its holomorphy envelope* $\mathcal{G}_2^\dagger$ *contains the set of all linear complex segments* $[\sigma\sigma']$ *whose end points* σ, σ' *can be linked by a finite "time-like" path belonging entirely to* R_2.

This theorem will be our basic tool in solving the problem of analytic completion of Sec. IV.[16]

Let us remark that initial domains of analyticity such as $\mathcal{G}_2$ appear naturally when one deals with analytic functions $F(Q)$ invariant under the rotation group which leaves unaltered some time-like vector (e.g., vertex functions). Taking the time-axis along the unaltered time-like vector, one can show[7] that, wherever F is analytic, it can be treated as an analytic function $\varphi(Z,Q^{(0)})$ of the two variables $Z=Q^2$ and $Q^{(0)}$. If the initial domain in Q space is the union of the tubes with a complex neighborhood of a domain R of real Q space,[6] then φ is analytic in a domain $\mathcal{G}_2$ of the type considered in Theorem III, if we set[17]

$$u=Z, \quad v=Q^{(0)},$$

[15] Otherwise stated, if, and only if, $[\sigma\sigma']$ intersects Γ. In case (a), $\langle\sigma\sigma'\rangle$ is a diamond bounded by segments of tangents to ReΓ issued from σ and σ'; in case (b), the boundary of $\langle\sigma\sigma'\rangle$ contains in addition a segment of ReΓ (cf. Fig. 7).

[16] The regions R_2 occuring in the applications to Sec. IV have simplifying features, and Theorem III, when applied to these particular cases, can be formulated in the following somewhat simpler way:

Call $\mathcal{R}_2$ the complex neighborhood of a domain R_2 of real (u,v) space which is (a) entirely inside ReΓ, (b) such that, if $\sigma, \sigma'\in R_2$, any *finite* "time-like" path linking σ to σ' belongs entirely to R_2. Then, the holomorphy envelope of the domain $\mathcal{G}_2\equiv\mathcal{T}_2\bigcup\mathcal{R}_2$ contains the set of all linear complex segments $[\sigma\sigma']$ which intersect Γ, and whose end points σ, σ' are in R_2.

[17] This domain is the cylindrical projection of $\mathcal{G}_5$ on the $(Z,Q^{(0)})$ plane. Cylindrical projections have not been considered in Sec. II. However, those used here are limiting cases of conical projections,

and if Γ is the parabola: $u=v^2$. Thus, direct application of Theorem III to the case of vertex functions yields physically relevant results, namely, the analyticity of the function on the cut planes defined by the complex segments $[\sigma\sigma']$.

2. In order to prove Theorem III, we have to choose a set of convenient disks and apply the continuity theorem.

Now, the use of disks on linear manifolds cannot lead us very far. Consider the restriction $\mathfrak{L}\bigcap\mathcal{J}_2$ of $\mathcal{J}_2$ to a certain real analytic manifold $\mathfrak{L}$ defined by Eq. (I.1). As long as Re$\mathfrak{L}$ does not intersect ReΓ, it is a cut plane in the variable t (with the cut on the real axis); but, as soon as Re$\mathfrak{L}$ intersects ReΓ, it does not contain any point of $\mathcal{T}_2$ and consists at most in isolated domains in neighborhoods of real finite segments. Thus, the Continuity Theorem cannot yield any result when applied to disks on linear manifolds which intersect ReΓ. In particular, no complex point outside $\mathcal{T}_2$ can be reached by direct linear completion. A similar situation is met when one tries to find the holomorphy envelope of a tube with a nonconvex base. In this case, a convenient trick[8] consists in slicing up the tube by slightly bent manifolds. We shall use essentially the same trick here.

Our first step is to define a convenient set of "elliptic" disks of the type discussed in Sec. I.

Call σ, σ' two points inside ReΓ which can be linked by a "time-like" path. The two tangents issued from σ' intersect the two tangents issued from σ at four points. Among these, we choose one which is a vertex of the "double light cone" $\langle\sigma\sigma'\rangle$; call it μ. [There is just one such point in case (b), there are two in case (a) (cf. Fig. 7)]. Note the following two geometrical properties of the triangle $\mu\sigma\sigma'$:

(i) Straight lines issued from μ outside the angle $(\mathfrak{y}\sigma, \mathfrak{y}\sigma')$ do not intersect ReΓ; their complex points are in $\mathcal{T}_2$.

(ii) If σ goes to σ' along a "time-like" path, μ goes to σ along the relevant tangent to ReΓ and the triangle $\mu\sigma\sigma'$ suffers a continuous deformation without ever becoming degenerate, i.e., none of its angles ever vanishes during the process.

The family of ellipses tangent to $\mu\sigma$, $\mu\sigma'$ at σ, σ', respectively, form a linear set. We consider on one of them the elliptic complex segment $[\sigma\frown\sigma']_p$ whose real points fall inside the triangle $\mu\sigma\sigma'$. p is the parameter labeling each ellipse of the set and plays the role of a bending parameter. To be specific, let us define it as the distance from the center of the ellipse to the middle of the straight segment $\sigma\sigma'$.[18]

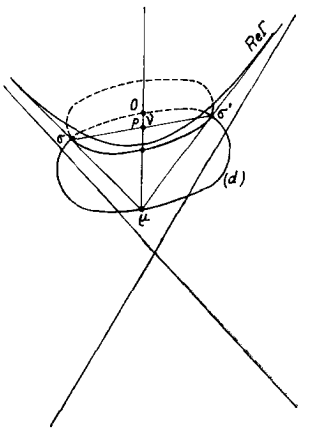

FIG. 8. Definition of the disk $D[\mu,\sigma\frown\sigma']_p$.

On each such elliptic segment, we consider the disk, henceforth denoted by $D[\mu,\sigma\frown\sigma']_p$, whose curve d (notations of Sec. I) is described by the origin of representative vectors whose supports contain μ (cf. Fig. 8).

Apart from the real points σ and σ', its boundary ∂D is formed by all the complex points whose supports contain μ. According to property (i) above, this complex part of ∂D is in $\mathcal{T}_2$. As a consequence, if σ and σ' are in R_2, then, for any given $p>0$, the boundary $\partial D[\mu,\sigma\frown\sigma']_p$ is entirely in $\mathcal{J}_2$. Further inspection shows that the disk $D[\mu,\sigma\frown\sigma']_p$ itself is entirely in $\mathcal{J}_2$ if Re$[\sigma\frown\sigma']_p$ is in R_2.

Now, assume that σ and σ' are linked by a finite "time-like" path entirely in R_2, and consider a disk $D[\mu,\sigma\frown\sigma']_p$ built on these two points. Keeping p fixed, we let σ' move along this path toward σ up to a position σ_1' sufficiently close to σ so that the elliptic segment Re$[\sigma\frown\sigma_1']_p$ be entirely in R_2 (cf. Fig. 9). During the motion, $D[\mu,\sigma\frown\sigma']_p$ suffers a continuous deformation, but it never ceases to be a disk [cf. property (ii) above] and its boundary ∂D remains inside $\mathcal{J}_2$. When σ' has reached the position σ_1', the whole disk is in $\mathcal{J}_2$. Applying the Continuity Theorem, we conclude that $D[\mu,\sigma\frown\sigma']_p$ belongs to the holomorphy envelope $\mathcal{J}_2\dagger$. Thus, if two points σ, σ' are linked by a finite "time-like" path entirely in R_2, for any given $p>0$, the corresponding disk $D[\mu,\sigma\frown\sigma']_p$ belongs to $\mathcal{J}_2\dagger$.

In order to complete the proof of Theorem III, we have to extend the above results to the linear complex segments $[\sigma\sigma']$ which are, crudely speaking, the limits of the disks $D[\mu,\sigma\frown\sigma']_p$ when the bending parameter p goes to zero. For this purpose, it suffices to show that each point $M\in[\sigma\sigma']$ belongs to a disk which fulfills the above conditions. The proof of this point, which is elementary, is outlined in Appendix A.

IV. THE JOST-LEHMANN-DYSON DOMAIN

1. The purpose of this section is to show that the holomorphy envelope $\mathcal{J}\dagger$ of the domain $\mathcal{J}$ defined in the Introduction is the domain predicted by the Jost-Lehmann-Dyson formula.

when the center of projection goes to infinity above ReP_5. All their geometrical properties can be obtained through this limiting process.

[18] This center is on the straight line $\mu\nu$, where ν is the middle of $\sigma\sigma'$. More precisely, its locus is the half-straight line starting from ν, which does not contain μ. When $p\to 0$, the ellipse reduces to the straight segment $\sigma\sigma'$. When $p\to\infty$, the ellipse tends to a parabola.

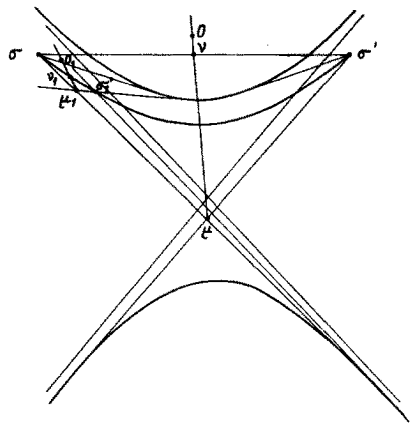

FIG. 9. The continuous deformation of a disk $D[\mu,\sigma\frown\sigma']_p$.

Most of the discussions and results will be given in the framework of the 5-dimensional (Z,Q) space introduced in Sec. II. We recall that $\mathcal{G}_5$, image of $\mathcal{G}$, is the domain on P_5 consisting in the union of the set $\mathcal{T}_5$ of all complex points whose representative vectors have their supports below $\mathrm{Re}P_5$, and of the complex neighborhood $\mathcal{R}_5$ of the "space-bounded" real domain R_5

$$\mathcal{G}_5 \equiv \mathcal{T}_5 \bigcup \mathcal{R}_5.$$

We denote by $\mathcal{G}_5$ the image of $\mathcal{G}^\dagger$.

2. Our first step is the establishment of a procedure of analytic completion. The result is contained in the following theorem:

Theorem IV. *The image $\mathcal{G}_5^\dagger$ in (Z,Q) space of the holomorphy envelope of $\mathcal{G}$ contains the enlarged domain $\mathfrak{h}(R_5)$ defined as the union of $\mathcal{G}_5$ with: (a) the set $h(R_5)$ made up with all real finite segments $\Sigma\frown\Sigma'$, Σ, and $\Sigma'\in R_5$, of hyperbolas (h) doubly inadmissible with respect to R_5; (b) all complex points of such hyperbolas; and (c) complex neighborhoods thereof.*

Proof of Theorem IV. Consider a two-dimensional plane (p) which intersects P_5 along a hyperbola (h) which is doubly inadmissible with respect to R_5. Then it is possible to find in $(h)\bigcap R_5$ four real points, two of them, A and A', on one branch of (h), the other two, B and B', on the other branch. We call $[h]$ the set which consists of (a) all points of the real finite segments of hyperbola AA' and BB'; (b) all complex points of (h). Theorem IV states that any such set $[h]$ is necessarily in $\mathcal{G}_5^\dagger$.

The proof of this statement is obtained essentially by performing a change of variables in which $[h]$ is transformed into a linear complex segment, and applying Theorem III to analytic functions of these new variables.

Call O the intersection of AA' and BB',[19] π_4 the plane conjugate to O with respect to $\mathrm{Re}P_5$. Clearly, O is above $\mathrm{Re}P_5$. The change of variables that we want to consider is the one associated with the conical projection from center O into the plane π_4. The projection of $[h]$ is obviously the linear complex segment $[ab]$ whose end points a, b are the projections of A and A', B and B', respectively (cf. Fig. 10).

Conical projections and associated changes of variables have been described in Sec. II. Unless otherwise stated, we take over the notations of Sec. II. The change of variables is the two-to-one change, $Q_\pm \leftrightarrow q$ defined by Eqs. (II.4) and (II.5). To each function $F(Q)$ analytic in domain $\mathcal{G}$, there corresponds the double-valued function $\varphi(q)$

$$\varphi(q)=F(Q)=F\left(\frac{q}{1\pm(1+q^2/\zeta)^{\frac{1}{2}}}\right). \qquad \text{(IV.1)}$$

$\varphi(q)$ is analytic on a two-sheeted domain built over the projection $\mathcal{G}_4$ of $\mathcal{G}_5$ into π_4.

The procedure of analytic continuation to $[ab]$ cannot be applied directly to $\varphi(q)$, since it is double valued. However, it is possible to express $\varphi(q)$ in terms of two single-valued functions $\varphi_e(q)$, $\gamma_e(q)$, to which this procedure can be applied. To this effect, we write $F(Q)$ as a sum of two terms which are, respectively, even and odd with respect to the conformal transformation

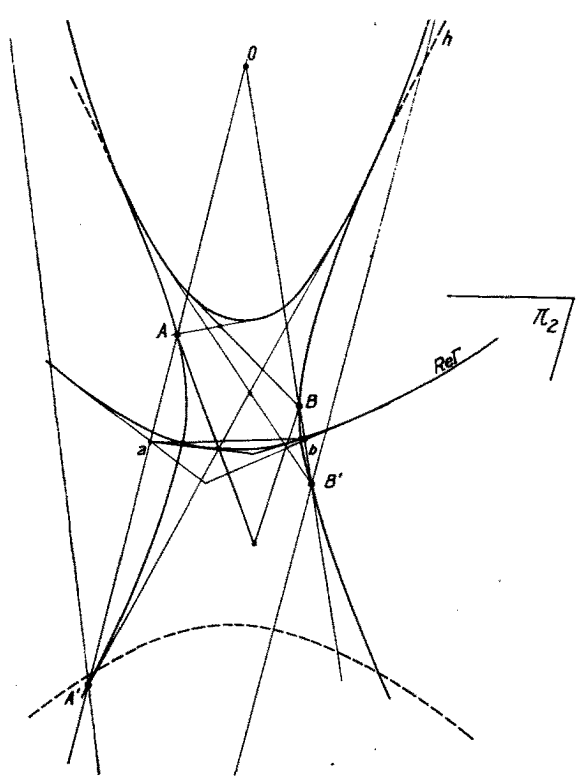

FIG. 10. $\mathfrak{h}$-completion in space E_3.

[19] The case when AA' and BB' are parallel straight lines, which would require a cylindrical rather a conical projection, need not to be considered here. For the purpose of proving Theorem IV, it is sufficient to consider on each doubly inadmissible hyperbola (h) sets of points A, A', B, B', such that AA' and BB' intersect at a finite distance.

$Q \leftrightarrow Q^c$ defined by Eq. (II.7)

$$F(Q)=F_e(Q)+\frac{Q^2+\zeta}{Q^2-\zeta}G_e(Q), \qquad \text{(IV.2)}$$

with

$$F_e(Q)=\tfrac{1}{2}[F(Q)+F(-\zeta Q/Q^2)], \qquad \text{(IV.3)}$$

$$G_e(Q)=\frac{1}{2}\frac{Q^2-\zeta}{Q^2+\zeta}[F(Q)-F(-\zeta Q/Q^2)]; \qquad \text{(IV.4)}$$

and we set

$$\varphi_e(q)=F_e(Q), \quad \gamma_e(q)=G_e(Q). \qquad \text{(IV.5)}$$

Equation (IV.1) can be more conveniently rewritten

$$\varphi(q)=\varphi_e(q)+[1+(q^2/\zeta)]^{\frac{1}{2}}\gamma_e(q). \qquad \text{(IV.6)}$$

Since $F_e(Q)$ and $G_e(Q)$ are invariant under the conformal transformation (II.7), $\varphi_e(q)$ and $\gamma_e(q)$ are single-valued. More precisely, $F_e(Q)$ and $G_e(Q)$ are well-defined analytic functions in the domain $\mathcal{D}\cap\mathcal{D}^c$, where $\mathcal{D}^c$ is the conformal transform of $\mathcal{D}$. That $G_e(Q)$ be analytic in this whole domain, including the manifold $Q^2+\zeta=0$, in spite of the occurrence of the denominator $Q^2+\zeta$ in Eq. (IV.4), follows from Weierstrass' preparation theorem,[10] which kills out any possible singularity on this manifold. As a consequence, $\varphi_e(q)$ and $\gamma_e(q)$ are well-defined analytic functions in the corresponding domain $\mathcal{D}_4\cap\mathcal{D}_4{}^c$, apart from possible singularities on the manifold $q^2+\zeta=0$ due to the occurrence of the square root in Eq. (II.5) which defines the change of variables. Now, according to Riemann's Theorem,[10] since these two functions are analytic and bounded in the neighborhood of all points of the manifold $q^2+\zeta=0$ belonging to the domain $\mathcal{D}_4\cap\mathcal{D}_4{}^c$, they are necessarily analytic at these points. In summary, $\varphi_e(q)$ and $\gamma_e(q)$ are single-valued analytic functions in $\mathcal{D}_4\cap\mathcal{D}_4{}^c$.

In order to perform the continuation, we may study the whole problem in the 3-dimensional plane E_3 defined by $(\mathfrak{p})$ and the vertical direction OZ. Figure 10 gives a view of the over-all situation in the 3-dimensional space E_3. We call P_3 and π_2 the intersection of P_5 and π_4 with E_3, respectively. We identify π_2 with the two-dimensional space of Sec. III.

We set $\Gamma\equiv P_3\cap\pi_2$ and $\mathcal{D}_2\equiv\mathcal{D}_4\cap\mathcal{D}_4{}^c\cap\pi_2$. Re$\Gamma$ is to be identified with the hyperbola of Sec. III, and $\mathcal{D}_2$ with the initial domain introduced in Theorem III. Indeed, we have

$$\mathcal{D}_2=\mathcal{T}_2\cup\mathcal{R}_2.$$

$\mathcal{T}_2$ is the projection of the E_3 intersection of the tubes, i.e., the set of all complex points whose representative vectors have their origins and supports inside ReΓ (cf. properties of the conical projections, Sec. II). $\mathcal{R}_2$ is the complex neighborhood of a real region R_2, which is the projection of the E_3 intersection of $R_5\cap R_5{}^c$.

Clearly, a and b belong to R_2. Better, they can be linked by a finite "time-like" path which belongs entirely to R_2. Indeed, it is certainly possible to draw on ReP_3 a "time-like" path linking A to B (cf. Fig. 10). Inspection easily shows that this path is entirely below the plane $Z=\zeta$. Therefore, its conical projection is a finite "time-like" path $a\frown b$. Now, the conformal transform $A'\frown B'$ of $A\frown B$ is a "time-like" path with the same projection. Since R_5 is space-bounded, both $A\frown B$ and $A'\frown B'$ belong entirely to R_5 and, since they are the conformal transforms of each other, both also belong to $R_5{}^c$. Consequently, their common projection $a\frown b$ is a "time-like" path which belongs entirely to R_2. Therefore, Theorem III is applicable to the linear complex segment $[ab]$; the latter belongs to the holomorphy envelope $\mathcal{D}_2{}^\dagger$ of the domain $\mathcal{D}_2$.

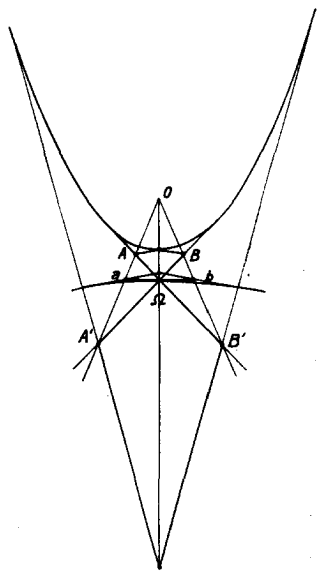

FIG. 11. $\mathfrak{h}$-completion in the degenerate case.

Otherwise stated, $\varphi_e(q)$ and $\gamma_e(q)$ can be analytically continued on the whole complex segment $[ab]$. Accordingly, $F_e(Q)$ and $G_e(Q)$, as defined by Eq. (IV.5), can be analytically continued on the set corresponding to $[ab]$ in the change of variables (II.4), since this set has no point on the manifold $Q^2-\zeta=0$, where singularities associated with this change of variables might occur. In the same way, $F(Q)$, as defined by Eq. (IV.2), can be continued on this set. In other words, the set $[h]$, reciprocal image of $[ab]$, belongs to $\mathcal{D}_5{}^\dagger$.

The foregoing argument is quite general. As can be easily verified, it holds true when $(\mathfrak{p})$ is tangent to ReP_5, in which case (h) degenerates into two straight lines (cf. Fig. 11). Q. E. D.

3. We now show that the $\mathfrak{h}$-completion process defined by Theorem IV can be iterated and that, after a finite number of iterations, one is led to a natural domain of holomorphy, namely the Jost-Lehmann-Dyson domain.

That the $\mathfrak{h}$ process can be iterated follows from the fact that the enlarged real domain $h(R_5)$ is "space-bounded," as well as R_5 itself. The proof that $h(R_5)$ is "space-bounded" is easily obtained in studying its

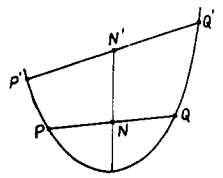

FIG. 12. NN' is a vertical segment above $\mathrm{Re}P_5$. N lies in $\mathcal{E}_5$, N' in $\partial\mathcal{E}_5$. $P'N'Q'$ is a "time-like" segment whose end points lie in the closure of $R_5{}^\dagger$. The figure is drawn in the vertical 2-plane of $NP'Q'$, which intersects $R_5{}^\dagger$ along a "time-like" segment of parabola limited at P' and Q'.

projection $h(R)$ in Q space; it then reduces to an elementary problem in three-dimensional geometry. This proof is given in Appendix B.

Repeated application of Theorem IV leads to larger and larger real domains $h(R_5)$, $h^2(R_5)\equiv h(h(R_5))$, $\cdots$, and to larger and larger domains of analyticity $\mathfrak{h}(R_5)$, $\mathfrak{h}^2(R_5)$, $\cdots$. Clearly, the largest possible real domain which could be obtained in this way is the domain $R_5{}^\dagger$ whose closure is the intersection with $\mathrm{Re}P_5$ of the *convex hull* $\mathcal{E}_5$ of the closure of R_5; indeed, any branch of doubly inadmissible hyperbola which is obtained by the $\mathfrak{h}$ process is obviously inside $R_5{}^\dagger$, since $h(R_5{}^\dagger)=R_5{}^\dagger$.

It turns out that $R_5{}^\dagger$ is actually reached after a finite number of iterations, i.e.[20]:

$$h^3(R_5)=R_5{}^\dagger. \tag{IV.7}$$

Property (IV.7) obtains because R_5 is space-bounded. Its proof, which is based on rather simple theorems about convex hulls, is given in Appendix C.

From (IV.7), we deduce

$$\mathfrak{h}^4(R_5)=\mathfrak{h}(R_5{}^\dagger).$$

As a consequence, $\mathfrak{h}^4(R_5)$ is the largest possible domain of analyticity which can be reached by repeated application of Theorem IV. Call $\mathcal{G}_5{}^\dagger$ this domain. Its real points are those of $R_5{}^\dagger$. Its complex points are the points of the tubes (i.e., the set $\mathcal{T}_5$) and the complex points of all hyperbolas which are doubly inadmissible with respect to $R_5{}^\dagger$. The latter may be alternately defined as the set of complex points of P_5 whose supports intersect the inside of $\mathcal{E}_5$ either above or on $\mathrm{Re}P_5$.

To show this last point, we note first that the support $v(M)$ of a complex point of a doubly inadmissible hyperbola necessarily intersects the inside of $\mathcal{E}_5$, for obvious geometrical reasons. Conversely, consider a complex point $M\in P_5$ for which $v(M)$ intersects the inside of $\mathcal{E}_5$ above $\mathrm{Re}P_5$. Call N a point on this intersection. Since N is inside $\mathcal{E}_5$ and above $\mathrm{Re}P_5$, there is certainly at least one segment PQ which contains it, and whose end points are in $R_5{}^\dagger$ and time-like distant. This can be seen by proving that the corresponding property holds for N on $\partial\mathcal{E}_5$ (cf. Appendix C, Point **2**), and using the space-boundedness of $R_5{}^\dagger$ (cf. Fig. 12). We then easily conclude that the plane (p) defined by the two straight lines PQ and $v(M)$ intersects P_5 along a hyperbola which is doubly inadmissible with respect to $R_5{}^\dagger$. The point M is one of the two complex points where $v(M)$ intersects this hyperbola.

In the limiting case, when N is on $\mathrm{Re}P_5$ (i.e., in $R_5{}^\dagger$), any 2-plane through $v(N)$ and tangent at N to $\mathrm{Re}P_5$ obviously intersects the latter along a degenerate doubly inadmissible hyperbola which contains M.

Q. E. D.

Summarizing, we have reached the following result (cf. Fig. 13).

Any function analytic in $\mathcal{G}$ is necessarily analytic in the larger domain $\mathcal{G}^\dagger$, whose image $\mathcal{G}_5{}^\dagger$ in (Z,Q)-space is defined as follows:

(i) Its real points form the domain $R_5{}^\dagger$, whose closure is the intersection of $\mathrm{Re}P_5$ with the convex hull $\mathcal{E}_5$ of the closure of R_5.

(ii) Its complex points are those whose representative vectors have supports either below $\mathrm{Re}P_5$ (i.e., points of $\mathcal{T}_5$) or intersecting the inside of $\mathcal{E}_5$.

Now, it has been shown in reference 7 that the complement of the above-defined domain is the union of all the hyperboloids admissible with respect to R. The Jost-Lehmann-Dyson formula provides examples of functions which are analytic in $\mathcal{G}^\dagger$ and nowhere else. This proves that $\mathcal{G}^\dagger$ is a natural domain of holomorphy, namely the Jost-Lehmann-Dyson domain.

CONCLUSION

We have solved in this paper a problem of analytic completion, the answer to which had been obtained by quite different techniques. The motivation for so doing was a methodological one. From a purely technical standpoint, the present case was moderately favorable: both the initial domain and its holomorphy envelope could be described in terms of linear or quadratic one-dimensional manifolds. Hence, the procedure of analytic completion appeared as a succession of simple geometrical steps.

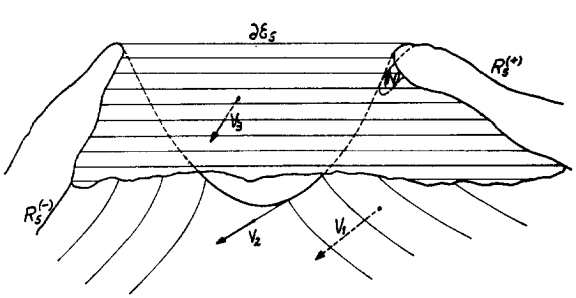

FIG. 13. A three dimensional view of the Jost-Lehmann-Dyson domain. Part of the boundary $\partial\mathcal{E}_5$ of $\mathcal{E}_5$ is represented. The completion of a "reentrant nose" (N) is indicated on the upper right part of $R_5{}^\dagger$. Three typical vectors are shown: V_1 belongs to one of the tubes; V_2 to the tubes boundary; V_3 has a space-like projection.

[20] When R is invariant under the rotations around a given time-like vector, it is only necessary to iterate the $\mathfrak{h}$ process once. In fact, the problem is then basically three-dimensional, and one can easily show that $h(R_5)=R_5{}^\dagger$. The physically relevant case, in which R is bounded by two (portions of) hyperboloids is subject to this simplification.[7]

However, it should be noted that we have only obtained part of the information which is contained in the Jost-Lehmann-Dyson representation. Besides analyticity properties, the latter puts bounds on the rate of growth of the analytic functions under consideration near the boundary of its holomorphy domain. In order to do so, we would have to make explicit use of the behavior of these functions near the boundary of the initial domain and apply the Maximum Modulus Theorem. This has clearly to be done, if one wants to derive rigorously such integral representations as dispersion relations.

ACKNOWLEDGMENTS

This work started from an illuminating discussion of one of us (R.S.) with Dr. A. S. Wightman at Les Houches Summer School of Theoretical Physics, about an unpublished work of Dr. N. Burgoyne.

We would also like to thank Dr. A. Beurling, Dr. H. Epstein, and Dr. L. Garding for communication of their work prior to publication, and Dr. V. Glaser for a stimulating discussion on the general edge-of-the-wedge problem.

APPENDIX A

The following lemma is used at the end of the proof of Theorem III:

Every point M of a linear complex segment $[\sigma\sigma']$, the end points of which can be linked by a finite "time-like" path belongs to a certain disk $D[\mu,\tau\frown\tau']_p$, whose end points, τ and τ', lie in neighborhoods of σ and σ', respectively.

Proof of the lemma: The disks to be considered are built on ellipses tangent to $\mu\sigma$ and $\mu\sigma'$, at τ and τ', respectively; μ is defined as in Sec. III. The ellipses are chosen in such a way that τ and τ' be outside the triangle $\mu\sigma\sigma'$. Thus, the intersection of $[\sigma\sigma']$ with the elliptic complex segment $[\tau\frown\tau']_p$ necessarily belongs to the disk $D[\mu,\tau\frown\tau']_p$. We call I the intersection of $\tau\tau'$ with $\sigma\sigma'$, J its harmonic conjugate with respect to σ, σ'.

(a) Real points of $[\sigma\sigma']$: We fix τ and τ' arbitrarily close to σ and σ' [cf. Fig. 14(a)]. When p goes from 0 to ∞, the elliptic segment $\mathrm{Re}[\tau\frown\tau']_p$ bends more and more; at a certain time it touches the segment $\sigma\sigma'$ at J and, from then on, intersects it at two points M and M', which stay harmonic conjugate of each other with respect to I, J, and tend to some limiting positions close to σ and σ', respectively, when p tends to ∞. Any real point of $[\sigma\sigma']$, no matter how close to σ and σ', can be reached in this way.

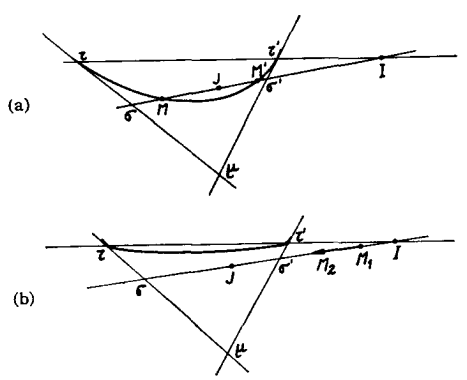

FIG. 14. Each point M of a complex segment $[\sigma\sigma']$ belongs to some disk $D[\mu,\tau\frown\tau']_p$. (a) M is real. (b) M is complex.

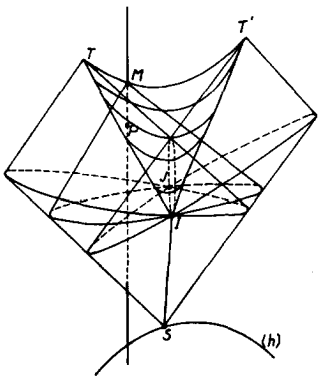

FIG. 15. $h(R)$ is "space-bounded." The figure represents the situation in the 3-dimensional plane defined by (h) and the semi-straight line MP. The double light cones $\langle ST\rangle$, $\langle ST'\rangle$, $\langle SM\rangle$ are shown; note that the ellipses edging these cones contain I and J. P is the point of intersection of the semi-straight line issued from M under consideration, with the region $\langle STT'\rangle$; here, P turns out to be on the past light cone of T. The "re-entrant nose" of $\langle STT'\rangle$ is removed in the $\mathfrak{h}$ process through the adding of region $(IJT\frown T')$.

(b) Complex points of $[\sigma\sigma']$: Call $\mathbf{M}=\mathbf{M}_1+i\mathbf{M}_2$ a complex point of $[\sigma\sigma']$ [cf. Fig. 14(b)]. As can be easily shown, for the one-parameter set of ellipses of the above-defined family which contain M (and its complex conjugate), the points I and J verify:

$$\mathbf{M}_1\mathbf{I}\cdot\mathbf{M}_1\mathbf{J}=-\mathbf{M}_2^2,$$

which fixes I and J unambiguously. To each couple τ, τ', such that the straight line $\tau\tau'$ contains I, there corresponds one ellipse of this particular set and τ, τ' may be chosen arbitrarily close to σ and σ', respectively. The corresponding elliptic complex segment $[\tau\frown\tau']_p$ fulfills the desired condition.

Q. E. D.

APPENDIX B

Proof that $h(R)$ is "space-bounded": The domain R of real Q space is bounded by two space-like surfaces, R_+ and R_-. $h(R)$ is deduced from it by applying the $\mathfrak{h}$ process as described in the text. The $\mathfrak{h}$ process yields two sorts of new points, according to whether they belong to future or past branches of doubly inadmissible hyperbolas; these new points are denoted M_+, M_-, respectively. We are going to show that, for any M_+ (respectively M_-), all semi-straight lines inside the past half light cone (respectively future half light cone) intersect R_+ (respectively R_-), and that the whole segment of this line between M_+ and the point of

intersection belongs to $h(R)$. Clearly, this property implies that $h(R)$ is "space-bounded."

Consider one of these new points, a point M_+ say. From now on, we drop the subscript $+$. By hypothesis, M is on a segment $T\frown T'$ of the future branch of a certain doubly inadmissible hyperbola (h), such that T, $T'\in R$. Call S, a point $\in R$ of the past branch. Since R is "space-bounded," the region

$$\langle STT'\rangle\equiv\langle ST\rangle\bigcup\langle ST'\rangle$$

is entirely in R. Therefore, it suffices to show that (a) each semi-straight line inside the past half light cone of M crosses $\langle STT'\rangle$ at a certain point P. (b) each point on the straight segment MP belongs to $h(R)$.

This problem is easily solved with the help of simple geometric arguments, in the 3-dimensional subspace containing (h) and the semi-straight line under consideration (cf. Fig. 15). Call I and J the two points of this subspace which are light-like distant from the points of (h). [I and J are symmetrical of each other with respect to the plane of (h); the middle of IJ is the center O of (h)]. The ellipses edging the double cones $\langle ST\rangle$, $\langle ST'\rangle$, $\langle SM\rangle$ all contain I and J. As a consequence (cf. Fig. 15), the ellipse edging $\langle SM\rangle$ is entirely inside $\langle STT'\rangle$. Point (i) follows from this by inspection.

Further inspection shows that MP is inside the region $(IJT\frown T')$ bounded by the triangles IJT, IJT' and by the conical surfaces described by the straight segments IQ, JQ when Q describes the hyperbolic segment $T\frown T'$. In order to prove (b), we show that this whole region belongs to $h(R)$.

To this effect, we note that the triangles IJS, IJT, IJT' are all in R. Now a homotecy of center O and coefficient λ, $(0<\lambda<1)$ followed by a homotecy of center I or J and coefficient μ, $(0<\mu<1)$, keep the points S, T, T' inside these triangles, respectively; therefore, the double-inadmissibility of (h) is conserved in any such transformation, and $T\frown T'$ goes into another hyperbolic segment belonging to $h(R)$.

Since the whole region $(IJT\frown T')$ can be described by performing such double homotecies on $T\frown T'$, it belongs to $h(R)$.

Q. E. D.

APPENDIX C[21]

Proof that $R_5{}^\dagger=h^3(R_5)$. First note that $h(\bar{R}_5)=\bar{h}(R_5)$. The proof of this point is left to the reader. Now, since the $\bar{h}$ process can be iterated, it follows that $h^3(\bar{R}_5)=\bar{h}^3(R_5)$. Therefore, the relation that we want to prove is equivalent to

$$\bar{R}_5{}^\dagger=h^3(\bar{R}_5).$$

Now, according to a remark which was made in Sec. IV, any point in $h^3(\bar{R}_5)$ necessarily belongs to $\bar{R}_5{}^\dagger$. Thus, it remains to show that, conversely, any point of $\bar{R}_5{}^\dagger$ belongs to $h^3(\bar{R}_5)$.

1. Let us take $M\in\bar{R}_5{}^\dagger$, therefore inside $\mathcal{E}_5$; the ascendant vertical of M intersects the boundary $\partial\mathcal{E}_5$ of $\mathcal{E}_5$ at some point N above $\mathrm{Re}P_5$.

Consider the supporting plane π[22] which contains N; we shall first prove that: (a) π intersects $\mathrm{Re}P_5$ along a two-sheeted hyperboloid H. (b) H is doubly inadmissible with respect to $\bar{R}_5$.

From the definition of π,[22] N lies inside the convex hull $\mathcal{E}_4$ of $\pi\bigcap\bar{R}_5$. Thus, N is the center of mass of some discrete positive distribution supported by points of $\pi\bigcap\bar{R}_5$. These points cannot all be space-like distant from one another, since N is above $\mathrm{Re}P_5$; therefore, there exist at least two points Σ^+, Σ^- in $\pi\bigcap\bar{R}_5$ which are time like distant.

Consider now the "double cone" $\langle\Sigma^+\Sigma^-\rangle$ on $\mathrm{Re}P_5$, which, in view of the spaceboundedness of R_5, lies entirely in $\bar{R}_5$. As π is a supporting plane, it leaves $\bar{R}_5$ entirely in one of the closed half-spaces which it bounds; therefore, π cannot intersect the open "double cone" $\langle\Sigma^+\Sigma^-\rangle$. Thus, the hyperboloid $H\equiv\pi\bigcap\mathrm{Re}P_5$ on which the points Σ^+, Σ^- lie, does not intersect the open "double cone" $\langle\Sigma^+\Sigma^-\rangle$ either. Here a little geometry in Q-space projection easily shows that H is necessarily two-sheeted, or as a limiting case, degenerates into a light cone.

Clearly, the time-like distant points Σ^+, Σ^- lie on different sheets, so that H is doubly inadmissible with respect to $\bar{R}_5$.

2. Still considering N as a center of mass, we may write

$$\mathbf{N}=\sum_i\lambda_i{}^+\mathbf{N}_i{}^+ +\sum_j\lambda_j{}^-\mathbf{N}_j{}^-,$$

$$\lambda_i{}^+\geqslant 0,\quad \lambda_j{}^-\geqslant 0,\quad \sum_{i,j}(\lambda_i{}^+ +\lambda_j{}^-)=1,$$

where the i's and j's form two finite sets of discrete indices.

The $N_i{}^+$ (resp. $N_j{}^-$) are points of $\bar{R}_5$ belonging to the future (resp. past) sheet of H; none of these two sets can be empty, since they each contain at least Σ^+ or Σ^-.

Let us call N^+, (N^-), the center of mass relative to the points $N_i{}^+$, $(N_j{}^-)$ with masses $\lambda_i{}^+$, $(\lambda_j{}^-)$; N^+ and N^- lie, respectively, inside the convex regions bounded by the corresponding sheets (i.e., below $\mathrm{Re}P_5$); therefore the segment N^+N^- is time-like and intersects H at two points P^+, P^- on different sheets. By construction, P^+, P^-, belong to $R_5{}^\dagger$ and the over-all center of mass N lies on the time-like segment P^+P^-. It is then easy to see, by projecting the figure in Q space, that M belongs to the "double-cone" $\langle P^+P^-\rangle$.

Now, if we can prove that the identity

$$\pi\bigcap\bar{R}_5{}^\dagger=h^3(\pi\bigcap\bar{R}_5)\qquad(C.1)$$

holds true, then the announced result will follow. Indeed, P^+, P^- will then lie in $h^3(\bar{R}_5)$, and also the whole

[21] In this appendix, the notation $\bar{S}$ means, for any set S, the closure of S.

[22] Bonnesen-Fenchel, *Konvexe Körper* (Chelsea Publishing Company, New York, 1948).

double cone $\langle P^+P^-\rangle$, since the iterated regions $h^n(R_5)$ are spacebounded (cf. Appendix B). Thus the point M will belong to $h^3(\bar{R}_5)$.

3. We proceed to prove (C.1). Call H a two sheeted hyperboloid in n-dimensional space

$$H: x^2 \equiv x^{(0)2} - \sum_{(i)=1}^{n-1} x^{(i)2} = \kappa^2, \quad \kappa^2 > 0.$$

H^+ and H^- are the future ($x^{(0)}>0$) and past ($x^{(0)}<0$) sheets of H and $\mathcal{K}_0$ the region of space $x^2 \leqslant \kappa^2$.

Let A be a bounded closed region on H which can be written $A \equiv A^+ \bigcup A^-$ with $A^\pm \subset H^\pm$, A^+ and A^- being assumed non empty; $\mathcal{E}(A)$ denotes the convex hull of A.

We define the "l_0 completion" of any set S in n-dimensional space as the completion of S by all the linear segments which lie entirely in $\mathcal{K}_0$ and have their end points in S; the completed region is noted $l_0(S)$. Then,

$$\mathcal{E}_0(A) \equiv \mathcal{E}(A) \bigcap \mathcal{K}_0 = l_0{}^n(A). \tag{C.2}$$

This is proved by induction as follows.

(a) It is clear by inspection that the property holds true when $n=2$.

(b) Suppose it holds in $(n-1)$-dimensional space, and consider the n-dimensional case ($n \geqslant 3$).

If $m \in l_0{}^n(A)$, then $m \in \mathcal{E}_0(A)$ in view of the barycentric construction of $\mathcal{E}(A)$.

Conversely, assume $m \in \mathcal{E}_0(A)$. Draw through m a straight line which lies entirely in $\mathcal{K}_0$. It intersects $\partial \mathcal{E}_0(A)$ at two points n_1, n_2, one on each side of m.

Consider the supporting planes ϖ_i at n_i ($i=1, 2$); they have the property[22]:

$$\varpi_i \bigcap \mathcal{E}(A) = \mathcal{E}(\varpi_i \bigcap A),$$

and thus $n_i \in \mathcal{E}(\varpi_i \bigcap A)$. As n_i lies in $\mathcal{K}_0$, it is then necessary that $\varpi_i \bigcap A$ have points on both sheets of H; it means that $\varpi_i \bigcap H$ is a two sheeted hyperboloid in $(n-1)$-dimensional space for which the theorem is assumed to hold; thus

$$n_i \in l_0{}^{n-1}(\varpi_i \bigcap A) \subset l_0{}^{n-1}(A).$$

As the segment $n_1 n_2$ lies entirely in $\mathcal{K}_0$, we have by definition of l_0:

$$m \in l_0{}^n(A).$$

This completes the proof of relation (C.2).

Consider now the $\mathfrak{h}$ completion of A on the hyperboloid H; it is clear by inspection that

$$l_0(h(A)) = l_0{}^2(A). \tag{C.3}$$

By application of (C.2) and iteration of (C.3), we have

$$\mathcal{E}_0(A) = l_0 h^{n-1}(A), \tag{C.4}$$

from which we deduce

$$A^\dagger \equiv \mathcal{E}(A) \bigcap H = \mathcal{E}_0(A) \bigcap H = h^{n-1}(A), \tag{C.5}$$

since the last operation l_0 occurring on the right-hand side of Eq. (C.4) cannot add points belonging to H.

Equation (C.5) reduces to (C.1) if we take for H the two-sheeted hyperboloid of points **1** and **2**, ($n=4$), together with $A = \pi \bigcap \bar{R}_5$ and $A^\dagger = \pi \bigcap R_5{}^\dagger$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER-OCTOBER, 1961

Role of the Asymptotic Condition in a Lagrangian Field Theory*

M. WELLNER and R. B. CURTIS
Indiana University, Bloomington, Indiana

(Received March 3, 1961)

The scattering operator of Feynman and Dyson for a self-interacting neutral scalar field is derived from a Lagrangian without the use of a canonical transformation between the Heisenberg and interaction pictures.

1. INTRODUCTION

IN the traditional formulation of quantum field theory, the dynamical behavior is specified by means of a Lagrangian. For simplicity let us restrict this discussion to the case of a single self-interacting neutral scalar field, denoted by $\phi(x)$ in the Heisenberg picture. Assuming the absence of bound states, we are led in the conventional theory to the formal expression[1]

$$S = T \exp i J_I\{A\} \tag{1.1}$$

for the scattering operator S. Time ordering is denoted by T; $A(x)$ is a free neutral scalar field with the mass m (the observable mass of the field ϕ); the functional $J_I\{\phi\}$ is what remains of the total action integral when the free-field part has been subtracted. The quantity $J_I\{A\}$ results from the substitution $\phi \to A$ in $J_I\{\phi\}$. The operator S is such that the transition amplitude $S_{\beta\alpha}$ from the state α to the state β after a long enough time is given by

$$S_{\beta\alpha} = \langle \beta | S | \alpha \rangle. \tag{1.2}$$

If A_+ (A_-) is the creation (annihilation) part of

$$A = A_+ + A_-, \tag{1.3}$$

then $|\alpha\rangle$ and $|\beta\rangle$ are constructed by repeated application

* Work supported in part by the National Science Foundation.

[1] See, for example, J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955), or N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959).

double cone $\langle P^+P^-\rangle$, since the iterated regions $h^n(R_5)$ are spacebounded (cf. Appendix B). Thus the point M will belong to $h^3(\bar{R}_5)$.

3. We proceed to prove (C.1). Call H a two sheeted hyperboloid in n-dimensional space

$$H: x^2 \equiv x^{(0)2} - \sum_{(i)=1}^{n-1} x^{(i)2} = \kappa^2, \quad \kappa^2 > 0.$$

H^+ and H^- are the future ($x^{(0)}>0$) and past ($x^{(0)}<0$) sheets of H and $\mathcal{K}_0$ the region of space $x^2 \leqslant \kappa^2$.

Let A be a bounded closed region on H which can be written $A \equiv A^+ \bigcup A^-$ with $A^\pm \subset H^\pm$, A^+ and A^- being assumed non empty; $\mathcal{E}(A)$ denotes the convex hull of A.

We define the "l_0 completion" of any set S in n-dimensional space as the completion of S by all the linear segments which lie entirely in $\mathcal{K}_0$ and have their end points in S; the completed region is noted $l_0(S)$. Then,

$$\mathcal{E}_0(A) \equiv \mathcal{E}(A) \bigcap \mathcal{K}_0 = l_0{}^n(A). \tag{C.2}$$

This is proved by induction as follows.

(a) It is clear by inspection that the property holds true when $n=2$.

(b) Suppose it holds in $(n-1)$-dimensional space, and consider the n-dimensional case ($n \geqslant 3$).

If $m \in l_0{}^n(A)$, then $m \in \mathcal{E}_0(A)$ in view of the barycentric construction of $\mathcal{E}(A)$.

Conversely, assume $m \in \mathcal{E}_0(A)$. Draw through m a straight line which lies entirely in $\mathcal{K}_0$. It intersects $\partial \mathcal{E}_0(A)$ at two points n_1, n_2, one on each side of m.

Consider the supporting planes ϖ_i at n_i ($i=1, 2$); they have the property[22]:

$$\varpi_i \bigcap \mathcal{E}(A) = \mathcal{E}(\varpi_i \bigcap A),$$

and thus $n_i \in \mathcal{E}(\varpi_i \bigcap A)$. As n_i lies in $\mathcal{K}_0$, it is then necessary that $\varpi_i \bigcap A$ have points on both sheets of H; it means that $\varpi_i \bigcap H$ is a two sheeted hyperboloid in $(n-1)$-dimensional space for which the theorem is assumed to hold; thus

$$n_i \in l_0{}^{n-1}(\varpi_i \bigcap A) \subset l_0{}^{n-1}(A).$$

As the segment $n_1 n_2$ lies entirely in $\mathcal{K}_0$, we have by definition of l_0:

$$m \in l_0{}^n(A).$$

This completes the proof of relation (C.2).

Consider now the $\mathfrak{h}$ completion of A on the hyperboloid H; it is clear by inspection that

$$l_0(h(A)) = l_0{}^2(A). \tag{C.3}$$

By application of (C.2) and iteration of (C.3), we have

$$\mathcal{E}_0(A) = l_0 h^{n-1}(A), \tag{C.4}$$

from which we deduce

$$A^\dagger \equiv \mathcal{E}(A) \bigcap H = \mathcal{E}_0(A) \bigcap H = h^{n-1}(A), \tag{C.5}$$

since the last operation l_0 occurring on the right-hand side of Eq. (C.4) cannot add points belonging to H.

Equation (C.5) reduces to (C.1) if we take for H the two-sheeted hyperboloid of points **1** and **2**, ($n=4$), together with $A = \pi \bigcap \bar{R}_5$ and $A^\dagger = \pi \bigcap R_5{}^\dagger$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER-OCTOBER, 1961

Role of the Asymptotic Condition in a Lagrangian Field Theory*

M. WELLNER and R. B. CURTIS
Indiana University, Bloomington, Indiana

(Received March 3, 1961)

The scattering operator of Feynman and Dyson for a self-interacting neutral scalar field is derived from a Lagrangian without the use of a canonical transformation between the Heisenberg and interaction pictures.

1. INTRODUCTION

IN the traditional formulation of quantum field theory, the dynamical behavior is specified by means of a Lagrangian. For simplicity let us restrict this discussion to the case of a single self-interacting neutral scalar field, denoted by $\phi(x)$ in the Heisenberg picture. Assuming the absence of bound states, we are led in the conventional theory to the formal expression[1]

$$S = T \exp i J_I\{A\} \tag{1.1}$$

for the scattering operator S. Time ordering is denoted by T; $A(x)$ is a free neutral scalar field with the mass m (the observable mass of the field ϕ); the functional $J_I\{\phi\}$ is what remains of the total action integral when the free-field part has been subtracted. The quantity $J_I\{A\}$ results from the substitution $\phi \to A$ in $J_I\{\phi\}$. The operator S is such that the transition amplitude $S_{\beta\alpha}$ from the state α to the state β after a long enough time is given by

$$S_{\beta\alpha} = \langle \beta | S | \alpha \rangle. \tag{1.2}$$

If A_+ (A_-) is the creation (annihilation) part of

$$A = A_+ + A_-, \tag{1.3}$$

then $|\alpha\rangle$ and $|\beta\rangle$ are constructed by repeated application

* Work supported in part by the National Science Foundation.

[1] See, for example, J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955), or N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959).

of A_+ on a vacuum, $|\text{vac}\rangle$, with the properties

$$\langle\text{vac}|\text{vac}\rangle=1 \tag{1.4}$$

and

$$A_-(x)|\text{vac}\rangle=0. \tag{1.5}$$

The existing derivations of (1.1) are objectionable because they make use of a canonical transformation between ϕ and A (i.e., between the Heisenberg and interaction pictures). Such a transformation does not exist in a strict mathematical sense[2] (except in the trivial case of a noninteracting ϕ) unless one resorts to the unphysical device of switching the interaction on and off adiabatically in the Heisenberg picture.[3] It is the purpose of this article to derive (1.1) without assuming any transformation between A and ϕ.

As is well known from perturbation theory, (1.1) is not mathematically defined unless one inserts a convergence factor, namely the above-mentioned switching, in the integrand of J_I, so that contributions are essentially limited to finite times. This factor is a peculiarity of the form (1.1) for the scattering operator. It does not occur in the physical formulation of the theory and need not occur in the calculation of physical processes. For example, all the perturbation calculations can be done without it. The reason for its introduction will become clear in the course of the derivation.

2. THE REDUCTION FORMULA

As the starting point of our derivation we shall use, instead of a canonical transformation, the prescription ("reduction formula") of Lehmann, Symanzik, and Zimmermann[4] for the operator S in terms of the vacuum expectation values of time-ordered products of $\phi(x)$. The formula is not explicitly dependent on any Lagrangian and has been derived by the above-named authors from their asymptotic condition. The result is[5]

$$S=1+\sum_{n=1}^{\infty}(n!)^{-1}\int dx_1\cdots\int dx_n:A(x_1)\cdots A(x_n):\\ iK_1\cdots iK_n\tau_n(x_1,\cdots,x_n). \tag{2.1}$$

The colons indicate Wick's normal ordering, and each integral is over space and time. The K_j are Klein-Gordon operators[6]:

$$K_j=\Box_j+m^2,\quad \Box=\partial_\mu\partial^\mu. \tag{2.2}$$

The τ_n are defined by

$$\left.\begin{aligned}&\tau_0=\langle 0|0\rangle=1,\\ &\tau_n(x_1,\cdots,x_n)=\langle 0|T(\phi(x_1)\cdots\phi(x_n))|0\rangle\quad(n\geq 1).\end{aligned}\right\} \tag{2.3}$$

Here $|0\rangle$ is the physical vacuum; it bears no relation to the vector $|\text{vac}\rangle$ of Sec. 1. The Heisenberg field ϕ is renormalized in that it satisfies the asymptotic condition.

3. RESULTS OF MATTHEWS, SALAM, AND COESTER

From the canonical commutation rules and the field equations, Matthews and Salam have derived a set of relations[7] among the τ functions. These relations can be put into compact form if, following Coester,[8] we introduce the set

$$\tau=\{\tau_0,\ \tau_1(x_1),\ \tau_2(x_1,x_2),\ \cdots\}, \tag{3.1}$$

which can be considered as representing a vector in a space rather similar to Fock space. An annihilation operator $\Phi^-(x)$ is defined in that space by specifying a new set $\Phi^-(x)\tau$, such that

$$(\Phi^-(x)\tau)_{n-1}(x_1,\cdots,x_{n-1})=\tau_n(x_1,\cdots,x_{n-1},x). \tag{3.2}$$

Similarly, let Φ_+ be a creation operator such that

$$\left.\begin{aligned}&(\Phi_+(x)\tau)_0=0,\\ &(\Phi_+(x)\tau)_{n+1}=\sum_{j=1}^{n+1}\delta(x_j-x)\\ &\qquad\times\tau_n(x_1,\cdots,x_{j-1},x_{j+1},\cdots,x_{n+1})\\ &\qquad\qquad(n\geq 0).\end{aligned}\right\} \tag{3.3}$$

The notation Φ^+ is reserved for later use. From these definitions the commutation rules

$$\left.\begin{aligned}&[\Phi^-(x),\Phi^-(y)]=[\Phi_+(x),\Phi_+(y)]=0,\\ &[\Phi^-(x),\Phi_+(y)]=\delta(x-y)\end{aligned}\right\} \tag{3.4}$$

follow simply.

Let J_I be the functional defined in Sec. 1. More explicitly, if

$$\mathfrak{L}_0\{\phi\}=\tfrac{1}{2}(\partial\phi)^2-\tfrac{1}{2}m^2\phi^2 \tag{3.5}$$

is the Lagrangian density for the free field, while

$$\mathfrak{L}=\mathfrak{L}_0+\mathfrak{L}_I \tag{3.6}$$

is the corresponding quantity for the interacting field, then we define the formal action integrals over all space-time:

$$\left.\begin{aligned}&J\{\phi\}=\int\mathfrak{L}\{\phi\}dx,\\ &J_0\{\phi\}=\int\mathfrak{L}_0\{\phi\}dx,\\ &J_I\{\phi\}=\int\mathfrak{L}_I\{\phi\}dx.\end{aligned}\right\} \tag{3.7}$$

It is convenient to think of $\mathfrak{L}$ as *not* being normally ordered, since the concept of normal ordering becomes

[2] R. Haag, Dan. Mat.-Fys. Medd. **29**, No. 12 (1955); O. W. Greenberg, Phys. Rev. **115**, 706 (1959).

[3] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), p. 194.

[4] H. Lehmann, K. Symanzik, and W. Zimmermann, Nuovo cimento **1**, 205 (1955).

[5] F. Coester, unpublished notes. Equation (2.1) is most simply compared to the formulation of reference 4 by noting that the matrix elements $\langle\beta|S|\alpha\rangle$ constructed from (2.1) satisfy the same recursion relations.

[6] We use the metric $g^{\mu\nu}=\text{diag}(1,-1,-1,-1)$, $(\mu,\nu=0,1,2,3)$.

[7] P. T. Matthews and A. Salam, Proc. Roy. Soc. (London) **A221**, 128 (1953).

[8] F. Coester, Phys. Rev. **95**, 1318 (1954).

more involved for an interacting field. The same purpose may be accomplished by including appropriate counter terms in $\mathfrak{L}$. We shall not be concerned here with the finiteness of these counter terms or of any other renormalization constants.

The equations of Matthews and Salam[7] for the τ functions are in this notation

$$\frac{\delta J\{\Phi^-\}}{\delta\Phi^-(x)}\tau=i\Phi_+(x)\tau \tag{3.8}$$

(at least if $\mathfrak{L}_I$ is a polynomial in ϕ).

The formal solution of (3.8) with "causal" boundary conditions has been given by Coester[8] as

$$\tau=(\exp iJ_I\{\Phi^-\})[\exp(-iJ_0\{\Phi^+\})]\Omega, \tag{3.9}$$

where

$$\Phi^+(x)=i\int dy\Delta_c(x-y)\Phi_+(y), \tag{3.10}$$

and where Ω is the "vacuum vector" defined by

$$\Omega_0=1, \quad \Omega_n=0 \quad (n\geqslant 1). \tag{3.11}$$

The causal propagator Δ_c is given by[9]

$$\Delta_c(x)=(2\pi)^{-4}\int dp e^{-ip\cdot x}(p^2-m^2+i\epsilon)^{-1}. \tag{3.12}$$

The splitting of J into J_0 and J_I in (3.9) ensures the correct free-field limit for the τ functions. The additive constant in $\mathfrak{L}$ is no longer arbitrary but must be adjusted so that $\tau_0=1$.

From the definition of Φ^+ and the identity

$$\int dz\left\{\left[\frac{\partial}{\partial z_\mu}\Delta_c(x-z)\right]\left[\frac{\partial}{\partial z^\mu}\Delta_c(y-z)\right]-m^2\Delta_c(x-z)\Delta_c(y-z)\right\}=\Delta_c(x-y), \tag{3.13}$$

it follows that

$$J_0\{\Phi^+\}=\tfrac{1}{2}\int dx\int dy\Phi_+(x)\Delta_c(x-y)\Phi_+(y). \tag{3.14}$$

The τ functions, and hence the scattering operator, can be unambiguously computed from (3.8) or (3.9) by perturbation theory, in spite of the fact that we never introduce any adiabatic switching. This is connected with the fact that these relations, unlike Eq. (1.1), do not restrict the τ_n to the mass shell. Hence, for example, the mass and wave function renormalizations can be carried out by applying Lehmann's results[10] for the spectrum of $\tau_2(x_1,x_2)$. The two relevant counter terms are adjusted so as to (a) eliminate the term in $(p^2-m^2+i\epsilon)^{-2}$ (mass renormalization), and (b) produce the correct residue for the term in $(p^2-m^2+i\epsilon)^{-1}$ (wave-function renormalization).

We close this summary by defining the Feynman amplitudes φ_n, a linear combination of the τ_n. The set φ is given by

$$\varphi=(\exp iJ_0\{\Phi^+\})\tau, \tag{3.15}$$

or, from (3.9) and with use of (I.5),[11]

$$\varphi=(\exp iJ_I\{\Phi\})\Omega, \tag{3.16}$$

where

$$\Phi=\Phi^-+\Phi^+. \tag{3.17}$$

Since Δ_c is an even function, we have

$$[\Phi(x),\Phi(y)]=0. \tag{3.18}$$

4. MODIFIED FORM OF THE REDUCTION FORMULA

We note that the τ_n and the φ_n may be used interchangeably in (2.1). Indeed, consider the exponential in (3.15) to be expanded in a power series. The difference between φ_n and $\tau_n(n\geqslant 1)$ is seen to be a sum of terms of the type

$$\Delta_c(x_j-x_k)\sigma, \tag{4.1}$$

where σ contains neither x_j nor x_k. Applying $\cdots iK_j\cdots iK_k\cdots$ as in (2.1), (4.1) becomes

$$[K_jK_k\Delta_c(x_j-x_k)]\sigma'=-[K_j\delta(x_j-x_k)]\sigma' \tag{4.2}$$

(σ' independent of x_j and x_k). Multiplying by

$$:\cdots A(x_j)\cdots A(x_k)\cdots:$$

and integrating over all variables as in (2.1), we find that the result vanishes because $KA=0$.

For our purpose, a more useful form of (2.1) can be obtained by defining a scalar product (f,g) between two sets of functions f_n, g_n:

$$(f,g)=(g,f)=f_0g_0+\sum_{n=1}^{\infty}(n!)^{-1}\int dx_1\cdots\int dx_n\times f_n(x_1,\cdots,x_n)g_n(x_1,\cdots,x_n). \tag{4.3}$$

Setting

$$\mathfrak{N}_n(x_1,\cdots,x_n)=:A(x_1)\cdots A(x_n): \tag{4.4}$$

and defining the operator $[iK]$ such that

$$\left.\begin{aligned}([iK]\tau)_0&=\tau_0,\\ ([iK]\tau)_n(x_1,\cdots,x_n)&=iK_1\cdots iK_n\tau_n(x_1,\cdots,x_n)\\ &\qquad(n\geqslant 1),\end{aligned}\right\} \tag{4.5}$$

we have for (2.1)

$$S=(\mathfrak{N},[iK]\tau), \tag{4.6}$$

or equivalently,

$$S=(\mathfrak{N},[iK]\varphi) \tag{4.7}$$

(keeping in mind that $\tau_0=\varphi_0=1$).

[9] The homogeneous Δ functions, to be used later, have such a sign that $\Delta_c(x)=-\Delta_+(x)$ $(x^0>0)$.

[10] H. Lehmann, Nuovo cimento **11**, 345 (1954).

[11] Equations numbered by Roman numerals are to be found in the appropriate appendix.

It is readily verified that Φ_+ and Φ^- are each other's adjoints under the scalar product (4.3):

$$(f, \Phi^-(x)g) = (\Phi_+(x)f, g). \tag{4.8}$$

5. PROOF OF EQUIVALENCE TO THE FEYNMAN-DYSON SCATTERING OPERATOR

We now show that (1.1) can be derived from (4.7) and (3.16). First we make use of the inverse of Wick's expansion (see Appendix III) to re-express the set $\mathfrak{N}$ [Eq. (4.4)] in terms of the set $\mathcal{T}$ of time-ordered products

$$\mathcal{T}_n(x_1, \cdots, x_n) = T(A(x_1) \cdots A(x_n)). \tag{5.1}$$

We have

$$\mathfrak{N} = (\exp iJ_0\{\Phi^+\})\mathcal{T}. \tag{5.2}$$

For any function $f(x)$, the set

$$F_n(x_1, \cdots, x_n) = f(x_1)f(x_2) \cdots f(x_n) \tag{5.3}$$

can be represented by

$$F = \left[\exp \int dx f(x)\Phi_+(x)\right]\Omega. \tag{5.4}$$

Since $A(x)$ behaves like a c number under the time-ordering symbol T, we can write, keeping (3.4) in mind,

$$\mathfrak{N} = T\left[\exp\left(iJ_0\{\Phi^+\} + \int dx A(x)\Phi_+(x)\right)\right]\Omega. \tag{5.5}$$

Inserting this result into (4.7) and making use of the hermiticity property (4.8), we obtain

$$S = T\left(\Omega, \left[\exp\left(iJ_0\{\Phi_-\} + \int dx A(x)\Phi^-(x)\right)\right][iK]\varphi\right), \tag{5.6}$$

where

$$\Phi_-(x) = i\int dy \Delta_c(x-y)\Phi^-(y). \tag{5.7}$$

At this stage the adiabatic switching must be introduced. In passing from (4.7) to (5.6) we have interchanged the order of two sets, (a) and (b), of integrations over all time. The first, (a), is the single integration which occurs in the exponent $iJ_I\{\Phi\}$ of φ; the second, (b), consists of the integrations involved in taking a scalar product of the type (f,g). It is clear that (a) is performed before (b) in (4.7), while in order to reach (5.6) via the hermiticity property we had to perform (b) before (a). The adiabatic factor makes this interchange harmless.

With use of (II.5), we have

$$(\exp iJ_0\{\Phi_-\})[iK]\varphi = [iK](\exp iJ_0\{\Phi^-\})\varphi, \tag{5.8}$$

$$= [iK](\exp iJ_I\{\Phi^+\})\Omega, \tag{5.9}$$

$$= (\exp iJ_I\{\Phi_+\})\Omega, \tag{5.10}$$

where (5.9) involves the use of (3.16) and (I.5), and where (5.10) is obtained from (II.6), bearing in mind that

$$[iK]\Omega = \Omega. \tag{5.11}$$

Equation (5.6) now becomes

$$S = T\left(\Omega, \left[\exp \int dx A(x)\Phi^-(x) \exp iJ_I\{\Phi_+\}\right]\Omega\right), \tag{5.12}$$

$$= T(\Omega, [\exp iJ_I\{\Phi_+ + A\}]\Omega) \tag{5.13}$$

by (I.6).

Thus

$$S = T \exp iJ_I\{A\}, \tag{5.14}$$

which was to be shown [cf. Eq. (1.1)].

ACKNOWLEDGMENT

One of us (M. W.) takes pleasure in thanking Professor J. R. Oppenheimer for his hospitality at the Institute for Advanced Study, where part of this work was carried out.

APPENDIX I. SOME PROPERTIES OF THE OPERATORS $\Phi^{\pm}$

From the commutation rules (3.4) it follows for any functional F that

$$[F\{\Phi^-\}, \Phi_+(x)] = \delta F\{\Phi^-\}/\delta\Phi^-(x). \tag{I.1}$$

Therefore, if

$$F\{\Phi^-\} = \exp\tfrac{1}{2}\int dx \int dy \Phi^-(x)\mathfrak{D}(x-y)\Phi^-(y) \tag{I.2}$$

for some symmetric kernel $\mathfrak{D}$, then

$$F\{\Phi^-\}\Phi_+(x) = \left[\Phi_+(x) + \int dy \mathfrak{D}(x-y)\Phi^-(y)\right]F\{\Phi^-\}. \tag{I.3}$$

More generally, for any $G\{\Phi_+, \Phi^-\}$ which can be expanded as a Volterra series (the functional analog of a Taylor series) in Φ_+, we have

$$F\{\Phi^-\}G\{\Phi_+(x), \Phi^-\} = G\left\{\Phi_+(x) + \int dy \mathfrak{D}(x-y)\Phi^-(y), \Phi^-\right\}F\{\Phi^-\}, \tag{I.4}$$

x being a dummy variable.

In an entirely similar fashion, it follows that

$$F\{\Phi_+\}G\{\Phi_+, \Phi^-(x)\} = G\left\{\Phi_+, \Phi^-(x) - \int dy \mathfrak{D}(x-y)\Phi_+(y)\right\}F\{\Phi_+\}. \tag{I.5}$$

Finally we note that, for any function $f(x)$ and functional H,

$$\left[\exp\int dx f(x)\Phi^-(x)\right]H\{\Phi_+\}$$

$$=H\{\Phi_+ + f\}\exp\int dx f(x)\Phi^-(x). \quad \text{(I.6)}$$

APPENDIX II. SOME PROPERTIES OF THE OPERATOR $[iK]$

We next show for any functional G as in Appendix I, that

$$G\{\Phi_-\}[iK]\tau=[iK]G\{\Phi^-\}\tau, \quad \text{(II.1)}$$

where

$$\Phi_-(x)=i\int dy\Delta_c(x-y)\Phi^-(x). \quad \text{(II.2)}$$

In particular, with a series expansion in mind, it is enough to take $G\{\Phi_-\}$ of the form $\Phi_-(y_1)\cdots\Phi_-(y_s)$. Then (II.1) is merely a rewriting of Zimmermann's result,[12]

$$\tau_n(x_1,\cdots,x_n)=-\int dx\Delta_c(x_n-x)$$

$$\times K_x\tau_n(x_1,\cdots,x_{n-1},x). \quad \text{(II.3)}$$

Since the φ_n are linear combinations of the τ_n with coefficients of the form $\Delta_c(x_j-x_k)\cdots\Delta_c(x_l-x_m)$, all the x's bearing different labels, it follows that (II.3) is equally valid for the φ_n:

$$\varphi_n(x_1,\cdots,x_n)$$

$$=-\int dx\Delta_c(x_n-x)K_x\varphi_n(x_1,\cdots,x_{n-1},x). \quad \text{(II.4)}$$

We have, corresponding to (II.1),

$$G\{\Phi_-\}[iK]\varphi=[iK]G\{\Phi^-\}\varphi. \quad \text{(II.5)}$$

[12] W. Zimmermann, Nuovo cimento **13**, 503 (1959), Eq. (25). This theorem follows from the asymptotic condition.

We finally note that, from an explicit use of the definitions,

$$[iK]G\{\Phi^+\}=G\{\Phi_+\}[iK]. \quad \text{(II.6)}$$

APPENDIX III. WICK'S EXPANSION OF TIME-ORDERED PRODUCTS[1]

We prove the relation

$$\mathcal{T}=[\exp(-iJ_0\{\Phi^+\})]\mathfrak{N} \quad \text{(III.1)}$$

or, in component form,

$$\mathcal{T}_n(x_1,\cdots,x_n)=\sum_{x\to y;n}\sum_{s=0}^{\leq n/2}\left[\prod_{j=1}^{s}i\Delta_c(y_{2j-1}-y_{2j})\right]$$

$$\times\mathfrak{N}_{n-2s}(y_{2s+1},\cdots,y_n). \quad \text{(III.2)}$$

The notation $x\to y;n$ refers to all the permutations $y_1,\cdots,y_n$ of $x_1,\cdots,x_n$ which give different expressions for the summand. By convention the product $\prod_{j=1}^{0}$ is unity.

Equation (III.2) is trivial for $n=0,1$. The general proof is by recursion. We consider $\mathcal{T}_{n+1}$ with

$$x_{n+1}{}^0\geqslant\cdots\geqslant x_1{}^0 \quad \text{(III.3)}$$

without loss of generality, since the $\mathcal{T}_n$ and the $\mathfrak{N}_n$ are symmetric in their variables. If we assume (III.2) for n, we have for $n+1$

$$\mathcal{T}_{n+1}=A(x_{n+1})\mathcal{T}_n \quad \text{(III.4)}$$

$$=\sum_{x\to y;n}\sum_{s=0}^{\leq n/2}\left[\prod_{j=1}^{s}i\Delta_c(y_{2j-1}-y_{2j})\right]$$

$$\{\mathfrak{N}_{n-2s+1}(y_{2s+1},\cdots,y_{n+1})$$

$$+[A_-(x_{n+1}),\mathfrak{N}_{n-2s}(y_{2s+1},\cdots,y_n)]\}. \quad \text{(III.5)}$$

Using

$$[A_-(x),\mathfrak{N}_k(x_1,\cdots,x_k)]=-\sum_{j=1}^{k}i\Delta_+(x-x_j)$$

$$\times\mathfrak{N}_{k-1}(x_1\cdots,x_{j-1},x_{j+1},\cdots,x_k) \quad \text{(III.6)}$$

and replacing Δ_+ by $-\Delta_c$ in view of (III.3), we obtain (III.2) with n replaced by $n+1$.

Acnodes and Cusps on Landau Curves

R. J. EDEN, P. V. LANDSHOFF, J. C. POLKINGHORNE, AND J. C. TAYLOR
Department of Applied Mathematics and Theoretical Physics, University of Cambridge

(Received May 2, 1961)

It is shown that the Landau curve for a reduced sixth-order diagram can acquire acnodes and real cusps as the masses are varied. They are associated with complex singularities that under certain conditions are in the physical sheet and cause a breakdown of the Mandelstam representation. The problem of obtaining general criteria for acnodes and cusps is discussed.

1. INTRODUCTION

IN previous papers[1,2] we studied the analytic properties of scattering amplitudes to all orders in perturbation theory. Under certain conditions, related to the masses involved and determined by the fourth-order graphs, we showed that the Mandelstam representation is true to every order of perturbation theory provided that the Landau curves have no acnodes lying in normal threshold cuts. (Acnodes are isolated points of real curves[3]: they are, of course, joined to the rest of the curve by complex surfaces.) The method was based on a consideration of how parts of a Landau curve can become singular on the physical sheet after the appearance of a crunode. (Crunodes are points where a real curve crosses itself.)

We previously believed[1] that acnodes only occurred when Landau curves were *accidentally* degenerate as a result of a fortuitous choice of the values of the masses. In fact, however, the leading curve of the Feynman diagram in Fig. 1 is found to have acnodes for a range

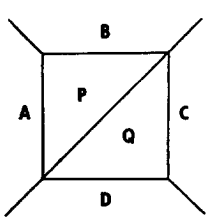

FIG. 1. The reduced Feynman graph under consideration.

of values of the masses. The proof of the Mandelstam representation in perturbation theory is, therefore, incomplete until it can be supplemented by a criterion excluding, for sufficiently small values of the external masses, the appearance of acnodes lying in normal threshold cuts.

We have not yet succeeded in finding such a general criterion. The conditions obtained from the graph of Fig. 1 involve relations between external and internal masses different from the fourth-order conditions and are even, in some instances, less stringent.

Part of the leading curve for Fig. 1 has been previously evaluated for the equal mass case[4] for which there are no acnodes. The amplitude then satisfies an ordinary Mandelstam representation and the boundary of the support of the spectral function is, as usual, a curve of the general shape of Fig. 2(a). When the external masses are increased, the singularities change in a complicated manner, that at first sight is paradoxical. The resolution of these complications involves some striking new features in the behavior of complex singularities which are discussed in Secs. 2–4 of this paper.[5]

As the external masses are increased, two acnodes appear. At the moment of appearance they coincide and then they separate. A further increase in the masses causes one of the acnodes to meet the curve that was the boundary of the support of the spectral function. This curve then develops two real cusps. The cusps may eventually project across the normal thresholds and the curve itself may lie outside both the normal threshold cuts, even for values of the masses that exclude the occurrence of anomalous thresholds for this diagram. Before the real cusps appear there are two complex cusps round which the Landau surface joining the acnodes to the curve is twisted. This twisting combined with a new aspect of singularities at cusps leads to a fully consistent picture.

A singular Landau curve is obtained when the Feynman integral for the amplitude has singularities that pinch the hypercontour of integration. At a cusp of the Landau curve a multiple pinch occurs in the space of the integration variables, corresponding to what happens in the case of a single integration when in the complex plane of the integration variable three singularities coincide instead of two. In passage through such a point the singularity of the amplitude may be "dissolved" because of a change in pairing of the singularities in such a way that the new pair subsequently does not pinch the contour of integration. This effect proves vital in an understanding of the analytic properties of the amplitude associated with

[1] P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, Nuovo cimento **19**, 939 (1961); R. J. Eden, Phys. Rev. **121**, 1567 (1961).
[2] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, Phys. Rev. **122**, 307 (1961).
[3] See, for example, H. Hilton, *Plane Algebraic Curves* (Oxford University Press, New York, 1920).
[4] R. Eden, C. Enz, and J. Lascoux, Bull. Am. Phys. Soc. **5**, 284 (1960); V. A. Kolkunov, L. B. Okun, and A. P. Rudik, Soviet Phys.—JETP **11**, 634 (1960); G. Wanders, Nuovo cimento **18**, 580 (1960).
[5] It should perhaps be remarked that these complications are not related to the observations of N. Nakanishi (Kyoto preprint), who has not made use of the crucial property that the Landau curve is *asymptotic* to the normal thresholds.

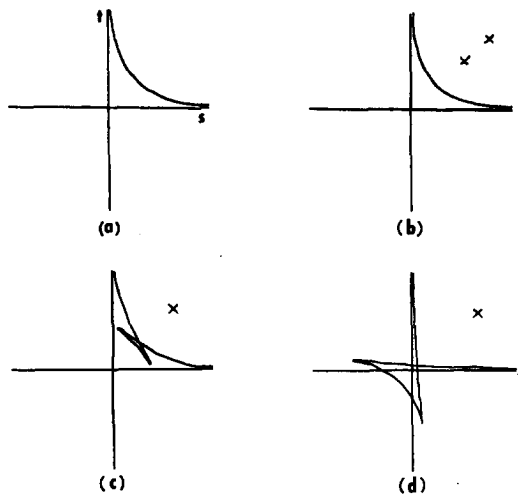

FIG. 2. Part of the Landau curve (in the real s, real t plane) associated with the graph in Fig. 1, showing positions of acnodes, crunodes and real cusps for four values of the external mass M.

Fig. 1. It is found that the complex part of the Landau curve joining the two acnodes is singular on the physical sheet. When the real cusps appear this singular surface connects the remaining acnode to the real curve joining the two cusps.

In Sec. 5 we allow a more extensive type of variation of the masses and obtain a more restrictive limitation on the masses for there to be no acnodes or complex singularities. In Sec. 6 we consider the general problem of obtaining criteria for the occurrence of nodes and cusps in higher orders in perturbation theory. It seems probable that nodes are a common feature of Landau curves, but cusps are comparatively rare. No feasible general method of ascertaining when the nodes are acnodes has been discovered.

2. THE PROBLEM

The Landau curves[6] for any Feynman diagram are the curves on which all singularities of the amplitude for the diagram are known to lie.[7] One of the central problems in discussing analyticity properties in perturbation theory is the determination of those parts of Landau curves that are singular on the physical sheet.[1] The curve in Fig. 2(a) is drawn in the real (s,t) plane and is a part of the Landau curve for the diagram of Fig. 1, in the equal mass case. This part is a curve of virtual singularity; that is, it is singular from the physical sheet in the limits $(s^0\pm i\epsilon,\ t^0\pm i\epsilon') \to (s^0,t^0)$, where s^0, t^0 are real, but not in the limits $(s^0\pm i\epsilon,\ t^0\mp i\epsilon') \to (s^0,t^0)$.

The Landau curve for this diagram also has complex parts, as can be seen by considering real search lines that successively intersect the real arc of Fig. 2(a), touch it, and then fail to intersect it. The complex surface joined to the arc lies in

$$\mathrm{Im}s/\mathrm{Im}t<0, \tag{1}$$

and comprises two complex conjugate parts.

Since the arc is a curve of virtual singularity, the attached surface is not singular on the physical sheet and the Mandelstam representation holds. The limit (1) on to the curve differs from the limit

$$\mathrm{Im}s/\mathrm{Im}t>0 \tag{2}$$

because the arc lies within the region of the real (s,t) plane where the cuts attached to two normal thresholds overlap. Thus the cuts separate the two limits.

The dual diagram[6,8] for Fig. 1, is drawn in Fig. 3. This can be discussed most conveniently as a diagram in a three-dimensional Euclidean space in which the coordinates in general take complex values. Formulas for the distance between points and equations for the position of a point given its distance from three other points are the same as if the space were real. So also are conditions for coplanarity, collinearity, etc. Trigonometrical formulas also are the same, except that sines and cosines may take values greater than one in

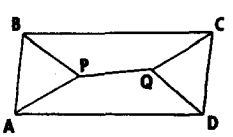

FIG. 3. The dual diagram for the graph in Fig. 1.

magnitude or be complex. The usual Landau equations imply that the lines PA, PB, PQ, and PQ, QC, QD lie, respectively, in two planes.[6,8] The squares of the lengths AC, BD give s and t. The two coplanarity conditions, the conditions following from the external mass values, and the use of symmetry in equating values of Feynman parameters, give a relation between s and t which is the equation of the Landau curve [Eq. (4)].

For any values of the masses there are values of (s,t) for which the dual diagram can be drawn in a single plane. Inspection of the dual diagram shows that the masses can be chosen so that in this situation:

(a) The coordinates of the vertices are all real. Then the corresponding values of (s,t) are less than the values they take at the normal thresholds. [The latter values are $(\mathrm{AP}+\mathrm{PQ}+\mathrm{QC})^2$ and $(\mathrm{BP}+\mathrm{PQ}+\mathrm{QD})^2$.]

(b) The points P, Q lie, respectively, inside the triangles ABQ, CDP. This means that all the Feynman parameters α are real and positive.

(c) There are no anomalous thresholds. The condition for this is that, if the configuration ABCPQ is constructed in a plane with PQ collinear, the point P lies outside the triangle ABC. Similar conditions apply to the configurations ABDPQ, BCDPQ, CDAPQ.

[6] L. D. Landau, Nuclear Phys. **13**, 181 (1959).

[7] J. C. Polkinghorne and G. R. Screaton, Nuovo cimento **15**, 289 and 925 (1960).

[8] J. C. Taylor, Phys. Rev. **117**, 261 (1960); P. V. Landshoff, Nuclear Phys. **20**, 129 (1960); L. B. Okun' and A. P. Rudik, Nuclear Phys. **15**, 261 (1960).

When these conditions are satisfied the dual diagram appears exactly in the configuration shown in the plane figure of Fig. 3. Then the corresponding point (s,t) lies on part of the Landau curve where all the Feynman parameters α are positive and in a region free from cuts, so that the two limits (1) and (2) on to the real s, t plane are the same.

It is known that the denominator of the Feynman integral for the graph is some power of

$$D=fs+gt+K, \tag{3}$$

where f, g are each a simple product of the integration parameters α, and f, g, K are each real when the α are real.[9] When the α are real and positive, f and g are both real and positive; hence, D does not vanish when the condition in Eq. (2) holds. Thus, the multidimensional hypercontour of integration in the Feynman integral remains undistorted as the limit (2) on to the real point s, t is approached. This means there is a pinch on this contour in the limit which must therefore be a singularity of the integral. Thus we have the result that a portion of a Landau curve and at least part of its attached surface is singular on the physical sheet although there are no anomalous thresholds. The manner in which this complex singularity develops is considered in the next section.

3. THE LANDAU CURVE

The equation of the part of the Landau curve which we will consider can be obtained from the dual diagram of Fig. 3. We take each mass equal to unity except for the two external masses BC and AD which we take to be M. We choose a solution having $\alpha_{AP}=\alpha_{CQ}$ and $\alpha_{BP}=\alpha_{DQ}$. Then the equation of the part of the curve we require is given parametrically by

$$s=5+4\cos\phi+2(2-\tfrac{1}{2}M^2+\cos\theta+\cos\phi)\sin\phi/\sin\theta, \tag{4a}$$

$$t=5+4\cos\theta+2(2-\tfrac{1}{2}M^2+\cos\theta+\cos\phi)\sin\theta/\sin\phi, \tag{4b}$$

where $\theta+\phi=\pi/3$. θ and ϕ are in fact the angles between PQ and AP, BP produced. On this part of the curve they are also equal to the angles between PQ and CQ, DQ produced.[10] It can be seen that when θ or ϕ tends to zero, the curve is asymptotic to the normal thresholds $t=9$ or $s=9$.

The Feynman parameters are given by

$$\frac{\alpha_{AP}}{\sin\phi}=\frac{\alpha_{BP}}{\sin\theta}=\frac{2\alpha_{PQ}}{\sqrt{3}}=\frac{\alpha_{CQ}}{\sin\phi}=\frac{\alpha_{DQ}}{\sin\theta}. \tag{5}$$

All the α are real and positive when

$$0<\theta<\tfrac{1}{3}\pi, \quad 0<\phi<\tfrac{1}{3}\pi. \tag{6}$$

The slope of the curve is

$$ds/dt=-\sin^2\phi/\sin^2\theta \quad (=-g/f), \tag{7}$$

which is always negative for real θ and ϕ.

The part of the real curve for which θ and ϕ are in the range (6) is drawn in Fig. 2 for various values of M^2. For $M^2<4+2\sqrt{2}=6.828$, it has the familiar form shown in Fig. 2(a).

To obtain the form as M^2 increases past this value, we take

$$\theta=\pi/6+i\eta, \quad \phi=\pi/6-i\eta. \tag{8}$$

When $4+2\sqrt{2}<M^2<4+5/\sqrt{3}$, there are real nonzero values of η so that

$$\mathrm{Im}s=\mathrm{Im}t=0. \tag{9}$$

Thus at $M^2=4+2\sqrt{2}$, coincident acnodes appear on the line $s=t$, and as M^2 increases further they separate giving the arc and two acnodes shown in Fig. 2(b). [For $M^2<4+2\sqrt{2}$, the equations formally identical with (9) have a solution for complex η, but they do not represent the vanishing of Ims, Imt. They therefore give complex nodes, which need not concern us.]

At the value $M^2=4+5/\sqrt{3}=6.887$, one of the acnodes lies on the curve. Above this value of M^2, this acnode has changed into a crunode and there are two ceratoid cusps on the curve as in Fig. 2(c). The positions of the cusps can be found from

$$ds/d\theta=dt/d\theta=0. \tag{10}$$

For $M^2<4+5/\sqrt{3}$, the equations (10) have a pair of solutions that correspond to complex s and t. They are complex cusps and they are important for our later discussion.

As M^2 increases further, the two cusps cross the normal thresholds; and for $7>M^2>4\sqrt{3}=6.928$, part of the arc lies outside the crossed cuts [Fig. 2(d)] although there are no anomalous thresholds. Above $M^2=7$, there will be anomalous thresholds and the character of the curve changes in a manner which we do not need to discuss here.

The curves drawn in Fig. 2 correspond to most of the masses being equal. A slightly more general situation is discussed in Sec. 5. No radical change in the character of the curves can be introduced by breaking the equalities between the masses because this part of the Landau curve can have no horizontal or vertical tangents. Because of the explicit form of f, g in the Feynman denominator D of Eq. (3), such tangents can only be the straight line Landau curves corresponding to the vertex graphs obtained from Fig. 1 by contracting one of the internal lines. These do not lie above the normal thresholds.

These details of the real section of the Landau curve are sufficient for our discussion, but we require also

[9] K. Symanzik, Prog. Theoret. Phys. (Kyoto) **20**, 690 (1958); J. D. Bjorken, Stanford preprint (1959); R. J. Eden, Phys. Rev. **119**, 1763 (1960).

[10] There are other parts of the curve for which the dual diagram does not have these symmetry properties. These may include arcs of virtual singularity and may provide a portion of the boundary of the support of the spectral function when the Mandelstam representation holds, but this is not known.

information about the attached complex surface, which is a more involved matter.[11] Consider intersections of the Landau curve with the searchlines

$$s+t=\lambda, \tag{11}$$

where λ is a real parameter. The paths traced out in the complex s (or t) plane as λ is varied are shown in Fig. 4, for the case of Fig. 2(b), that is for

$$4+2\sqrt{2}<M^2<4+5/\sqrt{3}.$$

The acnodes are labeled A, B, the midpoint of the real arc is denoted C, and D, D′ are complex cusps. The parts of the paths drawn in solid line correspond to real η, where η is defined by (8), and those in broken line correspond to complex η.

From Fig. 4, we see that it is possible to trace a path on the complex surface which joins the acnodes, passes through one of the complex cusps, and meets the real curve. It can be verified[12] that if a path on the Landau surface is traced that does not pass through a cusp [so that Eq. (11) is not satisfied], then its projection on

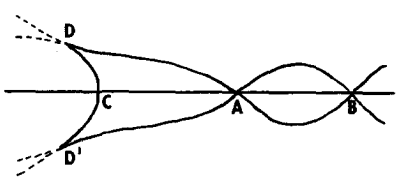

FIG. 4. The projection, in the complex s plane (or t plane) of a pair of complex conjugate paths on the Landau surface from the real branch (C), through the cusps (D or D') to the acnodes (A and B).

to the complex s plane, or on to the complex t plane, necessarily goes round the cusp in one or other of the two planes but not in both.

The only other feature of the complex surface that need be described here is its form near either an acnode or a real cusp. Again using searchlines, we see that in the neighborhood of such a point there are to be found portions of surface in both regions (1) and (2) of complex (s,t) space.

4. SINGULAR PARTS OF THE LANDAU CURVE

We will refer to a limit from the physical sheet on to the real s, t plane when (1) is satisfied as the *corresponding* limit, and when (2) is satisfied as the *opposite* limit. When the Landau curve has the form shown in Fig. 2(a), the real arc is singular in the opposite limit but is not singular in the corresponding limit. If now M^2 is increased, the Landau surface changes continuously and it is possible for a real or complex part to become singular on the physical sheet only if that part grows from zero size.[2] As M^2 increases past the value $4+2\sqrt{2}$, the two acnodes appear, at first coincident, and then they separate. Thus, the piece of the Landau surface between the acnodes grows from zero and may therefore be singular. We denote this piece of the Landau surface by ξ.

To determine whether ξ is in fact singular, we increase M^2 further until $M^2>4\sqrt{3}$, when the situation of Fig. 2(d) has been reached. During this variation of M^2, ξ remains finite in size and does not cross any branch cuts. The situation of Fig. 2(d) has part of the real Landau curve outside both normal threshold cuts, and on this part the Feynman parameters are positive. Hence, this part is singular in both corresponding and opposite limits, and the piece of the complex surface attached to it is also singular. This is the piece into which ξ has grown as M^2 was increased. Hence ξ is singular as soon as it appears. We conclude that there are complex singularities in the physical sheet for $M^2>4+2\sqrt{2}$.

This conclusion raises two difficulties. As we have said, in the neighborhood of the cusps a searchline of positive slope shows that there are parts of the Landau surface that satisfy condition (2). The cusp is singular in the corresponding limit since it has positive Feynman parameters and there is no distortion of the hypercontour of integration before the limit (2) on to the real plane is reached. But equally because there is no distortion of the hypercontour before this limit is reached, from the form of D in Eq. (3) it follows that there cannot be any singularities in the region (2) where s and t have imaginary parts of the same sign. Thus we have a real singularity on the curve near the cusp but the attached surface in region (2) is not singular. No end point (lower-order) singularities occur at the cusp, so we have a singularity that disappears in some hitherto unknown manner.

The second difficulty is that, in the case shown in Fig. 2(b), the real arc is not singular in the opposite limit. Thus, the part of the complex surface attached to it and lying in the neighboring unphysical sheet is also nonsingular. (This unphysical sheet is reached through the *two* normal threshold cuts that are along the real axes in the complex s, complex t planes.) We can follow a path on this surface, passing through the complex cusp D of Fig. 4 to reach the acnode A. We can then continue on this path back into the physical sheet onto ξ and conclude that ξ is not singular, unless along this path the singularity has disappeared in some manner not related either to normal threshold cuts or to anomalous thresholds.

To understand the resolution of these difficulties, it is helpful at first to imagine that the Feynman integration is over a single variable α only instead of many variables. The singular part of the Landau curve in complex s, t space is given by those points for which the contour of integration is "pinched" between two zeros of the Feynman denominator D.[7] If we start at a

[11] Compare a discussion of a curve very similar to that of Fig. 2 by F. Klein, Math. Ann. **10**, 365 (1876). We are grateful to Mr. F. P. White for drawing our attention to this paper.

[12] We are grateful to Mr. P. Swinnerton-Dyer and Mr. C. Paradine who performed computations for us on EDSAC II in connection with this problem.

singularity on the Landau curve and follow the curve, the singularity remains so long as the pinch persists. It can fail to persist if either (a) it falls off an end point of the contour of integration. This can occur only at a point of contact of the Landau curve with a normal threshold, which can only be at infinity in this example; or (b) the path we follow "steps over" a cut. This is equivalent to taking a new contour of integration by lifting the old contour across one or more zeros of D.

There is, however, a third method by which the singularity can disappear along the path, that has not previously been considered. This can occur at a "double pinch" where a pair of zeros of D that pinches the integration contour meets a third zero of D. One of the zeros forming the pinch can then change its mate and pair with the new zero. If the new zero is on the same side of the contour this process "dissolves" the pinch as illustrated in Fig. 5. It will be recalled that the Landau surface requires only that a pair of zeros is coincident; thus, if two pinching zeros meet a third zero, we have a singular point of the Landau surface (in the geometrical sense). Depending on how we pass this singular point, the pinch may be either maintained or dissolved.

As in cases (a) and (b), this third method can be generalized to a multidimensional integral, and this is further discussed in Sec. 6. The singular point at which the dissolution of the pinch takes place is a cusp on the Landau curve that we are following.

FIG. 5. A one-dimensional model of the process of dissolution of a pinch by re-pairing.

We now see that it is consistent for the part ξ of the surface to be singular since the path we have described linking it to the real arc passes through a complex cusp. At the cusp, the pinch is dissolved along this path and only the part of the path on the side of the cusp nearest to ξ is singular.

There is still an apparent difficulty if we trace a path on the Landau surface from ξ to the real arc so that it avoids the cusp. Then we do not have a double pinch at any point of this path. However, a detailed analysis of such a path shows that the Landau surface is twisted in the neighborhood of the cusp, so that in projection this path must go round the cusp either in the s plane or in the t plane. This ensures that all such paths must go through the three-dimensional cut attached to the singular part of the Landau surface in the neighborhood of the cusp. Thus, for any path that avoids the cusp we either step over this cut, which is a nonanalytic change and removes the singularity, or we go through it and end up on the wrong Riemann sheet.

When the real arc has two real cusps, the part of the arc between the cusps is attached directly to ξ. Thus it is singular in the opposite limit since ξ lies on this side of the normal threshold branch cuts. It is also singular in the corresponding limit since the Feynman parameters α are positive and this limit maintains an undistorted contour. Consistency is obtained in this situation by the different pairing of the three zeros of D in these two limits so that both are singular.

Finally, for a searchline of positive slope near a cusp, it is clear that one path becomes complex on the physical sheet while the other becomes complex on the neighboring unphysical sheet. On the former, the singularity disappears either because we go through the real cusp and the pinch is dissolved, or because we go through the attached branch cut. On the latter, the singularity is retained; if it were not retained we should have a violation of the disk theorem of many complex variable theory.[13]

5. CRITICAL VALUES FOR THE EXTERNAL MASSES

We now consider a case when not all the internal masses are equal. We retain some measure of symmetry by taking the masses for lines AD and BC equal to M, for AB and CD equal to m, for AP and QC equal to μ, for BP and QD equal to μ', and for PQ equal to 1. The equation for the part of the Landau curve under consideration is

$$s=1+4\mu^2+4\mu\cos\phi+\frac{\mu\sin\phi}{\mu'\sin\theta}\{-M^2+(1+\mu^2+\mu'^2)+2\mu\cos\phi+2\mu'\cos\theta\}+\mu^2\frac{\sin\phi}{\sin\theta}(2\cos\gamma), \quad (12)$$

where

$$\cos\gamma=\frac{\mu^2+\mu'^2-m^2}{2\mu\mu'}, \quad \theta+\phi=\gamma. \quad (13)$$

The parametric equation for t is obtained by interchanging ϕ with θ, and μ' with μ.

To find the simplest form that gives a critical value for m^2, we take $\mu=\mu'=1$ initially. Then this part of the Landau curve is symmetric about $\theta=\phi=\frac{1}{2}\gamma$. The nodes lie on the line of symmetry on the complex surface

$$\theta=\tfrac{1}{2}\gamma+i\eta, \quad \phi=\tfrac{1}{2}\gamma-i\eta \quad (14)$$

at the points given by the roots of

$$x^2\cos\gamma+\tfrac{1}{2}x(3-M^2+2\cos\gamma)\cos\tfrac{1}{2}\gamma+\cos^2\tfrac{1}{2}\gamma=0, \quad (15)$$

where $x=\cosh\eta$. They first appear as acnodes when the roots of (15) are equal, so that

$$(3-M^2+2\cos\gamma)^2\cos^2\tfrac{1}{2}\gamma=16\cos\gamma\cos^2\tfrac{1}{2}\gamma. \quad (16)$$

Stability of the particle of mass m implies that $0<\gamma<\pi$ so we may cancel the factor $\cos^2\frac{1}{2}\gamma$. On substituting

$$2\cos\gamma=2-m^2,$$

[13] See, for example, A. S. Wightman, Les Houches lectures, 1960 (to be published).

we obtain

$$\lambda_1 \equiv [5-(M^2+m^2)]^2-8(2-m^2)=0. \qquad (17)$$

When one root of (16) is $x=\cosh\eta=1$, the corresponding acnode lies on the real arc. This is the condition for the real cusps first to appear:

$$\lambda_2 \equiv 5+\frac{4-\frac{3}{2}m^2}{[(1-\frac{1}{4}m^2)]^{\frac{1}{2}}}-(M^2+m^2)=0. \qquad (18)$$

Equation (15) also indicates that when $\cos\gamma=0$, one acnode is at infinity. Then

$$\lambda_3 \equiv m^2-2=0. \qquad (19)$$

The curves λ_1 and λ_2, and the line λ_3 are shown in Fig. 6. The location of the nodes in the real s, t plane is shown in Fig. 7 for values of (M^2+m^2) and m^2 corresponding to regions A, B, C, D of Fig. 6. In region A, there are two complex nodes and no acnodes. In region B we have two acnodes to the right and above the curve. This is the situation that has been discussed in previous sections. Both acnodes are singular on the physical sheet and are connected by a surface singular in the physical sheet. On the line λ_3, one acnode is at infinity.

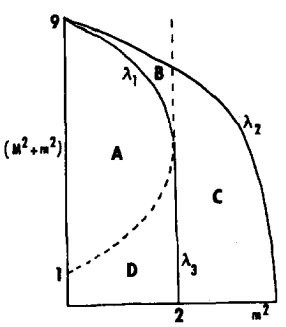

FIG. 6. When two masses are varied, the curves for coincident acnodes (λ_1) for an acnode becoming a cusp (λ_2) and an acnode at infinity (λ_3). In region B or C the Mandelstam representation does not apply.

In region C of Fig. 6, the node lies to the left of and below the curve and it is not singular when approached from the physical sheet.[14] However the swathe of complex surface that extends from the other acnode to infinity is singular in the physical sheet. Between regions C and D this acnode goes to infinity and reappears as an acnode not singular on the physical sheet. The curve λ_1 between regions D and A represents the transition points where both acnodes coincide and become complex nodes.

We conclude that complex singularities appear in the physical sheet for any values of M^2 and m^2 to the right of the heavy line shown in Fig. 6, which is made up of part of the curve λ_1 and part of the line λ_3. The conditions are therefore

(a) If $m^2>2$, there are always complex singularities in the physical sheet.

[14] This node may become a crunode on another part of the Landau curve.

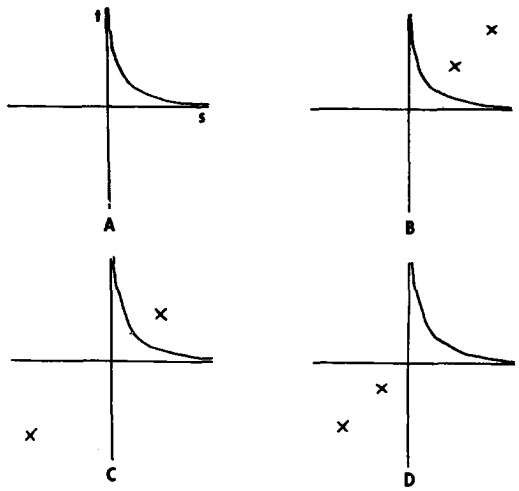

FIG. 7. The positions of the real nodes in relation to the important part of the real Landau curve, for the ranges of values of M^2 and m^2 shown in Fig. 6.

(b) If $m^2<2$, there are complex singularities if

$$(M^2+m^2)>5+2[(4-2m^2)]^{\frac{1}{2}}.$$

If (a) holds then as M^2 decreases (for fixed m^2) the complex singularities move further away, but for no value of M^2 do they disappear.

The more general masses mentioned earlier in this section do not give any substantial change. The condition (a) becomes (a′) if $m^2>\mu^2+\mu'^2$ there are always complex singularities in the physical sheet.

It will be noted that this corresponds to $\gamma=\frac{1}{2}\pi$ in Eq. (13). We conclude that the appearance of singular acnodes in this sixth-order diagram can occur for values of the masses that exclude the crunode transition process of reference 2 even in fourth order. Physical examples of processes leading to complex singularities on account of singular acnodes are given by $\pi-\Sigma$ and $P-\Lambda$ scattering. However, these singularities are so located as not to upset partial wave dispersion relations.

6. NODES AND CUSPS

Since the occurrence of nodes and cusps in Landau curves plays so important a part in determining the analytic properties of scattering amplitudes, it is necessary to discover the extent to which they appear in curves associated with graphs other than that of Fig. 1.

If the algebraic equation of a curve is $F(s,t)=0$, a node occurs at points for which

$$F=\partial F/\partial s=\partial F/\partial t=0. \qquad (20)$$

The possibility of the equations (20) being satisfied simultaneously implies a condition on the coefficients in the algebraic form F, so that most curves do not have nodes. It was previously thought[1,2] that, for Landau curves, this condition would only be satisfied for fortuitous values of the masses and that it could then

be broken by changing the masses a little. We have seen above that in fact this is not so in all cases.

In practice the explicit form of F for a Landau curve is not available and in any case it would not be attractive to use. We shall resort, therefore, to a discussion in terms of dual diagrams. We recall that these are vector diagrams, constructed to satisfy the Landau equations

$$q_i^2=m_i^2, \tag{21}$$

$$\sum_l \alpha_i q_i=0. \tag{22}$$

As usual, q_i, m_i, and α_i are, respectively, the momentum, mass, and Feynman parameter associated with the graph under consideration. Here, q_i may be regarded as a 3-vector. In (22), the summation is taken round the lth closed loop of the graph. Equation (22) serves to define the α_i and may also impose geometrical constraints on the dual diagram.

At a node, two parts of the Landau curve cross and so, for the given external momenta corresponding to that point, it is possible to construct two different dual diagrams with internal momenta $q_i^{(1)}$ and $q_i^{(2)}$. We shall define

$$q_i^{(\pm)}=q_i^{(1)}\pm q_i^{(2)} \tag{23}$$

and suppose that separate vector diagrams are now drawn for $q_i^{(+)}$ and $q_i^{(-)}$. The external momenta only

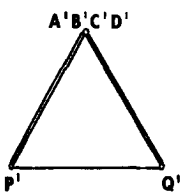

FIG. 8. The (−) diagram for determining the nodes on the Landau curve given by the dual diagram in Fig. 3. The lettering of the vertices corresponds to that in Fig. 3.

appear in the diagram for $q_i^{(+)}$; this diagram is similar in appearance to the original dual diagram but does not satisfy conditions (21) and (22). The general form of the vector diagram for $q_i^{(-)}$ is obtained by contracting the external lines in the original dual diagram. For Fig. 1 this gives Fig. 8, where the double lines represent two equal momenta.

Equations (21) and (22) are replaced by

$$\left.\begin{aligned} q_i^{(+)}\cdot q_i^{(-)}&=0 \\ q_i^{(+)2}+q_i^{(-)2}&=m_i^2 \end{aligned}\right\} \tag{21'}$$

and

$$\left.\begin{aligned} \sum_l (\alpha_i^{(+)}q_i^{(+)}+\alpha_i^{(-)}q_i^{(-)})&=0 \\ \sum_l (\alpha_i^{(-)}q_i^{(+)}+\alpha_i^{(+)}q_i^{(-)})&=0 \end{aligned}\right\}, \tag{22'}$$

where $\alpha_i^{(\pm)}$ have definitions similar to (23).

Inspection of these equations and consideration of the number of geometrical constraints that they impose on the (±) vector diagrams does not enable us to rule out the possibility that in general they determine a finite number of nodes on the Landau curve. They fail to do so, however, if the (−) vector diagram contains a trebled line, since then the perpendicularity condition in (21′) cannot be satisfied in the three-dimensional space.

Thus it is possible that many graphs give Landau curves having nodes. The important question will be to determine whether these nodes are acnodes and, if so, whether they are in a "dangerous" part of the real (s,t) plane. If a node is real, the corresponding external momenta may be represented by real vectors in a space of appropriate metric. (When it is an acnode, $q_i^{(\pm)}$ are just the real and imaginary parts of the q_i.) The most dangerous situation is probably that an acnode occurs in the crossed cuts, and then the metric is non-Euclidean. This makes the use of some sort of majorisation procedure seem difficult.

We now turn to cusps. At a cusp, in addition to (20), there is the condition that the two branches have a common tangent

$$\left(\frac{\partial^2 F}{\partial s\partial t}\right)^2=\frac{\partial^2 F}{\partial s^2}\cdot\frac{\partial^2 F}{\partial t^2}. \tag{24}$$

Thus, cusps in algebraic curves are even less likely than nodes.

If D is the homogeneous Feynman denominator of Eq. (3), the Landau equations (21) may be written as

$$D_i\equiv\partial D/\partial\alpha_i=0. \tag{25}$$

Hence, on the Landau curve,

$$0=dD_i=f_i ds+g_i dt+D_{ij}d\alpha_j. \tag{26}$$

At a cusp $ds=dt=0$ so that

$$D_{ij}d\alpha_j=0. \tag{27}$$

Thus, the $n\times n$ matrix D_{ij}, which is normally of rank $(n-1)$ at a point of the Landau curve, becomes of rank $(n-2)$. D_{ij} determines the form of the surface $D=0$ in the neighborhood of the critical point in α space. When the rank of D_{ij} is $(n-1)$, this form is locally conelike, and the possibility of a singularity arises from pinching of the multidimensional hypercontour of integration between the two halves of the cone. When the rank of D_{ij} is $(n-2)$, this cone degenerates and the intersection of D with the hypercontour involves more points than necessary for the pinch.

To translate this into the dual diagram formalism we may apply the differentiation procedure of Eq. (26) instead directly to Eqs. (21) and (22). We then obtain

$$q_i\cdot dq_i=0 \tag{21''}$$

and

$$\sum_l (q_i d\alpha_i+dq_i\alpha_i)=0. \tag{22''}$$

Corresponding to the (+) vector diagram in the node case, we now have the ordinary dual diagram itself; while corresponding to the (−) diagram, we have the vector diagram for the increments dq_i. Now the con-

straint equation (22) is also operative, and, in general, the number of conditions becomes too great to be satisfied for arbitrary external masses. An exception is provided by the graph of Fig. 1 for which the two coplanarity conditions for the dual diagram make one of the perpendicularity conditions for the dq_i vector diagram dependent on the other two.

Thus it appears that cusps are not a general feature of Landau curves. Note, however, that if a graph G provides a cusp, then any Feynman graph within which the graph G is embedded has cusps on its leading Landau curve.

ACKNOWLEDGMENTS

We wish to thank Professor R. Karplus, Professor S. Mandelstam, and Professor K. Stein for discussions. One of us (P.V.L.) is grateful to the D. S. I. R. for a grant.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

Wave Equations for Scalar and Vector Particles in Gravitational Fields

STIG HJALMARS
Royal Institute of Technology, Stockholm, Sweden
(Received April 28, 1961)

An earlier representation of the Kemmer wave equation in Riemann space is modified so as to remove the matrices, which in general have to be added to the differential operators of the equation. It is pointed out that the possibility of this removal is equivalent to the fact, known from Maxwell's equations, that if only scalars, vectors, and antisymmetrical tensors are involved, the field equations can be written with ordinary derivatives without explicit use of the affine connection. The wave equation is written in component form, and the photon zero mass case is obtained by means of the most general matrix mass term, without any questionable limiting process.

1. INTRODUCTION AND SUMMARY

THE relativistic wave equation of the electron in Riemann space has been investigated by many authors since 1928, e.g., by E. Schrödinger[1] and V. Bargmann.[2] As is well known, the space-dependent matrix coefficients γ^k of this equation have to satisfy a generally covariant commutation rule, obtained from the commutation rule for the $\gamma^\circ{}_k{:}s$ of the ordinary Dirac equation by changing the $\gamma^\circ{}_k{:}s$ into $\gamma^k{:}s$ and the $\delta_{kl}{:}s$ into $g^{kl}{:}s$. Moreover, certain matrices Γ_k must be added to the partial derivatives of the equation in order to obtain covariant operators. The $\Gamma_k{:}s$ are usually obtained from the expression for the derivative of γ_k. The equation is covariant under arbitrary space-dependent similarity transformations as well as under simultaneous general coordinate transformations, the latter transforming the $\gamma^k{:}s$ as vector components. No connection can be assumed between the similarity transformation and the general coordinate transformation. Nevertheless, as shown by Bargmann,[2] a metric form can be defined, which permits the construction of c-number tensor fields, e.g., a current vector with vanishing covariant divergence.

As shown by O. Brulin and S. Hjalmars[3] in 1951, the first-order wave equation for scalar and vector particles, given by N. Kemmer,[4] can be generalized in the same way (II.1). The transformation properties of this generalized equation are summarized in Sec. II of the present paper. The coefficient matrices $\beta^\circ{}_k$, satisfying the Petiau[5] commutation rule, (II.4), are again to be replaced by space dependent matrices β^k, satisfying the corresponding, generally covariant commutation rule (II.3). The formula for the covariant derivative of β_k (II.5), as well as for the transformation of β^k (II.8) and Γ_k (II.9) are seen to have the same form as in the electron case.

Since the components of the wave function ψ for integer spin particles have essentially tensor transformation properties, the similarity transformation can, in this case, be connected to the general coordinate transformation. The matrix of the similarity transformation is then, of course, a representation of the full linear group. It is shown in Sec. III that, as soon as a connection is made, a representation of the $\beta^k{:}s$ and the $\Gamma_k{:}s$ can be constructed. Such a construction was explicitly made[3] by use of a convenient representation of the $\beta^\circ{}_k{:}s$[6] and of a similarity transformation, which transforms the five ψ components in the scalar case as the contravariant components of a 4-vector and a scalar and the ten ψ components in the vector case as the contravariant components of an antisymmetrical tensor of second rank and the contravariant components

[1] E. Schrödinger, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl. XI, 105 (1932).
[2] V. Bargmann, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl. XI, 346 (1932).
[3] O. Brulin and S. Hjalmars, Arkiv Fysik 5, 163 (1952).
[4] N. Kemmer, Proc. Roy. Soc. (London) A173, 91 (1939).
[5] G. Petiau, Compt. rend. 200, 374 (1935).
[6] S. Hjalmars, Arkiv Fysik 1, 41 (1949).

straint equation (22) is also operative, and, in general, the number of conditions becomes too great to be satisfied for arbitrary external masses. An exception is provided by the graph of Fig. 1 for which the two coplanarity conditions for the dual diagram make one of the perpendicularity conditions for the dq_i vector diagram dependent on the other two.

Thus it appears that cusps are not a general feature of Landau curves. Note, however, that if a graph G provides a cusp, then any Feynman graph within which the graph G is embedded has cusps on its leading Landau curve.

ACKNOWLEDGMENTS

We wish to thank Professor R. Karplus, Professor S. Mandelstam, and Professor K. Stein for discussions. One of us (P.V.L.) is grateful to the D. S. I. R. for a grant.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

Wave Equations for Scalar and Vector Particles in Gravitational Fields

STIG HJALMARS
Royal Institute of Technology, Stockholm, Sweden
(Received April 28, 1961)

An earlier representation of the Kemmer wave equation in Riemann space is modified so as to remove the matrices, which in general have to be added to the differential operators of the equation. It is pointed out that the possibility of this removal is equivalent to the fact, known from Maxwell's equations, that if only scalars, vectors, and antisymmetrical tensors are involved, the field equations can be written with ordinary derivatives without explicit use of the affine connection. The wave equation is written in component form, and the photon zero mass case is obtained by means of the most general matrix mass term, without any questionable limiting process.

1. INTRODUCTION AND SUMMARY

THE relativistic wave equation of the electron in Riemann space has been investigated by many authors since 1928, e.g., by E. Schrödinger[1] and V. Bargmann.[2] As is well known, the space-dependent matrix coefficients γ^k of this equation have to satisfy a generally covariant commutation rule, obtained from the commutation rule for the $\gamma^\circ{}_k{:}s$ of the ordinary Dirac equation by changing the $\gamma^\circ{}_k{:}s$ into $\gamma^k{:}s$ and the $\delta_{kl}{:}s$ into $g^{kl}{:}s$. Moreover, certain matrices Γ_k must be added to the partial derivatives of the equation in order to obtain covariant operators. The $\Gamma_k{:}s$ are usually obtained from the expression for the derivative of γ_k. The equation is covariant under arbitrary space-dependent similarity transformations as well as under simultaneous general coordinate transformations, the latter transforming the $\gamma^k{:}s$ as vector components. No connection can be assumed between the similarity transformation and the general coordinate transformation. Nevertheless, as shown by Bargmann,[2] a metric form can be defined, which permits the construction of c-number tensor fields, e.g., a current vector with vanishing covariant divergence.

As shown by O. Brulin and S. Hjalmars[3] in 1951, the first-order wave equation for scalar and vector particles, given by N. Kemmer,[4] can be generalized in the same way (II.1). The transformation properties of this generalized equation are summarized in Sec. II of the present paper. The coefficient matrices $\beta^\circ{}_k$, satisfying the Petiau[5] commutation rule, (II.4), are again to be replaced by space dependent matrices β^k, satisfying the corresponding, generally covariant commutation rule (II.3). The formula for the covariant derivative of β_k (II.5), as well as for the transformation of β^k (II.8) and Γ_k (II.9) are seen to have the same form as in the electron case.

Since the components of the wave function ψ for integer spin particles have essentially tensor transformation properties, the similarity transformation can, in this case, be connected to the general coordinate transformation. The matrix of the similarity transformation is then, of course, a representation of the full linear group. It is shown in Sec. III that, as soon as a connection is made, a representation of the $\beta^k{:}s$ and the $\Gamma_k{:}s$ can be constructed. Such a construction was explicitly made[3] by use of a convenient representation of the $\beta^\circ{}_k{:}s$[6] and of a similarity transformation, which transforms the five ψ components in the scalar case as the contravariant components of a 4-vector and a scalar and the ten ψ components in the vector case as the contravariant components of an antisymmetrical tensor of second rank and the contravariant components

[1] E. Schrödinger, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl. XI, 105 (1932).
[2] V. Bargmann, Sitzber. preuss. Akad. Wiss., Physik.-math. Kl. XI, 346 (1932).
[3] O. Brulin and S. Hjalmars, Arkiv Fysik 5, 163 (1952).
[4] N. Kemmer, Proc. Roy. Soc. (London) A173, 91 (1939).
[5] G. Petiau, Compt. rend. 200, 374 (1935).
[6] S. Hjalmars, Arkiv Fysik 1, 41 (1949).

of a 4-vector. By means of this connection, the β^k:s were constructed[3] as functions (III.16) of the $\beta^\circ{}_k$:s and the metric fundamental tensor g^{kl}, and also the Γ_k:s as functions (III.17) of the $\beta^\circ{}_k$:s and the affine connection $\Gamma_{lm}{}^k$.

In a recent paper, A. da Silveira[7] has treated the Kemmer wave equation in Riemann space by means of the four-leg formalism. It appears from this treatment that a representation of the Kemmer equation exists, in which the Γ_k:s have disappeared. In Sec. IV of the present paper, it is shown explicitly that the vanishing of the Γ_k:s from the wave equation can be afforded by a similarity transformation, which transforms the five ψ components in the scalar case as the contravariant components of a 4-vector density and a scalar, and the ten ψ components in the vector case as the contravariant components of an antisymmetrical tensor density of second rank and the covariant components of a 4-vector. In fact, it is shown that this connection determines a representation of the β^k:s, (IV.4), and of the Γ_k:s, (IV.5), for which the expression $\beta^k\Gamma_k$ of the wave equation vanishes, (IV.6), as required.

In Sec. V, the corresponding wave equation is written in component form. It is seen from these component equations that the possibility of removing the Γ_k:s from the wave equation is equivalent to the fact, known from Maxwell's equations, that if only scalars, vectors, and antisymmetrical tensors are involved, the field equations in Riemann space can be written with ordinary derivatives only, without explicit use of the affine connection $\Gamma_{lm}{}^k$. Consequently, this possibility does not exist for particles of higher integer spin than 1, e.g., spin 2.[8]

It is also shown that the photon zero mass case can, just as in special relativity,[9,6] be obtained by means of the most general matrix mass term, without any questionable limiting process.

II. TRANSFORMATION PROPERTIES OF THE WAVE EQUATION

It has been shown[3] that the Kemmer[4] wave equation for scalar and vector particles can, in Riemann space, be written in the form

$$[\beta^k(\partial_k-\Gamma_k)-\mu c/\hbar]\psi=0, \tag{II.1}$$

with

$$\partial_k=\partial/\partial x^k, \quad x^4=ict. \tag{II.2}$$

The matrices β^k of (II.1) satisfy

$$\beta^k\beta^l\beta^m+\beta^m\beta^l\beta^k=\beta^k g^{lm}+\beta^m g^{kl}. \tag{II.3}$$

This rule is obtained from Petiau's[5] commutation rule,

$$\beta^{\circ k}\beta^{\circ l}\beta^{\circ m}+\beta^{\circ m}\beta^{\circ l}\beta^{\circ k}=\beta^{\circ k}\delta^{lm}+\beta^{\circ m}\delta^{kl}, \tag{II.4}$$

by the same method of generalization as is used in the electron case.[1,2] The superscript $^\circ$ in (II.4) is used to denote quantities, belonging to the orthogonal systems of special relativity. The matrices $\beta^{\circ k}$ are thus the ordinary Kemmer matrices with coordinate-independent matrix elements.

As in the electron case, the matrices Γ_k are introduced to secure the covariance of the derivatves ∂_k of ψ. They can be shown[3] to satisfy the same differential equations as in the electron case, i.e.,

$$(\partial\beta_k/\partial x^l)-\Gamma_{kl}{}^s\beta_s+\beta_k\Gamma_l-\Gamma_l\beta_k=0, \tag{II.5}$$

where the β_k:s are the covariant components of the β^k:s, obtained by lowering the indices in the ordinary way.

The quantity μ is in general a matrix, the form of which will be discussed in Sec. V.

Now, perform a general coordinate transformation $x^k\rightarrow x^{k'}(x^k)$, with

$$dx^{k'}=a_l{}^k dx^l, \quad dx^k=b_l{}^k dx^{l'}, \tag{II.6}$$

and a simultaneous similarity transformation,

$$\psi'=S^{-1}\psi, \quad \psi=S\psi'. \tag{II.7}$$

Then the β^k:s are assumed to be transformed as

$$\beta^{k'}=a_l{}^k S^{-1}\beta^l S, \tag{II.8}$$

i.e., as vector components under the coordinate transformation. It is proved[3] that (II.8) and (II.5) give the same transformation formula for Γ_k as in the electron case

$$\Gamma_k{}'=b_k{}^l(S^{-1}\Gamma_l S-S^{-1}\partial_l S). \tag{II.9}$$

It follows[3] from (II.7)–(II.9) that the wave equation (II.1) is covariant under the simultaneous coordinate and similarity transformation, provided that μ is transformed according to

$$\mu'=S^{-1}\mu S, \tag{II.10}$$

i.e., as a scalar under the coordinate transformation.

Moreover, the existence[3] of a metric form β with the transformation property

$$\beta'=S^+\beta S \tag{II.11}$$

makes it possible to construct[1,2] a current density $\psi^+\beta\beta^k\psi$ which, in virtue of (II.1), has vanishing covariant divergence.

III. CONSTRUCTION OF A REPRESENTATION

In Sec. II, no connection between the coordinate transformation and the similarity transformation was assumed. For integer spin particles, where the components of ψ have essentially the properties of tensor components, it is always possible to connect S to the coordinate transformation in such a way that S is a function $S(a_l{}^k,b_l{}^k)$ of the transformation coefficients $a_l{}^k$ and $b_l{}^k$ of (II.6). This, of course, implies that $S(a_l{}^k,b_l{}^k)$ is a representation of the full linear group.

[7] A. da Silveira, J. Math. Phys. **1**, 489 (1960).
[8] O. Brulin and S. Hjalmars, Arkiv Fysik **14**, 49 (1958); **16**, 19 (1959); **18**, 209 (1960).
[9] Harish-Chandra, Proc. Roy. Soc. (London) **A186**, 502 (1946).

$S(a_l{}^k, b_l{}^k)$ has thus to fulfil the group conditions, which can be stated as follows, cf. (II.7):

$$S(\delta_l{}^k, \delta_l{}^k) = 1, \tag{III.1}$$

$$S(\bar{a}_l{}^k, \bar{b}_l{}^k) S(\bar{\bar{a}}_l{}^k, \bar{\bar{b}}_l{}^k) = S(\bar{\bar{a}}_r{}^k \bar{a}_l{}^r, \bar{\bar{b}}_l{}^r \bar{b}_r{}^k). \tag{III.2}$$

For scalar and vector particles, this possibility has been proved[3] by explicit construction of S as a possible function of the $\beta^\circ{}_k$:s of (II.4) and the $a_l{}^k$:s and $b_l{}^k$:s.

In order to make the choice of S as natural as possible, it is advisable to start with a representation of the $\beta^\circ{}_k$:s, which directly takes into account the tensor properties of the ψ components under the orthogonal group. Such a representation is developed by S. Hjalmars[6] by means of a 5-dimensional formalism, introduced by O. Klein. In this representation, the five components of ψ in the scalar case are assumed to transform under orthogonal transformations as the components ψ_κ (greek indices from 1 to 5) of a 5-dimensional vector, which under the 4-dimensional orthogonal group splits in a 4-vector ψ_k (latin indices from 1 to 4) and a scalar ψ_5. The ten components of ψ in the vector case are assumed to transform as the ten independent components of an antisymmetrical, 5-dimensional tensor, $\psi_{\kappa\lambda} = -\psi_{\lambda\kappa}$, of second rank, which under the 4-dimensional orthogonal group splits into an antisymmetrical, 4-dimensional tensor ψ_{kl} of second rank and a 4-vector ψ_{5k}. The matrix elements of $\beta^\circ{}_k$ are then,[6] in the scalar case,

$$(\kappa|\beta^\circ{}_k|\lambda) = i[1-(k5)]\delta_{\kappa k}\delta_{\lambda 5} \tag{III.3}$$

and, in the vector case,

$$(\kappa\lambda|\beta^\circ{}_k|\mu\nu) = i[1-(\mu\nu)][1-(\kappa\lambda)][1-(k5)]\delta_{\kappa k}\delta_{\mu 5}\delta_{\lambda\nu}, \tag{III.4}$$

where $(\alpha\beta)$ denotes the permutation operator on the indices α, β. By means of the explicit representations (III.3) and (III.4), all relations of the Petiau-Kemmer algebra can be evaluated.

In the following, we need certain invariant projection operators $\omega^\circ{}_S$, $\omega^\circ{}_V$, $\omega^\circ{}_4$, $\omega^\circ{}_5$ projecting respectively into the subspaces, spanned by the ψ components of the scalar particle (ψ_k and ψ_5), by the ψ components of the vector particle (ψ_{kl} and ψ_{5k}), by the ψ components with no index higher than 4 (ψ_k and ψ_{kl}), and by the ψ components with one index 5 (ψ_5 and ψ_{5k}). These projection operators can be expressed as polynomials in the invariant $\beta^\circ{}_k\beta^{\circ k}$. They can be proved[3,6] to have the following properties:

$$[\omega_M, \omega_N] = 0, \quad (M, N = S, V, 4, 5), \tag{III.5}$$

$$\omega^\circ{}_M\omega^\circ{}_N = \omega^\circ{}_M\delta_{MN}, \quad \omega^\circ{}_M + \omega^\circ{}_N = 1, \quad (M, N = S, V) \text{ or } (M, N = 4, 5), \tag{III.6}$$

$$\omega^\circ{}_S\beta^\circ{}_k - \beta^\circ{}_k\omega^\circ{}_S = \omega^\circ{}_V\beta^\circ{}_k - \beta^\circ{}_k\omega^\circ{}_V = 0, \tag{III.7}$$

$$\omega^\circ{}_4\beta^\circ{}_k - \beta^\circ{}_k\omega^\circ{}_5 = \omega^\circ{}_5\beta^\circ{}_k - \beta^\circ{}_k\omega^\circ{}_4 = 0, \tag{III.8}$$

$$\beta^\circ{}_k\beta^{\circ k}\omega^\circ{}_4\omega^\circ{}_S = \omega^\circ{}_4\omega^\circ{}_S, \quad \beta^\circ{}_k\beta^{\circ k}\omega^\circ{}_5\omega^\circ{}_S = 4\omega^\circ{}_5\omega^\circ{}_S, \tag{III.9}$$

$$\beta^\circ{}_k\beta^{\circ k}\omega^\circ{}_4\omega^\circ{}_V = 2\omega^\circ{}_4\omega^\circ{}_V, \quad \beta^\circ{}_k\beta^{\circ k}\omega^\circ{}_5\omega^\circ{}_V = 3\omega^\circ{}_5\omega^\circ{}_V, \tag{III.10}$$

$$(\beta^\circ{}_k\beta^\circ{}_l - \delta_{kl})\omega^\circ{}_5\omega^\circ{}_S = 0, \tag{III.11}$$

$$[(\beta^\circ{}_k\beta^\circ{}_l - \delta_{kl})(\beta^\circ{}_r\beta^\circ{}_s - \delta_{rs}) + (\beta^\circ{}_r\beta^\circ{}_l - \delta_{rl})\delta_{ks}]\omega^\circ{}_5 = 0. \tag{III.12}$$

As soon as the $\beta^\circ{}_k$ representation is fixed and S is chosen as a possible function of $\beta^\circ{}_k$, $a_l{}^k$, and $b_l{}^k$, a representation of β^k, Γ_k, and β can be determined. In fact, if we want to calculate β^k, Γ_k, and β in a certain world point and in a given general coordinate system, we start with a coordinate system which is orthogonal and geodetic in this point. With this system and in this point we can put $\beta^k = \beta^{\circ k}$, $\Gamma_k = 0$, and $\beta^\circ = 1 - 2\beta^\circ{}_4{}^2$. The β^k, Γ_k, and β in the general coordinate system can then be calculated by means of (II.8), (II.9), (II.11) and the following relations for the transformation coefficients from an orthogonal ($^\circ$) and also geodetic ($^-$) system to a general system

$$a^\circ{}_r{}^k a^\circ{}_s{}^l \delta^{rs} = g^{kl}, \quad b^\circ{}_k{}^r b^\circ{}_l{}^s \delta_{rs} = g_{kl}, \tag{III.13}$$

$$\partial\bar{a}^\circ{}_s{}^l/\partial\bar{x}^{\circ r} = -\bar{a}^\circ{}_r{}^i \bar{a}^\circ{}_s{}^k \Gamma_{ik}{}^l, \tag{III.14}$$

$$\partial\bar{b}^\circ{}_s{}^l/\partial\bar{x}^{\circ r} = \bar{a}^\circ{}_r{}^k \bar{b}^\circ{}_i{}^l \Gamma_{sk}{}^i. \tag{III.15}$$

It was now shown[3] that to every general coordinate transformation a transformation matrix S^{-1} can be found, which by (II.7) in the scalar case transforms ψ^k as a contravariant vector and ψ^5 as a scalar, and in the vector case transforms ψ^{kl} as an antisymmetrical contravariant tensor of second rank and ψ^{5k} as a contravariant vector. By means of the construction method indicated above, this S^{-1} gives the following expressions for β^k and Γ_k:

$$\beta^k = \beta^{\circ k}\omega^\circ{}_4 + g^{kl}\beta^\circ{}_l\omega^\circ{}_5, \tag{III.16}$$

$$\Gamma_k = -\Gamma_{kl}{}^r[\beta^\circ{}_r\beta^{\circ l}\omega^\circ{}_4 - (\beta^{\circ l}\beta^\circ{}_r - g_r{}^l)\omega^\circ{}_5]. \tag{III.17}$$

IV. KEMMER WAVE EQUATION WITHOUT Γ_k

In a recent paper, A. da Silveira[7] has treated the Kemmer wave equation in Riemann space by means of the four-leg formalism. It can be seen from this treatment that the generalized Kemmer equation can be cast in a form, where the Γ_k:s of (II.1) have disappeared. It is the aim of the present paper to show that this can be afforded by another choice of S than that used before.[3]

We introduce a transformation matrix S^{-1} which, in the scalar case, transforms ψ^k as a contravariant vector density and ψ^5 as a scalar, and which, in the vector case, transforms ψ^{kl} as a contravariant antisymmetrical tensor density of second rank and ψ_{5k} as a covariant vector. This S^{-1} is given by the expression

$$S^{-1} = (g'/g)^{\frac{1}{4}} a_l{}^k \beta^\circ{}_k \beta^{\circ l} \omega^\circ{}_4 \omega^\circ{}_S + \omega^\circ{}_5\omega^\circ{}_S + \tfrac{1}{2}(g'/g)^{\frac{1}{4}} a_l{}^k a_l{}^r \beta^\circ{}_k (\beta^{\circ l}\beta^\circ{}_r - g_r{}^l)\beta^\circ{}_t \omega^\circ{}_4\omega^\circ{}_V - b_l{}^k(\beta^\circ{}_k\beta^{\circ l} - g_k{}^l)\omega^\circ{}_5\omega^\circ{}_V. \tag{IV.1}$$

In fact, (II.7) gives, by means of (IV.1) and (III.3), for the scalar case,

$$\psi^{k'}=(g'/g)^{\frac{1}{2}}a_l{}^k\psi^l, \quad \psi^{5'}=\psi^5; \tag{IV.2}$$

and by means of (IV.1) and (III.4), for the vector case,

$$\psi^{kl'}=(g'/g)^{\frac{1}{2}}a_r{}^k a_s{}^l\psi^{rs}, \quad \psi_{5k}{}'=b_k{}^l\psi_{5l}; \tag{IV.3}$$

which are the required transformation properties.

The procedure for the construction of β^k and Γ_k, indicated in Sec. III, now gives, after some calculations,

$$\begin{aligned}\beta^k=&\beta^{\circ k}\omega^\circ{}_4\omega^\circ{}_S/\sqrt{g}+g^{kr}\beta^\circ{}_r\omega^\circ{}_5\omega^\circ{}_S\sqrt{g}\\&+g_{st}(\beta^{\circ k}\beta^{\circ s}-\beta^{\circ s}\beta^{\circ k})\beta^{\circ t}\omega^\circ{}_4\omega^\circ{}_V/\sqrt{g}\\&-g^{kr}g^{st}\beta^\circ{}_t(\beta^\circ{}_r\beta^\circ{}_s-\beta^\circ{}_s\beta^\circ{}_r)\omega^\circ{}_5\omega^\circ{}_V\sqrt{g},\end{aligned} \tag{IV.4}$$

$$\Gamma_k=-\Gamma_{qk}{}^r(\beta^\circ{}_r\beta^{\circ q}-g_r{}^q)=-\Gamma_{qk}{}^r(\beta_r\beta^q-g_r{}^q). \tag{IV.5}$$

A comparison with (III.16), (III.17) shows that the more simple and symmetrical form (IV.5) of Γ_k is bought at the expense of the more complicated form (IV.4) of β^k.

As could be foreseen, each Γ_k does not vanish, but if we calculate the expression $\beta^k\Gamma_k$ of (II.1), we obtain by (II.3)

$$\begin{aligned}\beta^k\Gamma_k=&-\Gamma_{qk}{}^r\beta^k(\beta_r\beta^q-g_r{}^q)\\=&-\tfrac{1}{2}\Gamma_{qk}{}^r(\beta^k\beta_r\beta^q+\beta^q\beta_r\beta^k-\beta^k g_r{}^q-\beta^q g_r{}^k)=0.\end{aligned} \tag{IV.6}$$

Consequently, with the present representation, the Γ_k:s disappear from the generalized Kemmer wave equation.

By means of (II.11) we can as before[3] construct the corresponding representation of the metric form β.

V. WAVE EQUATION IN COMPONENT FORM

With the representation, constructed in Sec. IV, the wave equation can be written

$$(\beta^k\partial_k-\mu c/\hbar)\psi=0. \tag{V.1}$$

In special relativity we know[6,9] that the most general form of μ is

$$\mu=\mu_4\omega^\circ{}_4+\mu_5\omega^\circ{}_5, \tag{V.2}$$

where μ_4 and μ_5 are arbitrary constants. Moreover,[6] if both μ_4 and μ_5 are different from zero, they can, without lack of generality, be assumed to have the same value. These statements are true also in Riemann space, since both the coordinate transformation and the similarity transformation leave $\omega^\circ{}_4$, $\omega^\circ{}_5$ and thus also μ unchanged.

If we insert (IV.4) and (V.2) in (V.1) we obtain by (III.3) for the scalar case:

$$\frac{1}{g^{\frac{1}{2}}}\partial_k(g^{\frac{1}{2}}U^k)-\frac{i\mu_5 c}{\hbar}U=0, \tag{V.3}$$

$$g^{kl}\partial_l U+\frac{i\mu_4 c}{\hbar}U^k=0, \tag{V.4}$$

where we, according to the transformation properties (IV.2), have put

$$\psi^k=U^k\sqrt{g}, \quad \psi^5=U. \tag{V.5}$$

By means of (III.4), we obtain for the vector case

$$g^{kr}g^{ls}(\partial_r U_s-\partial_s U_r)+\frac{i\mu_4 c}{\hbar}U^{kl}=0, \tag{V.6}$$

$$\frac{1}{g^{\frac{1}{2}}}g_{ks}\partial_r(g^{\frac{1}{2}}U^{rs})-\frac{i\mu_5 c}{\hbar}U_k=0, \tag{V.7}$$

where we, according to (IV.3), have put

$$\psi^{kl}=g^{\frac{1}{2}}U^{kl}, \quad \psi_{5k}=U_k. \tag{V.8}$$

From these component equations, it is seen that the possibility of removing the Γ_k:s from the Kemmer equation is identical with the fact, well-known from Maxwell's equations, that, as far as only scalars, vectors, and antisymmetrical tensors are involved, the field equations in Riemann space can, by suitable introduction of scalar, vector, and tensor densities, be written with only ordinary derivatives without any explicit introduction of the affinity $\Gamma_{kl}{}^i$. Consequently, this possibility does not exist for particles with higher integer spin than 1, e.g., spin 2.[8]

It is also seen from (V.6) and (V.7) that the photon zero mass case can, just as in special relativity,[6,9] be obtained by putting $\mu_5=0$, without any questionable limiting process.

New Approach to Einstein's Empty Space Field Equations*

E. T. NEWMAN AND L. A. TAMBURINO
University of Pittsburgh, Pittsburgh, Pennsylvania
(Received December 8, 1960)

Tetrad formalism is used to derive a set of 36 scalar field equations which correspond to the ordinary field equations $R_{\mu\nu}=0$. The scalar equations are obtained by beginning with a given Petrov type of empty-space Riemann tensor and applying the Ricci identity to each of the tetrad vectors. The unknowns or field variables become the 24 Ricci rotation coefficients, the number of which can always be reduced by the Bianchi identities and occasionally by tetrad transformations which leave the form of the Riemann tensor invariant. The use of these scalar field equations is illustrated by their application to a degenerate case of Petrov type I. It is believed that by this method all possible solutions of this particular case have been found.

I. INTRODUCTION

PETROV[1] has used a tetrad formalism to classify empty-space Riemann tensors into three algebraic types, the physical significance of which has been discussed by Pirani.[2] This paper introduces a new approach to solving the Einstein field equations of empty space by beginning with a given Petrov class and using a set of scalar equations obtained from the Ricci identity applied to each of the tetrad vectors. In these scalar equations the unknowns or field variables are the 24 Ricci rotation coefficients. This number can always be reduced by employing the Bianchi identities and occasionally by tetrad transformations which leave the form of the Riemann tensor invariant. Finally, the remaining Ricci rotation coefficients are related to the metric (essentially by coordinate conditions) and the scalar equations used to solve for the components of the metric.

The method outlined above will be applied in Sec. III to a degenerate case of Petrov type I.

II. DERIVATION OF FIELD EQUATIONS

Capital Latin letters will be used to label tetrad quantities and the Greek alphabet to label tensor indices. Tetrad vectors are written as $\lambda_{A\mu}$: A is the tetrad label and μ is the vector index. The tetrad system used consists of two null vectors ($A=1,2$) and two space-like vectors ($A=3,4$). They satisfy the following orthogonality relations:

$$\lambda_{A\mu}\lambda_{B}{}^{\mu}=\eta_{AB}=\eta^{AB}, \tag{1}$$

where

$$\eta_{AB}=\eta^{AB}=\begin{Vmatrix}0&1&0&0\\1&0&0&0\\0&0&-1&0\\0&0&0&-1\end{Vmatrix}. \tag{2}$$

η_{AB} and η^{AB} are used to raise and lower tetrad labels.[3]

$$\lambda^{A}{}_{\mu}=\eta^{AB}\lambda_{B\mu},\qquad \lambda_{A}{}^{\mu}=\eta_{AB}\lambda^{B\mu}. \tag{3}$$

The tensor labels are raised and lowered in the usual manner by $g_{\mu\nu}$. It is easily shown[4] that

$$\lambda^{A\mu}\lambda_{B\mu}=\delta^{A}{}_{B}, \tag{4}$$

$$\lambda^{A}{}_{\mu}\lambda_{A\nu}=g_{\mu\nu}, \tag{5}$$

$$\lambda^{A}{}_{\mu}\lambda_{A}{}^{\nu}=\delta_{\mu}{}^{\nu}, \tag{6}$$

$$\lambda^{A\mu}\lambda_{A}{}^{\nu}=g^{\mu\nu}. \tag{7}$$

The Ricci rotation coefficients are the set of scalars defined by

$$\gamma_{A}{}^{MN}=\lambda_{A\mu;\nu}\lambda^{M\mu}\lambda^{N\nu}, \tag{8}$$

where $\lambda_{A\mu;\nu}$ is the covariant derivative of $\lambda_{A\mu}$. This can be rewritten as

$$\lambda_{A\mu;\nu}=\gamma_{A}{}^{MN}\lambda_{M\mu}\lambda_{N\nu}. \tag{9}$$

From (3) and (8), we have

$$\gamma^{ABM}=\eta^{AN}\gamma_{N}{}^{BM}. \tag{10}$$

Covariant differentiation of $\lambda^{A\mu}\lambda^{B}{}_{\mu}=\eta^{AB}$ yields the following identities,

$$\gamma^{ABM}=-\gamma^{BAM}. \tag{11}$$

The total number of independent γ's is easily seen to be 24.

The mixed curvature tensor is introduced by the Ricci identity

$$\lambda_{A\rho}R^{\rho}{}_{\mu\alpha\beta}=\lambda_{A\mu;\alpha\beta}-\lambda_{A\mu;\beta\alpha}. \tag{12}$$

The following expression for the covariant curvature tensor results from multiplying (12) by $\lambda^{A}{}_{\sigma}$ and substituting (9):

$$R_{\sigma\mu\alpha\beta}=\lambda^{A}{}_{\sigma}[(\gamma_{A}{}^{MN}\lambda_{M\mu}\lambda_{N\alpha})_{;\beta}-(\gamma_{A}{}^{MN}\lambda_{M\mu}\lambda_{N\beta})_{;\alpha}]. \tag{13}$$

By using (8), (9), (11), and rearranging terms, Eq. (13) may be re-expressed as follows:

$$R_{\sigma\mu\alpha\beta}=2\lambda_{A\sigma}\lambda_{B\mu}\lambda_{[M\alpha}\lambda_{N]\beta}\times[\gamma^{AKM}\gamma_{K}{}^{BN}+\gamma^{ABK}\gamma_{K}{}^{MN}+\gamma^{ABM:N}], \tag{14}$$

where[5]

$$\lambda_{[M\alpha}\lambda_{N]\beta}=\tfrac{1}{2}(\lambda_{M\alpha}\lambda_{N\beta}-\lambda_{M\beta}\lambda_{N\alpha}),$$

$$\gamma^{ABM:N}=(\gamma^{ABM})_{;\rho}\lambda^{N\rho}.$$

* This work was supported by the Aeronautical Research Laboratories, Wright-Patterson Air Force Base.

[1] A. Z. Petrov, Sci. Not. Kazan State Univ. **114**, 55 (1954).

[2] F. A. E. Pirani, Acta Phys. Polon. **XV**, 389 (1956); Phys. Rev. **105**, 1089 (1957).

[3] Summation convention is to be understood for tetrad indices as well as for tensor indices.

[4] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1960), Chap. 3.

[5] $\gamma^{ABM:N}$ is called the "intrinsic derivative."

By noting that the last two terms in brackets in (14) are antisymmetrical in A and B, and using the identity

$$\lambda_{[M\alpha}\lambda_{N]\beta}(\gamma^{AKM}\gamma_K{}^{BN}-\gamma_K{}^{AN}\gamma^{KBM})=0,$$

it follows that

$$R_{\sigma\mu\alpha\beta}=2\lambda_{[A\sigma}\lambda_{B]\mu}\lambda_{[M\alpha}\lambda_{N]\beta} \times(\gamma^{AKM}\gamma_K{}^{BN}+\gamma^{ABK}\gamma_K{}^{MN}+\gamma^{ABM:N}). \quad (15)$$

The sum over $ABMN$ consists of 144 terms. It will be convenient to rewrite (15) so that the sum is over ordered pairs of A and B and M and N, the summation now consisting of 36 terms. This can be effected by the following convention. Let

$$\lambda_{[A\sigma}\lambda_{B]\mu}\rightarrow M_{s[\sigma\mu]}$$
$$\lambda_{[M\alpha}\lambda_{N]\beta}\rightarrow M_{t[\alpha\beta]};$$

ordered pairs AB; MN: 13 41 23 24 34 12
s ; t : 1 2 3 4 5 6.

Equation (15) then becomes

$$R_{\sigma\mu\alpha\beta}=M_{s[\sigma\mu]}M_{t[\alpha\beta]}V^{st}, \quad (16)$$

where

$$V^{st}\equiv 4(\gamma_K{}^{AN}\gamma^{KBM}-\gamma_K{}^{AM}\gamma^{KBN}+\gamma^{ABK}\gamma_K{}^{MN} -\gamma^{ABK}\gamma_K{}^{NM}+\gamma^{ABM:N}-\gamma^{ABN:M}). \quad (17)$$

From another point of view, all three Petrov types of Riemann tensors can be written in the form[6]

$$R_{\sigma\mu\alpha\beta}=M_{s[\sigma\mu]}M_{t[\alpha\beta]}U^{st}, \quad (18)$$

where

$$U^{st}=U^{ts}.$$

A particular Petrov type is specified by a set of symmetrical coefficients U^{st} of which many are equal to zero.[7]

Equating the coefficients from Eqs. (16) and (18), we now have 36 scalar equations

$$V^{st}=U^{st}, \quad (19)$$

which we shall call the scalar field equations (SFE). The 36 SFE consist of a group of equations for which the left-hand side of (19) equals the nonvanishing scalars U^{st} and a large group of equations for which the left-hand side equals zero.

Finally, the Riemann tensor must satisfy the Bianchi identities which are written as

$$R^{\alpha}{}_{\mu\nu\rho;\alpha}=0 \quad (20)$$

because the Ricci tensor $R_{\mu\nu}=0$. This yields further conditions for the γ's which augment the 36 field equations.

[6] E. T. Newman, J. Math. Phys. 2, 327 (1961).
[7] For example, Petrov type II null is specified by only two nonvanishing coefficients, i.e., U^{11} and U^{22}.

III. APPLICATION TO A SPECIAL CASE OF PETROV TYPE I

The special case of Petrov type I which we shall consider, expressed in the form of (18), is[6]

$$R=\rho[M_1M_3+M_3M_1-M_2M_4-M_4M_2 +2(M_5M_5-M_6M_6)], \quad (21)$$

where the tensor indices are suppressed and ρ is a scalar. Therefore we have, by using (19), the following SFE:

$$\begin{aligned} V^{13}&=V^{31}=\rho \\ V^{24}&=V^{42}=-\rho \\ V^{55}&=2\rho \\ V^{66}&=-2\rho; \end{aligned} \quad (22a)$$

all other

$$V^{st}=0. \quad (22b)$$

It will be expedient to display the γ's in the form of matrices. By incorporating the identities (11) we obtain four rotation matrices:

$$\|\gamma_1{}^{AB}\|=\begin{Vmatrix} \gamma_1{}^{11} & \gamma_1{}^{12} & \gamma_1{}^{13} & \gamma_1{}^{14} \\ 0 & 0 & 0 & 0 \\ \gamma_1{}^{31} & \gamma_1{}^{32} & \gamma_1{}^{33} & \gamma_1{}^{34} \\ \gamma_1{}^{41} & \gamma_1{}^{42} & \gamma_1{}^{43} & \gamma_1{}^{44} \end{Vmatrix};$$

$$\|\gamma_2{}^{AB}\|=\begin{Vmatrix} 0 & 0 & 0 & 0 \\ -\gamma_1{}^{11} & -\gamma_1{}^{12} & -\gamma_1{}^{13} & -\gamma_1{}^{14} \\ \gamma_2{}^{31} & \gamma_2{}^{32} & \gamma_2{}^{33} & \gamma_2{}^{34} \\ \gamma_2{}^{41} & \gamma_2{}^{42} & \gamma_2{}^{43} & \gamma_2{}^{44} \end{Vmatrix};$$

$$\|\gamma_3{}^{AB}\|=\begin{Vmatrix} \gamma_2{}^{31} & \gamma_2{}^{32} & \gamma_2{}^{33} & \gamma_2{}^{34} \\ \gamma_1{}^{31} & \gamma_1{}^{32} & \gamma_1{}^{33} & \gamma_1{}^{34} \\ 0 & 0 & 0 & 0 \\ \gamma_3{}^{41} & \gamma_3{}^{42} & \gamma_3{}^{43} & \gamma_3{}^{44} \end{Vmatrix};$$

$$\|\gamma_4{}^{AB}\|=\begin{Vmatrix} \gamma_2{}^{41} & \gamma_2{}^{42} & \gamma_2{}^{43} & \gamma_2{}^{44} \\ \gamma_1{}^{41} & \gamma_1{}^{42} & \gamma_1{}^{43} & \gamma_1{}^{44} \\ -\gamma_3{}^{41} & -\gamma_3{}^{42} & -\gamma_3{}^{43} & -\gamma_3{}^{44} \\ 0 & 0 & 0 & 0 \end{Vmatrix};$$

with A designating the rows and B the columns.

By inspection of the matrices, one can perceive the identities, e.g., $\|\gamma_2{}^{AB}\|$:

$$\gamma_2{}^{1N}=0, \quad \gamma_2{}^{2N}=-\gamma_1{}^{1N}$$

or

$$\gamma^{11N}=0, \quad \gamma^{12N}=-\gamma^{21N}.$$

Also, by inspection of all four rotation matrices, it is seen that there are 24 independent γ's.

Further identities are obtained by requiring that (21) satisfy the Bianchi identities (20).[8] The results are

[8] R. Sachs (private communication).

summerized below:

$$\|\gamma_1{}^{AB}\| = \begin{Vmatrix} \gamma_1{}^{11} & \gamma_1{}^{12} & \gamma_1{}^{13} & \gamma_1{}^{14} \\ 0 & 0 & 0 & 0 \\ \gamma_1{}^{31} & 0 & \gamma_1{}^{33} & 0 \\ \gamma_1{}^{41} & 0 & 0 & \gamma_1{}^{33} \end{Vmatrix};$$

$$\|\gamma_2{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ -\gamma_1{}^{11} & -\gamma_1{}^{12} & -\gamma_1{}^{13} & -\gamma_1{}^{14} \\ 0 & \gamma_1{}^{31} & \gamma_2{}^{33} & 0 \\ 0 & \gamma_1{}^{41} & 0 & \gamma_2{}^{33} \end{Vmatrix};$$

$$\|\gamma_3{}^{AB}\| = \begin{Vmatrix} 0 & \gamma_1{}^{31} & \gamma_2{}^{33} & 0 \\ \gamma_1{}^{31} & 0 & \gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & 0 \\ \gamma_3{}^{41} & \gamma_3{}^{42} & \gamma_3{}^{43} & \gamma_3{}^{44} \end{Vmatrix};$$

$$\|\gamma_4{}^{AB}\| = \begin{Vmatrix} 0 & \gamma_1{}^{41} & 0 & \gamma_2{}^{33} \\ \gamma_1{}^{41} & 0 & 0 & \gamma_1{}^{33} \\ -\gamma_3{}^{41} & -\gamma_3{}^{42} & -\gamma_3{}^{43} & -\gamma_3{}^{44} \\ 0 & 0 & 0 & 0 \end{Vmatrix}.$$

There are four differential Bianchi identities relating ρ and four γ's:

$$3\gamma_2{}^{33} = (\ln\rho)^{:1}; \quad 3\gamma_1{}^{33} = (\ln\rho)^{:2}; \quad 3\gamma_1{}^{31} = (\ln\rho)^{:3}; \quad 3\gamma_1{}^{41} = (\ln\rho)^{:4}. \tag{23}$$

There are now 12 independent γ's.

There exist two one-parameter transformations (rotations) of the tetrad vectors which leave the form of (21) invariant. It will be shown how a judicious choice of parameters and use of the SFE greatly reduce the number of independent γ's. From a combined study of rotation and scalar field equations, it has been found that there exist three possible sets of initial assumptions concerning the vanishing of certain γ's, for which the SFE would support nonflat space solutions. The details of the rotations and their applications in connection with the SFE are discussed in the Appendix. In this connection, the SFE were used at times to prove integrability conditions for the existence of particular transformation parameters which yielded simplifying relationships among the γ's. Secondly, after a specific choice of rotation parameters, frequently the number of independent γ's were reduced directly by linear combinations of a few SFE. The results obtained in the Appendix are summarized below. It will be noted that, in each case, the number of independent γ's have been reduced to four.

Case I.

$$\|\gamma_1{}^{AB}\| = \begin{Vmatrix} \gamma_1{}^{11} & \gamma_1{}^{11} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \gamma_1{}^{31} & 0 & \gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & \gamma_1{}^{33} \end{Vmatrix};$$

$$\|\gamma_2{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ -\gamma_1{}^{11} & -\gamma_1{}^{11} & 0 & 0 \\ 0 & \gamma_1{}^{31} & -\gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & -\gamma_1{}^{33} \end{Vmatrix};$$

$$\|\gamma_3{}^{AB}\| = \begin{Vmatrix} 0 & \gamma_1{}^{31} & -\gamma_1{}^{33} & 0 \\ \gamma_1{}^{31} & 0 & \gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_3{}^{44} \end{Vmatrix};$$

$$\|\gamma_4{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & -\gamma_1{}^{33} \\ 0 & 0 & 0 & \gamma_1{}^{33} \\ 0 & 0 & 0 & -\gamma_3{}^{44} \\ 0 & 0 & 0 & 0 \end{Vmatrix}.$$

The four differential Bianchi identities are

$$3\gamma_1{}^{33} = -(\ln\rho)^{:1} = (\ln\rho)^{:2}; \quad 3\gamma_1{}^{31} = (\ln\rho)^{:3}; \quad \rho^{:4} = 0.$$

Case II.

$$\|\gamma_1{}^{AB}\| = \begin{Vmatrix} \gamma_1{}^{11} & \gamma_1{}^{12} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \gamma_1{}^{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix};$$

$$\|\gamma_2{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ -\gamma_1{}^{11} & -\gamma_1{}^{12} & 0 & 0 \\ 0 & \gamma_1{}^{31} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix};$$

$$\|\gamma_3{}^{AB}\| = \begin{Vmatrix} 0 & \gamma_1{}^{31} & 0 & 0 \\ \gamma_1{}^{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_3{}^{44} \end{Vmatrix};$$

$$\|\gamma_4{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\gamma_3{}^{44} \\ 0 & 0 & 0 & 0 \end{Vmatrix}.$$

The four differential Bianchi identities are

$$3\gamma_1{}^{31} = (\ln\rho)^{:3}; \quad \rho^{:1} = \rho^{:2} = \rho^{:4} = 0.$$

Case III.

$$\|\gamma_1{}^{AB}\| = \begin{Vmatrix} \gamma_1{}^{11} & \gamma_1{}^{11} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & \gamma_1{}^{33} \end{Vmatrix};$$

$$\|\gamma_2{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & 0 \\ -\gamma_1{}^{11} & -\gamma_1{}^{11} & 0 & 0 \\ 0 & 0 & -\gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & -\gamma_1{}^{33} \end{Vmatrix};$$

$$\|\gamma_3{}^{AB}\| = \begin{Vmatrix} 0 & 0 & -\gamma_1{}^{33} & 0 \\ 0 & 0 & \gamma_1{}^{33} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_3{}^{43} & \gamma_3{}^{44} \end{Vmatrix};$$

$$\|\gamma_4{}^{AB}\| = \begin{Vmatrix} 0 & 0 & 0 & -\gamma_1{}^{33} \\ 0 & 0 & 0 & \gamma_1{}^{33} \\ 0 & 0 & -\gamma_3{}^{43} & -\gamma_3{}^{44} \\ 0 & 0 & 0 & 0 \end{Vmatrix}.$$

The four differential Bianchi identities are

$$3\gamma_1{}^{33}{}_{;1}=-(\ln\rho)_{;1}{}^{;1}=(\ln\rho)_{;2}{}^{;2};$$
$$\rho^{;3}=\rho^{;4}=0.$$

An important theorem that greatly simplifies the calculation of the metric is the following[4]:

Necessary and sufficient conditions for

$$\lambda_{N\mu}=\zeta\phi_{;\mu}, \quad \text{(where } \zeta \text{ and } \phi \text{ are scalars),}$$

are

$$\gamma_N{}^{AB}=\gamma_N{}^{BA}, \quad (A, B=1, \ldots N-1).$$

By a coordinate transformation, $\tilde{\chi}_N=\phi(x\text{'s})$, one can set

$$\lambda_{N\mu}=\zeta\delta_{N\mu}.$$

It is seen by inspection that in each of the three cases all four of the tetrad vectors satisfy the conditions of the theorem, so that in each case we have:

$$\lambda_{N\mu}=\zeta_N\delta_{N\mu}, \quad (N=1, 2, 3, 4).$$

We can now use (5) to express the covariant metric in terms of the four scalars (ζ_N's),

$$g_{\mu\nu}=\eta^{AB}\lambda_{B\mu}\lambda_{A\nu}=\begin{Vmatrix} 0 & \zeta_1\zeta_2 & 0 & 0 \\ \zeta_1\zeta_2 & 0 & 0 & 0 \\ 0 & 0 & -(\zeta_3)^2 & 0 \\ 0 & 0 & 0 & -(\zeta_4)^2 \end{Vmatrix}. \tag{24}$$

It is then a straightforward calculation to obtain the Christoffel symbols and to relate the γ's to the four scalars ζ_N by substituting the covariant derivative $\lambda_{A\mu;\nu}$ into (8). Thus, the relationship between the rotation coefficients and the metric (the four scalars) has now been established and it remains to solve the field equations which are a set of simultaneous differential equations involving five scalars, i.e., the four tetrad scalars and ρ.

In our calculations, instead of relating the γ's to the nondiagonal metric (24) discussed in the preceeding paragraph, it was found convenient to relate the same γ's to a diagonal metric. This was accomplished by defining a new tetrad system expressed below in terms of the original tetrad:

$$\tilde{\lambda}_{0\mu}=(1/\sqrt{2})(\lambda_{1\mu}+\lambda_{2\mu}); \quad \tilde{\lambda}_{2\mu}=\lambda_{3\mu};$$
$$\tilde{\lambda}_{1\mu}=(1/\sqrt{2})(\lambda_{1\mu}-\lambda_{2\mu}); \quad \tilde{\lambda}_{3\mu}=\lambda_{4\mu}. \tag{25}$$

The new $\tilde{\gamma}$'s are linear functions of the γ's and also satisfy in each of the three cases the conditions of the theorem for all four vectors. Now we have

$$\tilde{\lambda}_{N\mu}=e_N H_N\delta_{N\mu}, \quad (N=0, 1, 2, 3),$$

where

$$e_N=1, \quad N=0$$
$$-1, \quad N=1, 2, 3,$$

and the H's are scalars.

The formalism of Eqs. (1)–(11) remains the same in the new system except that now

$$\tilde{\eta}_{AB}=\begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{Vmatrix} \quad (A, B=0, 1, 2, 3,),$$

and the tetrad consists of one time-like and three space-like vectors. It is easily shown that[9]

$$g_{\mu\mu}=(g^{\mu\mu})^{-1}=e_N(H_N)^2\delta_{N\mu};$$
$$g_{\mu\nu}=g^{\mu\nu}=0 \quad (\mu\neq\nu), \tag{27}$$

and

$$\tilde{\gamma}_M{}^{NM}=\frac{e_N(\ln H_N)_{,M}}{H_M} \quad (N\neq M). \tag{28}$$

With the use of identities (11), it can easily be seen that Eqs. (28) include all of the independent $\tilde{\gamma}$'s. Finally, we can equate the original γ's to functions of the H's, thus associating them with the diagonal metric (27), by substituting into Eqs. (28) the $\tilde{\gamma}$'s as linear functions of the γ's. Because there are only four independent γ's in each case, a number of the 12 $\tilde{\gamma}_M{}^{NM}$'s in Eqs. (28) are equal to zero, in which cases $H_{N;M}=0$.

TABLE I.

	V	W	h_0	h_1	h_2	h_3
Case I	$\neq 0$	$\neq 0$	$h_0(x_1)$	$h_1(x_1)$	$h_2(x_2)$	$h_3(x_2)$
Case II	$=0$	$\neq 0$	$h_0(x_0,x_1)$	$h_1(x_0,x_1)$	$h_2(x_2)$	$h_3(x_2)$
Case III	$\neq 0$	$=0$	$h_0(x_1)$	$h_1(x_1)$	$h_2(x_2,x_3)$	$h_3(x_2,x_3)$

(29d)

The γ's are now substituted into scalar equations (22b), which have been greatly reduced in number, and into the differential Bianchi identities. These two sets of equations are readily solved by simple integration and yield the functional forms of the five scalars. The resulting forms for each case are:

$$H_0=\beta e^{h_0}; \quad H_2=\beta e^{h_2}; \quad \rho=2K\beta^{-3};$$
$$H_1=\beta e^{h_1}; \quad H_3=\beta e^{h_3}, \tag{29a}$$

where

$$\beta=[V(x_1)+W(x_2)]^{-1};$$
$$\epsilon_1 e^{h_1}=e^{-h_0}V_{,1}; \tag{29b}$$
$$\epsilon_2 e^{h_2}=e^{-h_3}W_{,2};$$

and

K is a nonvanishing constant;

$$\epsilon_1=1, \quad V\neq 0$$
$$0, \quad V=0; \tag{29c}$$
$$\epsilon_2=1, \quad W\neq 0$$
$$0, \quad W=0.$$

Equations (29) and Table I establish the functional forms for each case.

[9] This notation is similar to Eisenhart's (see reference 4).

The functional forms (29) are then substituted into Eqs. (22a), the number of which has been reduced to three in each case. Before discussing the solutions of these scalar equations, it is an interesting digression to compare them with the ordinary field equations

$$R_{\mu\nu}=0, \tag{30}$$

where $R_{\mu\nu}$ is the Ricci tensor. When (29) were substituted into Eqs. (30), it was found that all but two of these equations were satisfied identically and that these remaining two were linear combinations of the three SFE (22a).

These remaining equations (22a), shown to be equivalent to the ordinary field equations (30), proved to be advantageous for it was found that one of our three SFE was separable and also the first integral of the other two. In the following list are the solutions to these separable equations where a and b are separation constants and K is as defined above.

Case I.

$$\begin{aligned} e^{2h_0}&=KV^3+aV^2+bV \\ e^{2h_3}&=KW^3-aW^2+bW. \end{aligned} \tag{31a}$$

Case II.

$$\begin{aligned} &e^{-2h_1}[h_{0,11}+h_{0,1}(h_{0,1}-h_{1,1})] \\ &\qquad -e^{-2h_0}[h_{1,00}+h_{1,0}(h_{1,0}-h_{0,0})]=a, \\ &\qquad e^{2h_3}=KW^3-aW^2. \end{aligned} \tag{31b}$$

Case III.

$$\begin{aligned} &e^{2h_0}=KV^3+aV^2 \\ &e^{-2h_2}[h_{3,22}+h_{3,2}(h_{3,2}-h_{2,2})] \\ &\qquad +e^{-2h_3}[h_{2,33}+h_{2,3}(h_{2,3}-h_{3,3})]=-a. \end{aligned} \tag{31c}$$

We can always carry out a coordinate transformation such that

$$\begin{aligned} \bar{V}(\bar{x}_1)&=(\bar{x}_1)^{-1}=V(x_1), \\ \bar{W}(\bar{x}_2)&=(\bar{x}_2)^{-1}=W(x_2), \end{aligned}$$

for cases in which V and W are not equal to zero. Therefore, setting $V=(x_1)^{-1}$ and $W=(x_2)^{-1}$ is equivalent to picking a particular coordinate system in which we shall write the solution. Combining (29) and (31) we obtain the following metrics:

Case I.

$$\begin{aligned} g_{00}&=\frac{(x_2)^2}{\Delta}\left(\frac{K}{x_1}+a+bx_1\right) \\ g_{11}&=-\frac{(x_2)^2}{\Delta}\left(\frac{K}{x_1}+a+bx_1\right)^{-1}, \\ g_{22}&=-\frac{(x_1)^2}{\Delta}\left(\frac{K}{x_2}-a+bx_2\right)^{-1} \\ g_{33}&=-\frac{(x_1)^2}{\Delta}\left(\frac{K}{x_2}-a+bx_2\right), \end{aligned} \tag{32}$$

where

$$\Delta=(x_1+x_2)^2.$$

Case II.

$$\begin{aligned} g_{00}&=(x_2)^2e^{2h_0}, \qquad g_{22}=-\left(\frac{K}{x_2}-a\right)^{-1}, \\ g_{11}&=-(x_2)^2e^{2h_1}, \quad g_{33}=-\left(\frac{K}{x_2}-a\right), \end{aligned} \tag{33}$$

where h_0 and h_1 must satisfy

$$\begin{aligned} a=&e^{-2h_1}[h_{0,11}+h_{0,1}(h_{0,1}-h_{1,1})] \\ &-e^{-2h_0}[h_{1,00}+h_{1,0}(h_{1,0}-h_{0,0})]. \end{aligned}$$

Case III.

$$\begin{aligned} g_{00}&=\frac{K}{x_1}+a, \qquad g_{22}=-(x_1)^2e^{2h_2}, \\ g_{11}&=-\left(\frac{K}{x_1}+a\right)^{-1}, \quad g_{33}=-(x_1)^2e^{2h_3}, \end{aligned} \tag{34}$$

where h_2 and h_3 must satisfy

$$\begin{aligned} -a=&e^{-2h_2}[h_{3,22}+h_{3,2}(h_{3,2}-h_{2,2})] \\ &+e^{-2h_3}[h_{2,33}+h_{2,3}(h_{2,3}-h_{3,3})]. \end{aligned}$$

IV. DISCUSSION

Further study of rotations and SFE reveals that there are initial assumptions different from the three cases considered in the Appendix which will also support nonflat space solutions; however, they can be shown to be equivalent to or a subcase of one of the solutions of (32), (33), or (34). We believe that all possible initial assumptions have been exhausted and hence all solutions to (21) have been obtained.

We will consider six particular metrics from Cases II and III. The first three metrics are from Case II and the remaining ones from Case III.

Letting $h_0=h_1=f(x_1\pm x_0)$, where f is arbitrary, we obtain from (33)

$$ds^2=(x_2)^2e^{2f}(dx_0^2-dx_1^2)-\frac{x_2}{K}dx_2^2-\frac{K}{x_2}dx_3^2. \tag{35}$$

Letting $h_0=0$ and $h_1=\ln(\sin x_0)$, we obtain from (33)

$$\begin{aligned} ds^2=&(x_2)^2[dx_0^2-(\sin^2x_0)dx_1^2] \\ &-\left(\frac{K}{x_2}-1\right)^{-1}dx_2^2-\left(\frac{K}{x_2}-1\right)dx_3^2. \end{aligned} \tag{36}$$

Letting $h_0=0$ and $h_1=\ln(\sinh x_0)$, we obtain from (33)

$$\begin{aligned} ds^2=&(x_2)^2[dx_0^2-(\sinh x_0)^2dx_1^2] \\ &-\left(\frac{K}{x_2}+1\right)^{-1}dx_2^2-\left(\frac{K}{x_2}+1\right)dx_3^2. \end{aligned} \tag{37}$$

Letting $h_2=0$ and $h_3=\ln(\sin x_2)$, we obtain from (34)

$$ds^2=\left(1+\frac{K}{x_1}\right)dx_0^2-\left(1+\frac{K}{x_1}\right)^{-1}dx_1^2 - (x_1)^2[dx_2^2+(\sin x_2)^2dx_3^2]. \quad (38)$$

Letting $h_2=0$ and $h_3=\ln(\sinh x_2)$, we obtain from (34)

$$ds^2=\left(\frac{K}{x_1}-1\right)dx_0^2-\left(\frac{K}{x_1}-1\right)^{-1}dx_1^2 - (x_1)^2[dx_2^2+(\sinh x_2)^2dx_3^2]. \quad (39)$$

Letting $h_2=h_3=0$, we obtain from (34)

$$ds^2=\frac{K}{x_1}dx_0^2-\frac{x_1}{K}dx_1^2-(x_1)^2(dx_2^2+dx_3^2). \quad (40)$$

Metrics (36)–(39) are equivalent to B. K. Harrison's metrics III-8, III-7, III-9, and III-10, respectively.[10] Metric (38) is the familiar form of the Schwarzschild solution. Metric (40) can be related to Harrison's metric III-1 by a complex coordinate transformation.

It should be noted that only for solutions from Case II can there be vanishing divergencies of the null vectors defined by (1) and (2). The restriction needed for one of the null vectors to have a vanishing divergence is

$$(e^{h_1})_{,0}=\pm(e^{h_0})_{,1} \quad (41)$$

where $\lambda_1{}^{\mu}{}_{;\mu}=0$, + sign and $\lambda_2{}^{\mu}{}_{;\mu}=0$, − sign. If (41) is substituted into (33), we can rewrite (33) as

$$a=\pm e^{-(h_0+h_1)}(h_{1,01}-h_{0,10}). \quad (42)$$

Metric (35) is a solution in which one of the null vectors has a vanishing divergence. When $h_0=h_1=f(x_0\pm x_1)$, it is seen to satisfy Eqs. (41) and (42). A special case of (35) is $f=0$ in which both null vectors have vanishing divergencies. Furthermore, if $f\neq 0$, it can be set equal to zero by coordinate transformations.[11] Consider, for example, $f=f(x_1+x_0)$.

The transformation

$$\xi=x_1+x_0, \quad \eta=x_1-x_0$$

gives the metric

$$ds^2=\tfrac{1}{2}(x_2)^2e^{2f(\xi)}d\xi d\eta-\frac{x_2}{K}dx_2^2-\frac{K}{x_2}dx_3^2.$$

Now let $d\bar{\xi}=e^{2f(\xi)}d\xi$, which yields

$$ds^2=\frac{(x_2)^2}{2}d\bar{\xi}d\eta-\frac{x_2}{K}dx_2^2-\frac{K}{x_2}dx_3^2.$$

The formalism of Sec. I was also applied to two other degenerate cases of Petrov type I. The first case can be expressed as

$$R=\rho(\lambda R^V+R^{V*}), \quad (43)$$

where λ equals zero or a real constant. To facilitate writing, tensor (21) was designated as

$$R=\rho R^V \quad (21')$$

and its dual by

$$R^*=\rho R^{V*},$$

where the dual is defined by

$$R_{\alpha\beta\gamma\delta}=\tfrac{1}{2}(-g)^{\frac{1}{2}}\epsilon_{\gamma\delta\mu\nu}R_{\alpha\beta}{}^{\mu\nu}.$$

This dual of R^V is[6]

$$R^{V*}=M_1M_4+M_4M_1+M_2M_3+M_3M_2+2(M_5M_6+M_6M_5). \quad (44)$$

It was readily found, by study of the Bianchi identities and several of the SFE, that the only solutions to (43) were flat-space solutions.

The second case considered can be expressed as

$$R=\alpha(R^{\mathrm{I}}+R^{\mathrm{II}})+\beta(R^{\mathrm{I}*}+R^{\mathrm{II}*}), \quad (45)$$

where

$$R^{\mathrm{I}}=M_1M_1-M_2M_2$$

and

$$R^{\mathrm{II}}=M_3M_3-M_4M_4.$$

It is possible to write the Riemann tensor of type II-null as

$$R=\sigma(\cos\theta R^{\mathrm{I}}+\sin\theta R^{\mathrm{I}*})$$

or

$$R=\sigma(\cos\theta R^{\mathrm{II}}+\sin\theta R^{\mathrm{II}*}).$$

Therefore, (45) is the linear combination of type II-null traveling in opposite directions. It was readily found that the only solutions to (45) were flat-space solutions. (This result was also obtained by J. N. Goldberg and R. P. Kerr.)[12]

APPENDIX

A. Rotation of $\lambda_{1\mu}$ and $\lambda_{2\mu}$

The form of R/ρ, (21), remains invariant under the following transformation[13]:

$$\bar{\lambda}_{1\mu}=\alpha\lambda_{1\mu}, \quad \bar{\lambda}_{2\mu}=\alpha^{-1}\lambda_{2\mu},$$
$$\bar{\lambda}_{3\mu}=\lambda_{3\mu}, \quad \bar{\lambda}_{4\mu}=\lambda_{4\mu}, \quad (A1)$$

where α is an arbitrary scalar.

The following transformation equations for the γ's are obtained by taking the covariant derivatives of (A1), using Eq. (9), and then multiplying both sides of the equations by $\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}$.

$$\bar{\gamma}_1{}^{AB}=\alpha\gamma_1{}^{MN}\lambda_{M\mu}\lambda_{N\nu}\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}+(\ln\alpha)_{,\nu}\bar{\lambda}^{B\nu}\delta_1{}^A,$$
$$\bar{\gamma}_2{}^{AB}=\alpha^{-1}\gamma_2{}^{MN}\lambda_{M\mu}\lambda_{N\nu}\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}-(\ln\alpha)_{,\nu}\bar{\lambda}^{B\nu}\delta_2{}^A, \quad (A2)$$
$$\bar{\gamma}_P{}^{AB}=\gamma_P{}^{MN}\lambda_{M\mu}\lambda_{N\nu}\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}, \quad (P=3,4).$$

[10] B. K. Harrison, Phys. Rev. **116**, 1285 (1959).
[11] Transformations suggested by I. Robinson.
[12] J. N. Goldberg and R. P. Kerr (preprint).
[13] For rotations in Secs. A and B, the transformed entities, designated by bars, satisfy Eqs. (1)–(11).

B. Rotation of $\lambda_{3\mu}$ and $\lambda_{4\mu}$

The second tetrad transformation is

$$\begin{aligned}\bar{\lambda}_{3\mu}&=(\cos\theta)\lambda_{3\mu}+\lambda_{4\mu}\sin\theta, \quad \bar{\lambda}_{1\mu}=\lambda_{1\mu},\\ \bar{\lambda}_{4\mu}&=-\lambda_{3\mu}\sin\theta+\lambda_{4\mu}\cos\theta, \quad \bar{\lambda}_{2\mu}=\lambda_{2\mu},\end{aligned} \tag{B1}$$

where θ is an arbitrary scalar.

It follows that

$$\begin{aligned}\bar{\gamma}_P{}^{AB}&=\gamma_P{}^{MN}\lambda_{M\mu}\lambda_{N\nu}\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}, \quad (P=1,2),\\ \bar{\gamma}_3{}^{AB}&=(\gamma_3{}^{MN}\cos\theta+\gamma_4{}^{MN}\sin\theta)\lambda_{M\mu}\lambda_{N\mu}\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}\\ &\qquad+\theta_{,\nu}\bar{\lambda}^{B\nu}\delta_4{}^A,\\ \bar{\gamma}_4{}^{AB}&=(-\gamma_3{}^{MN}\sin\theta+\gamma_4{}^{MN}\cos\theta)\lambda_{M\mu}\lambda_{N\nu}\bar{\lambda}^{A\mu}\bar{\lambda}^{B\nu}\\ &\qquad-\theta_{,\nu}\bar{\lambda}^{B\nu}\delta_3{}^A.\end{aligned} \tag{B2}$$

C. Case I

The initial set of assumptions is

$$\gamma_1{}^{33}\neq 0, \quad \gamma_2{}^{33}\neq 0, \quad \gamma_1{}^{11}\neq 0,$$
$$\gamma_1{}^{31}\neq 0, \quad \gamma_3{}^{44}\neq 0.$$

From (A2), let us consider

$$\bar{\gamma}_1{}^{33}=\alpha\gamma_1{}^{33}$$

and

$$\bar{\gamma}_2{}^{33}=\alpha^{-1}\gamma_2{}^{33}.$$

Letting

$$\alpha^2=-\gamma_2{}^{33}/\gamma_1{}^{33}$$

is equivalent to setting

$$\bar{\gamma}_1{}^{33}=-\bar{\gamma}_2{}^{33}.$$

With this identity it follows directly, using various combinations of eight of the SFE (22), that

$$\bar{\gamma}_1{}^{11}=\bar{\gamma}_1{}^{12}, \quad \bar{\gamma}_1{}^{13}=0, \quad \text{and} \quad \bar{\gamma}_1{}^{14}=0.$$

From (B2), let us consider

$$\bar{\gamma}_1{}^{31}=\gamma_1{}^{31}\cos\theta+\gamma_1{}^{41}\sin\theta,$$

and

$$\bar{\gamma}_1{}^{41}=-\gamma_1{}^{31}\sin\theta+\gamma_1{}^{41}\cos\theta.$$

Letting

$$\tan\theta=\gamma_1{}^{41}/\gamma_1{}^{31}$$

is equivalent to setting

$$\bar{\gamma}_1{}^{41}=0.$$

Because of the initial assumption $\gamma_1{}^{31}\neq 0$, which implies $\bar{\gamma}_1{}^{31}=0$, it follows directly from three of the SFE that

$$\bar{\gamma}_3{}^{41}=\bar{\gamma}_3{}^{42}=\bar{\gamma}_3{}^{43}=0.$$

Thus, in Case I we have only four independent coefficients:

$$\gamma_1{}^{11}, \gamma_1{}^{31}, \gamma_1{}^{33}, \gamma_3{}^{44} \text{ (suppressing the bars).}$$

D. Case II

The initial set of assumptions is

$$\gamma_1{}^{33}=\gamma_2{}^{33}=0, \quad \gamma_1{}^{31}\neq 0, \quad \text{and} \quad \gamma_3{}^{44}=0.$$

By letting

$$\tan\theta=\gamma_1{}^{41}/\gamma_1{}^{31},$$

it follows, as discussed in Sec. C, that

$$\bar{\gamma}_1{}^{41}=\bar{\gamma}_3{}^{41}=\bar{\gamma}_3{}^{42}=\bar{\gamma}_3{}^{43}=0.$$

Let us derive the integrability conditions of the intrinsic derivatives which we shall use below. The first and second intrinsic derivatives of an arbitrary scalar α are

$$\alpha^{:A}=\alpha_{;\mu}\lambda^{A\mu}$$

and

$$\alpha^{:AB}=\alpha_{;\mu\nu}\lambda^{A\mu}\lambda^{B\nu}+\alpha_{;\mu}\lambda^{A\mu}{}_{;\nu}\lambda^{B\nu}.$$

Because

$$\alpha_{;\mu\nu}-\alpha_{;\nu\mu}=0,$$

it follows that

$$\alpha^{:AB}-\alpha^{:BA}=\alpha_{;\mu}\lambda_M{}^{\mu}(\gamma^{AMN}\lambda_{N\nu}\lambda^{B\nu}-\gamma^{BMN}\lambda_{N\nu}\lambda^{A\nu}).$$

Hence

$$\alpha^{:AB}-\alpha^{:BA}=\alpha^{:M}(\gamma_M{}^{BA}-\gamma_M{}^{AB}). \tag{D1}$$

Consider the following two equations from (A2):

$$\bar{\gamma}_1{}^{13}=\gamma_1{}^{13}+\alpha_{,\nu}\lambda^{3\nu}/\alpha$$

and

$$\bar{\gamma}_1{}^{14}=\gamma_1{}^{14}+\alpha_{,\nu}\lambda^{4\nu}/\alpha.$$

We wish to show that there exists an α that will enable one to set $\bar{\gamma}_1{}^{13}=0$ and $\bar{\gamma}_1{}^{14}=0$. This α is specified by the equations

$$\alpha^{:3}=-\alpha\gamma_1{}^{13}$$

and

$$\alpha^{:4}=-\alpha\gamma_1{}^{14}. \tag{D2}$$

The integrability condition (D1) for $\alpha^{:3}$ and $\alpha^{:4}$ is

$$\alpha^{:34}-\alpha^{:43}=\alpha^{:M}(\gamma_M{}^{43}-\gamma_M{}^{34}); \tag{D3}$$

which in this case becomes

$$\alpha^{:34}-\alpha^{:43}=\alpha^{:4}\gamma_3{}^{44}=-\alpha\gamma_1{}^{14}\gamma_3{}^{44}. \tag{D4}$$

From (D2), we obtain

$$\alpha^{:34}-\alpha^{:43}=\alpha(\gamma_1{}^{14:3}-\gamma_1{}^{13:4}). \tag{D5}$$

Therefore, from (D4) and (D5) we must have

$$\gamma_1{}^{14:3}-\gamma_1{}^{13:4}=-\gamma_1{}^{14}\gamma_3{}^{44}. \tag{D6}$$

Equation (D6) is satisfied for it is identically one of the SFE (22).

Thus, the four independent γ's for Case II are

$$\gamma_1{}^{11}, \quad \gamma_1{}^{12}, \quad \gamma_1{}^{31}, \quad \gamma_3{}^{44}.$$

E. Case III

The initial set of assumptions is

$$\gamma_1{}^{33}\neq 0, \quad \gamma_2{}^{33}\neq 0, \quad \gamma_1{}^{11}=0,$$
$$\gamma_1{}^{31}=0, \quad \gamma_1{}^{41}=0.$$

By letting

$$\alpha^2=-\gamma_2{}^{33}/\gamma_1{}^{33},$$

it follows, as discussed in Sec. C, that

$$\bar{\gamma}_1{}^{33}=-\bar{\gamma}_2{}^{33},\quad \bar{\gamma}_1{}^{11}=\bar{\gamma}_1{}^{12},$$
$$\bar{\gamma}_1{}^{13}=\bar{\gamma}_1{}^{14}=0.$$

Consider the following two equations from (B2):

$$\bar{\gamma}_3{}^{41}=\gamma_3{}^{41}+\theta_{,\nu}\lambda^{1\nu},$$

and

$$\bar{\gamma}_3{}^{42}=\gamma_3{}^{42}+\theta_{,\nu}\lambda^{2\nu}.$$

We wish to show that there exists a θ that will enable one to set simultaneously $\bar{\gamma}_3{}^{41}=0$ and $\bar{\gamma}_3{}^{42}=0$. This θ is specified by the equations

$$\theta^{:1}=-\gamma_3{}^{41}$$

and

$$\theta^{:2}=-\gamma_3{}^{42}. \tag{E1}$$

The integrability condition (D1) for $\theta^{:1}$ and $\theta^{:2}$ is

$$\theta^{:12}-\theta^{:21}=\theta^{:M}(\gamma_M{}^{21}-\gamma_M{}^{12}), \tag{E2}$$

which in this case becomes

$$\theta^{:12}-\theta^{:21}=-\gamma_1{}^{11}(\theta^{:1}+\theta^{:2})=\gamma_1{}^{11}(\gamma_3{}^{41}+\gamma_3{}^{42}). \tag{E3}$$

From (E1) we obtain

$$\theta^{:12}-\theta^{:21}=\gamma_3{}^{42:1}-\gamma_3{}^{41:2}. \tag{E4}$$

Therefore, from (E3) and (E4) we must have

$$\gamma_3{}^{42:1}-\gamma_3{}^{41:2}=\gamma_1{}^{11}(\gamma_3{}^{41}+\gamma_3{}^{42}). \tag{E5}$$

Equation (E5) is satisfied for it is identically one of the SFE (22).

Thus the four independent γ's for Case III are:

$$\gamma_1{}^{11},\quad \gamma_1{}^{33},\quad \gamma_3{}^{43},\quad \gamma_3{}^{44}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

New Approach to the Einstein and Maxwell-Einstein Field Equations*

EZRA T. NEWMAN†
Department of Physics, Syracuse University, Syracuse 10, New York
(Received December 28, 1960)

The components of a "vierbein" system are introduced as field variables in place of the metric tensor in a Riemannian space. The Riemann tensor, which is then written in terms of these new variables, is used to reformulate the Einstein and Einstein-Maxwell equations with or without a cosmological constant. These field equations have as solutions metrics with Riemann tensor of predetermined algebraic properties.

INTRODUCTION

IN a previous paper[1] we developed a new point of view for studying the Einstein empty-space field equation. In the present paper, we further develop this approach and generalize it to include the Maxwell-Einstein equations with or without a cosmological term. In the first section we derive an expression for the Riemann tensor in terms of the components of a set of vierbein vectors and their ordinary derivatives. In the second section, we use the previously derived form of the Riemann tensor in order to formulate field equations, which are equivalent to the usual formulation, with the advantage that we can ask for metrics such that the Riemann tensor takes a particular algebraic form.

SECTION I. RIEMANN TENSOR

We introduce four linearly independent tetrad vectors $\lambda_{A\mu}$ (the tetrad index A and the tensor index μ both range from 1-4) and define a metric tensor by

$$g_{\mu\nu}\equiv\lambda_{A\mu}\lambda^{A}{}_{\nu}, \tag{1.1}$$

where the tetrad index has been raised by means of a constant Minkowski metric η^{AB}.[1] (The identity symbol will be used in this section to denote a definition.) Next we define covariant differentiation, in terms of 64 equations

$$\lambda_{A\mu;\nu}\equiv\gamma_A{}^{BC}\lambda_{B\mu}\lambda_{C\nu}, \tag{1.2}$$

with the covariant derivative of a scalar being equal to the ordinary derivative, and assume that the product rule for differentiation holds. For later use, we can define the affine connection $\Gamma_{\mu\nu}{}^{\rho}$ by $\gamma_A{}^{BC}\lambda_{B\mu}\lambda_{C\nu}=\lambda_{A\mu,\nu}-\Gamma_{\mu\nu}{}^{\rho}\lambda_{A\rho}$. The $\gamma_A{}^{BC}$'s (Ricci rotation coefficients) are at present 64 arbitrary functions [Equation (1.2) does not yet imply that the affine connection be symmetric.] It is required that

$$g_{\mu\nu;\sigma}=0. \tag{1.3}$$

This restricts the 64 γ's to only 24 by means of the relations[2]

$$\gamma^{ABC}=-\gamma^{BAC}. \tag{1.4}$$

Using the definition (1.2), we define the tensor

$$\tilde{R}_{\alpha\beta\gamma\delta}\equiv 2\lambda^{A}{}_{\alpha}\lambda_{A\beta;[\gamma\delta]}. \tag{1.5}$$

* This work has been supported by the U. S. Air Force Office of Scientific Research.

† On leave of absence from the University of Pittsburgh.

[1] E. Newman and L. Tamburino, J. Math. Phys. 2, 667 (1961).

[2] Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1960), Chap. III.

By letting

$$\alpha^2=-\gamma_2{}^{33}/\gamma_1{}^{33},$$

it follows, as discussed in Sec. C, that

$$\bar{\gamma}_1{}^{33}=-\bar{\gamma}_2{}^{33},\quad \bar{\gamma}_1{}^{11}=\bar{\gamma}_1{}^{12},$$
$$\bar{\gamma}_1{}^{13}=\bar{\gamma}_1{}^{14}=0.$$

Consider the following two equations from (B2):

$$\bar{\gamma}_3{}^{41}=\gamma_3{}^{41}+\theta_{,\nu}\lambda^{1\nu},$$

and

$$\bar{\gamma}_3{}^{42}=\gamma_3{}^{42}+\theta_{,\nu}\lambda^{2\nu}.$$

We wish to show that there exists a θ that will enable one to set simultaneously $\bar{\gamma}_3{}^{41}=0$ and $\bar{\gamma}_3{}^{42}=0$. This θ is specified by the equations

$$\theta^{:1}=-\gamma_3{}^{41}$$

and

$$\theta^{:2}=-\gamma_3{}^{42}. \tag{E1}$$

The integrability condition (D1) for $\theta^{:1}$ and $\theta^{:2}$ is

$$\theta^{:12}-\theta^{:21}=\theta^{:M}(\gamma_M{}^{21}-\gamma_M{}^{12}), \tag{E2}$$

which in this case becomes

$$\theta^{:12}-\theta^{:21}=-\gamma_1{}^{11}(\theta^{:1}+\theta^{:2})=\gamma_1{}^{11}(\gamma_3{}^{41}+\gamma_3{}^{42}). \tag{E3}$$

From (E1) we obtain

$$\theta^{:12}-\theta^{:21}=\gamma_3{}^{42:1}-\gamma_3{}^{41:2}. \tag{E4}$$

Therefore, from (E3) and (E4) we must have

$$\gamma_3{}^{42:1}-\gamma_3{}^{41:2}=\gamma_1{}^{11}(\gamma_3{}^{41}+\gamma_3{}^{42}). \tag{E5}$$

Equation (E5) is satisfied for it is identically one of the SFE (22).

Thus the four independent γ's for Case III are:

$$\gamma_1{}^{11},\quad \gamma_1{}^{33},\quad \gamma_3{}^{43},\quad \gamma_3{}^{44}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

New Approach to the Einstein and Maxwell-Einstein Field Equations*

EZRA T. NEWMAN†
Department of Physics, Syracuse University, Syracuse 10, New York
(Received December 28, 1960)

The components of a "vierbein" system are introduced as field variables in place of the metric tensor in a Riemannian space. The Riemann tensor, which is then written in terms of these new variables, is used to reformulate the Einstein and Einstein-Maxwell equations with or without a cosmological constant. These field equations have as solutions metrics with Riemann tensor of predetermined algebraic properties.

INTRODUCTION

IN a previous paper[1] we developed a new point of view for studying the Einstein empty-space field equation. In the present paper, we further develop this approach and generalize it to include the Maxwell-Einstein equations with or without a cosmological term. In the first section we derive an expression for the Riemann tensor in terms of the components of a set of vierbein vectors and their ordinary derivatives. In the second section, we use the previously derived form of the Riemann tensor in order to formulate field equations, which are equivalent to the usual formulation, with the advantage that we can ask for metrics such that the Riemann tensor takes a particular algebraic form.

SECTION I. RIEMANN TENSOR

We introduce four linearly independent tetrad vectors $\lambda_{A\mu}$ (the tetrad index A and the tensor index μ both range from 1-4) and define a metric tensor by

$$g_{\mu\nu}\equiv\lambda_{A\mu}\lambda^{A}{}_{\nu}, \tag{1.1}$$

where the tetrad index has been raised by means of a constant Minkowski metric η^{AB}.[1] (The identity symbol will be used in this section to denote a definition.) Next we define covariant differentiation, in terms of 64 equations

$$\lambda_{A\mu;\nu}\equiv\gamma_A{}^{BC}\lambda_{B\mu}\lambda_{C\nu}, \tag{1.2}$$

with the covariant derivative of a scalar being equal to the ordinary derivative, and assume that the product rule for differentiation holds. For later use, we can define the affine connection $\Gamma_{\mu\nu}{}^{\rho}$ by $\gamma_A{}^{BC}\lambda_{B\mu}\lambda_{C\nu}=\lambda_{A\mu,\nu}-\Gamma_{\mu\nu}{}^{\rho}\lambda_{A\rho}$. The $\gamma_A{}^{BC}$'s (Ricci rotation coefficients) are at present 64 arbitrary functions [Equation (1.2) does not yet imply that the affine connection be symmetric.] It is required that

$$g_{\mu\nu;\sigma}=0. \tag{1.3}$$

This restricts the 64 γ's to only 24 by means of the relations[2]

$$\gamma^{ABC}=-\gamma^{BAC}. \tag{1.4}$$

Using the definition (1.2), we define the tensor

$$\tilde{R}_{\alpha\beta\gamma\delta}\equiv 2\lambda^{A}{}_{\alpha}\lambda_{A\beta;[\gamma\delta]}. \tag{1.5}$$

* This work has been supported by the U. S. Air Force Office of Scientific Research.

† On leave of absence from the University of Pittsburgh.

[1] E. Newman and L. Tamburino, J. Math. Phys. 2, 667 (1961).

[2] Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1960), Chap. III.

This tensor will eventually become the Riemann tensor. By substituting (1.2) into (1.5) and regrouping terms[1] we get

$$\bar{R}_{\alpha\beta\gamma\delta}=V^{st}\underset{s}{L}_{\alpha\beta}\underset{t}{L}_{\gamma\delta}, \tag{1.6}$$

where s and t run from 1 to 6, the $L_{s\alpha\beta}$ are the 6 bivectors constructed from $\lambda_{A\alpha}$

$$\underset{s}{L}_{\alpha\beta}\rightarrow\lambda_{A[\alpha}\lambda_{B\beta]} \tag{1.7}$$

(in other words, the s or t stands for an ordered pair of indices A and B) and

$$V^{st}\rightarrow 4(\gamma^{ABM:N}-\gamma^{ABN:M}+\gamma_K{}^{AN}\gamma^{KBM}-\gamma_K{}^{AM}\gamma^{KBN}-2\gamma^{ABK}\gamma_K{}^{[NM]}, \tag{1.8}$$

with

$$\gamma^{ABM:N}\equiv\gamma^{ABM}{}_{,\mu}\lambda^{N\mu}.$$

At this point, the tensor $\bar{R}_{\alpha\beta\gamma\delta}$ cannot be considered as the Riemann tensor of the metric $g_{\mu\nu}$ (1.1), as the affine connection is not yet symmetric.

The necessary and sufficient condition that the affine connection be the Cristoffel symbol and hence that (1.6) be the Riemann tensor is that

$$\lambda_{A[\alpha;\beta]}=\lambda_{A[\alpha,\beta]}. \tag{1.9}$$

In order to prove this, let

$$G_{A\alpha\beta}=\lambda_{A\alpha;\beta}-\lambda_{A\alpha,\beta}. \tag{1.10}$$

We also have

$$G_{A\alpha\beta}=G_{A[\alpha\beta]}+G_{A(\alpha\beta)}. \tag{1.11}$$

From (1.9), $G_{A[\alpha\beta]}=0$ and hence

$$\lambda_{A\alpha;\beta}=\lambda_{A\alpha,\beta}+G_{A(\alpha\beta)}. \tag{1.12}$$

The $G_{A(\alpha\beta)}$ can be rewritten as

$$\begin{aligned}G_{A(\alpha\beta)}&=G_{B(\alpha\beta)}\delta_A{}^B\\&=G_{B(\alpha\beta)}\lambda^{B\mu}\lambda_{A\mu}\\&=-\Gamma_{(\alpha\beta)}{}^{\mu}\lambda_{A\mu},\end{aligned} \tag{1.13}$$

with $\Gamma_{(\alpha\beta)}{}^{\mu}\equiv -G_{B(\alpha\beta)}\lambda^{B\mu}$.
Hence,

$$\lambda_{A\alpha;\beta}=\lambda_{A\alpha,\beta}-\Gamma_{(\alpha\beta)}{}^{\mu}\lambda_{A\mu}. \tag{1.14}$$

This form, with condition (1.3), completes our proof (see, for example, Schrödinger[3]).

We will translate this condition into a form that is convenient for use in (1.6). By defining

$$C^{ABC}\equiv\gamma^{A[BC]}, \tag{1.15}$$

we have the identity[4]

$$\gamma^{ABC}=C^{ABC}+C^{CBA}+C^{BCA}. \tag{1.16}$$

Now, by multiplying (1.9) by $\lambda^{B\alpha}$ and $\lambda^{C\beta}$, we obtain

$$\gamma_A{}^{[BC]}\equiv C_A{}^{BC}=\lambda_{A[\alpha,\beta]}\lambda^{B\alpha}\lambda^{C\beta}. \tag{1.17}$$

[3] E. Schrödinger, *Space-Time Structure* (Cambridge University Press, New York, 1950), Chap. VI.
[4] This identity was pointed out to us by R. Kerr.

If (1.16) is used with (1.17) in (1.6) and (1.8), we obtain the result that $\bar{R}_{\alpha\beta\gamma\delta}=R_{\alpha\beta\gamma\delta}$, the Riemann tensor. This is the form of the Riemann tensor that we will use in the next section to reformulate the Einstein field equations.

SECTION II. FIELD EQUATIONS

Consider a tensor

$$S_{\alpha\beta\gamma\delta}=U^{st}\underset{s}{L}_{\alpha\beta}\underset{t}{L}_{\gamma\delta} \tag{2.1}$$

with all the algebraic properties of the Riemann tensor, the U^{st} being functions of position.

Eventually, after imposing further algebraic conditions on $S_{\alpha\beta\gamma\delta}$, we will consider as field equations

$$R_{\alpha\beta\gamma\delta}=S_{\alpha\beta\gamma\delta}. \tag{2.2}$$

The further conditions arise from the algebraic properties of the Ricci tensor $R^{\alpha}{}_{\beta\gamma\alpha}\equiv R_{\beta\gamma}$. We consider the following three cases:

$$\text{I.}\quad R_{\alpha\beta}=0 \tag{2.3}$$

$$\text{II.}\quad R_{\alpha\beta}=\kappa T_{\alpha\beta} \tag{2.4}$$

$$\text{III.}\quad R_{\alpha\beta}=\Lambda g_{\alpha\beta}. \tag{2.5}$$

The $T_{\alpha\beta}$ is the stress tensor of the Maxwell field

$$T_{\alpha\beta}=(F_{\alpha\mu}F^{\mu}{}_{\beta}+\tfrac{1}{4}g_{\alpha\beta}F_{\mu\nu}F^{\mu\nu}), \tag{2.6}$$

and Λ is the cosmological constant.

Case I leads to the condition on $S_{\alpha\beta\gamma\delta}$

$$S^{\mu}{}_{\alpha\beta\mu}=0, \tag{2.7}$$

the algebra of which has been completely analyzed by Petrov.[5] In addition, the form of the U^{st} in (2.1) has also been given.[6,7] This particular $S_{\alpha\beta\gamma\delta}$ will be denoted by $C_{\alpha\beta\gamma\delta}$.

Case II splits into two subcases depending on whether the Maxwell field is null or non-null. (For a complete discussion of the algebra of these two fields, see Synge.[8]) In the former case our condition is

$$S^{\mu}{}_{\alpha\beta\mu}=\sigma^2k_{\alpha}k_{\beta},\quad k_{\mu}k^{\mu}=0 \tag{2.8}$$

and in the latter

$$\begin{aligned}&S^{\mu}{}_{\alpha\beta\mu}=\kappa(2k_{(\alpha}l_{\beta)}+m_{\alpha}m_{\beta}+n_{\alpha}n_{\beta})\\&k_{\mu}k^{\mu}=l_{\mu}l^{\mu}=k_{\mu}m^{\mu}=k_{\mu}n^{\mu}=l_{\mu}m^{\mu}=l_{\mu}n^{\mu}=m_{\mu}n^{\mu}=0\\&k_{\mu}l^{\mu}=-m_{\mu}m^{\mu}=-n_{\mu}n^{\mu}=1.\end{aligned} \tag{2.9}$$

(This vierbein need not be the one used in Sec. I.)

In case III, our conditions become

$$S^{\mu}{}_{\alpha\beta\mu}=\Lambda\lambda_{A\alpha}\lambda^{A}{}_{\beta}. \tag{2.10}$$

[5] A. Z. Petrov, Sci. Not. Kazan State Univ. 114, 55 (1954).
[6] E. Newman, J. Math. Phys. 2, 324 (1961).
[7] J. F. Schell, "A classification of four-dimensional Riemannian spaces" (preprint).
[8] J. L. Synge, *Relativity, the Special Theory* (North-Holland Publishing Company, Amsterdam, 1959), Chap. IX.

From the conditions on S, in cases II and III, it is an easy matter to calculate the full S by means of the following equation:

$$S_{\alpha\beta\gamma\delta}=\tfrac{1}{2}(g_{\alpha\delta}S^{\mu}{}_{\beta\gamma\mu}-g_{\alpha\gamma}S^{\mu}{}_{\beta\delta\mu}+g_{\beta\gamma}S^{\mu}{}_{\alpha\delta\mu}-g_{\beta\delta}S^{\mu}{}_{\alpha\gamma\mu}) +\tfrac{1}{6}S(g_{\alpha\delta}g_{\beta\gamma}-g_{\alpha\gamma}g_{\beta\delta})+C_{\alpha\beta\gamma\delta}, \quad (2.11)$$

with $g_{\alpha\beta}=\lambda_{A\alpha}\lambda^{A}{}_{\beta}$ and $S=S^{\mu\nu}{}_{\mu\nu}$.

This equation is nothing but a restatement of the equation for the Riemann tensor in terms of the Ricci tensor and the Weyl tensor.[9] It should be noted that any S from case I can be added to a case II or III and that the Riemann scalar R vanishes in case II.

It is easily seen what the field equations (2.2) mean. They are second-order differential equations for the $\lambda_{A\mu}$, such that the solution has a Riemann tensor of a particular algebraic form. From (2.2) with (1.6) and (2.1) we have the alternate form of the field equations,

$$U^{st}=V^{st}. \quad (2.12)$$

With case II, these equations must be supplemented by Maxwell's equations

$$F^{\mu\nu}{}_{;\nu}=F^{\mu\nu}{}_{;\nu}{}^{*}=0, \quad (2.13)$$

the star indicating the dual. For the null field

$$F_{\mu\nu}=\sigma k_{[\mu}m_{\nu]}, \quad (2.14)$$

and for the non-null field.

$$F_{\mu\nu}=\alpha k_{[\mu}l_{\nu]}+\beta m_{[\mu}n_{\nu]}. \quad (2.15)$$

DISCUSSION

Pirani,[10] several years ago, gave a discussion of gravitational radiation which was based on the Petrov classification of empty-space Riemann tensors or of Weyl tensors. The classification allows one to choose uniquely or almost uniquely (up to a one or two parameter rotation) a vierbein system in terms of the eigenbivectors of the Riemann tensor. This vierbein is called the canonical vierbein. Frequently, because of Pirani, one can identify one of the canonical vierbein vectors as a propagation vector for gravitational radiation or as a polarization vector.

Although the choice of the vierbein to be used in the field equations (2.2) is arbitrary, it is often convenient (though not always, even in empty space) to choose them as the canonical vierbein. The formalism presented here has, in addition to the advantage of giving the equations for metrics with a Riemann tensor of specific algebraic type, the advantage of allowing one to predetermine certain differential properties of the propagation vector (for example, whether it is a geodesic, whether it has a vanishing curl and/or divergence). The value of being able to ask for solutions with these prescribed characteristics is obvious.

An important special case of the field equations is when $S_{\alpha\beta\gamma\delta}=0$, i.e., flat-space, but with a Maxwell field. There will be two different sets of equations, one for the null and one for the non-null field. In both cases, the vierbein in the Riemann tensor should be chosen to be the one appearing in Eq. (2.9).

It appears that the field equations (2.2) offers, computationally, a simpler method of obtaining metrics, than the usual field equations allow. The method has been successfully applied to Petrov I degenerate with real eigenvalue and is being applied to all Petrov I degenerate, II and III as well as the null Maxwell field in flat space.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Dr. Ivor Robinson and Dr. Peter G. Bergmann for many valuable discussions and comments while this manuscript was in preparation.

[9] Reference 2, Chap. II.
[10] F. A. E. Pirani, Phys. Rev. **105**, 1089 (1957).

Coordinate Conditions*

J. PLEBANSKI
Institute of Theoretical Physics, University of Warsaw, Warsaw, Poland

AND

J. RYTEN
Institute of Physics of the Polish Academy of Sciences, Warsaw, Poland

(Received March 16, 1961)

Some extremal properties of coordinates imposed by coordinate conditions are studied. In particular, variational principles leading to the de Donder condition, the generalized de Donder condition, $[(-g)^w g^{\alpha\beta}]_{|\beta}=0$; Einstein's old condition $(-g)^{\frac{1}{2}}=1$; and the Einstein-Infeld conditions $[(-g)^{\frac{1}{2}}g^{0\rho}]_{|\rho}=0$, $[(-g)^{\frac{1}{2}}g^{ab}]_{|b}=0$ are given. The class of generalized de Donder conditions is examined in the case of Schwarzschild's solution. The values of $w=0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1$ are here distinguished. For an arbitrary w the mathematical problem is reduced to the study of certain solutions of the hypergeometric equation.

1. INTRODUCTION

As is well known, the de Donder condition[1] $\mathcal{G}^{\alpha\beta}{}_{|\beta} \equiv [(-g)^{\frac{1}{2}}g^{\alpha\beta}]_{|\beta}=0$ forms a convenient technical tool for treating some mathematical problems of the theory of relativity. First of all, after using it one gets the leading term in the Einstein field equations in the form of a wave operator.[2,3] Secondly, the equation for harmonic coordinates $\bar{x}^\alpha$ as functions of arbitrary ones x^α are *linear* and have the simple form

$$\Box\bar{x}^\alpha \equiv -(-g)^{-\frac{1}{2}}[(-g)^{\frac{1}{2}}g^{\rho\sigma}\bar{x}^\alpha{}_{|\sigma}]_{|\rho}=0. \tag{1.1}$$

In the Schwarzschild case, one easily gets harmonic coordinates by a shift of the standard Schwarzschildian r through m, $r_h=r-m$, and subsequent transition from quasi-spherical to quasi-cartesian coordinates.

These coordinate conditions preserve 4-dimensional symmetry. It is sometimes useful to apply the Einstein-Infeld conditions (cf. reference 4 and in particular reference 5), where this symmetry is broken:

$$[(-g)^{\frac{1}{2}}g^{0\rho}]_{|\rho}=0, \quad [(-g)^{\frac{1}{2}}g^{ab}]_{|b}=0. \tag{1.2}$$

After developing $(-g)^{\frac{1}{2}}g^{\alpha\beta}$ in a power series in $\lambda=c^{-1}$ as $\eta^{\alpha\beta}+\gamma_1{}^{\alpha\beta}+\gamma_2{}^{\alpha\beta}+\cdots$, these conditions take (in terms of γ's) a more familiar form.

In the older literature, the simple condition of Einstein[6]

$$(-g)^{\frac{1}{2}}=1 \tag{1.3}$$

had also often been used. Some interest is also attached to the investigation of a class of coordinate conditions which are a simple generalization of these of de Donder

$$[(-g)^w g^{\alpha\beta}]_{|\beta}=0, \quad w \text{ arbitrary}. \tag{1.4}$$

Conditions (1.4) will be called super-harmonic coordinate conditions.

For $w=\frac{1}{2}$ they coincide with these of de Donder. The value of $w=1$ is also distinguished. Observe that the following equation holds identically:

$$(-g)^{\frac{1}{2}}R=-[(-g)^{-\frac{1}{2}}[(-g)g^{\alpha\beta}]_{|\beta}]_{|\alpha} + (-g)^{\frac{1}{2}}g^{\mu\nu}(\Gamma_{\nu\sigma}{}^\rho\Gamma_{\mu\rho}{}^\sigma-\Gamma_{\nu\mu}{}^\rho\Gamma_{\sigma\rho}{}^\sigma). \tag{1.5}$$

Hence, by assuming (1.4) with $w=1$, one kills all second derivatives in the density of the curvature scalar, i.e., in the Lagrange function of the theory of gravitation.

Moreover, as it is trivial to see, defining $g^{*\alpha\beta}=(-g)^{\frac{1}{2}}\times g^{\alpha\beta}$ one gets

$$\mathrm{Det}\|g^{*\alpha\beta}\|=1. \tag{1.6}$$

Therefore, only nine out of the ten $g^{*\alpha\beta}$ can be regarded as independent; they may be understood to be connected with the "polarizational" degrees of freedom of the metric, so to speak. Then $g=\mathrm{Det}\|g^{\alpha\beta}\|$ must be here comprehended as an independent quantity. It seems to be interesting to investigate the coordinate condition $g^{*\alpha\beta}{}_{|\beta}=0$ which does not affect directly this degree of freedom which corresponds to g.

It seems to be an interesting problem to study the question whether there exist some variational principles which would lead to the coordinate conditions mentioned above.

More generally, it would certainly be highly interesting to get some variational principle which would lead to "optimal" coordinates, which in some sense were as close to cartesian coordinates as is possible, consistently with the curvature of the space-time. It seems that such a principle could be constructed in accordance with the idea that if the curvature is not too large, and the coordinates are close to cartesian, the world lines of light rays are almost "straight" (in the sense of a linear relationship between the coordinates of the ray). Such coordinates, in which light rays were as straight as possible, would be excellent optimal coordinates. The process of projection of motions in Riemannian

* Part of this work was carried out by J. Plebanski during a stay at Lehigh University, Bethlehem, Pennsylvania, and was supported by the Air Force Research Division, Air Research and Development Command, U. S. Air Force.

[1] De Donder, *La Gravifique Einsteinienne* (Paris, 1921).
[2] V. Fock, *The Theory of Space, Time and Gravitation* (Pergamon Press, New York, 1960).
[3] J. Plebanski and B. Bertotti, Ann. Phys. **11**, 169 (1960).
[4] A. Einstein and L. Infeld, Can. J. Math. **1**, 209 (1949).
[5] L. Infeld, Ann. Phys. **6**, 341 (1959).
[6] A. Einstein, Ann. Phys. **49**, 769 (1916).

space-time on planes at infinity (which is important from the point of view of the physical interpretation, cf. chapter V of reference 7) would be very easy to execute.

Unfortunately, we have not succeeded in constructing such a "physical" principle. We are able, however, to write some more formal variational principles which lead to the previously discussed formal coordinate conditions. This is discussed in Sec. 2 and, we hope, may prove useful in giving some hints for the problem of the construction of the "physical" principle discussed above.

In Sec. 3, in order to get some ideas of the significance of the "super-harmonic" coordinates, we study the condition (1.4) in the case of Schwarzschild's solution. It has been found, in this case, that the structure of the condition (1.4) distinguishes the values of $w=0$, $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1$. We also proved that, in the case of an arbitrary w, one can get the coordinate transformation leading from Schwarzschildian coordinates to the "super-harmonic" ones in an explicit way; the mathematical problem is reduced to the study of some solutions of a hypergeometric equation. Strangely enough, for some w, there exist well defined regions of Schwarzschildian coordinates where the transition to super-harmonic coordinates is impossible; whereas in other regions, it still could be done.

2. VARIATIONAL PRINCIPLES

To every coordinate system $\{x^\alpha\}$ let there correspond a number W defined by the integral

$$W=\int d_4x(-g)^{\frac{1}{2}}L(g^{\mu\nu}) \tag{2.1}$$

(the integration runs over the whole range of the given coordinates), where $g^{\mu\nu}$ are the components of the metric tensor in *this* coordinate system and $L(g^{\mu\nu})$ always has the same from as a function of the $g^{\mu\nu}$. Thus W is, for a given world, a functional of the coordinate system $\{x^\alpha\}$. We are looking for such a coordinate system in which this functional has a stationary value.

Let $\{x^\alpha\}$ be this coordinate system that we are looking for, and $\{x'^\alpha\}$ an arbitrary one. Of course

$$W'\equiv W[x'^\alpha]=\int d_4x'(-g')^{\frac{1}{2}}L(g'^{\mu\nu})$$

$$=\int d_4x(-g)^{\frac{1}{2}}L(x_{|\rho}{}'^{\nu}x_{|\sigma}{}'^{\mu}g^{\rho\sigma}). \tag{2.2}$$

Now let us suppose that the difference between x^α and x'^α is small: $x'^\alpha-x^\alpha=\delta x^\alpha(x^\beta)$. Thus,

$$W'-W\cong\delta W=\int d_4x(-g)^{\frac{1}{2}}\frac{\partial L}{\partial g^{\mu\nu}}(\delta x^\nu{}_{|\rho}g^{\rho\mu}+g^{\nu\rho}\delta x^\mu{}_{|\rho})$$

$$=\oint d_3x(-g)^{\frac{1}{2}}\frac{\partial L}{\partial g^{\mu\nu}}(\delta x^\nu g^{\rho\mu}+g^{\nu\rho}\delta x^\mu)n_\rho$$

$$-2\int d_4x\left[(-g)^{\frac{1}{2}}\frac{\partial L}{\partial g^{\mu\nu}}g^{\rho\mu}\right]_{|\rho}\delta x^\nu. \tag{2.3}$$

Now if our coordinates $\{x^\alpha\}$ are stationary, δW must vanish for arbitrary δx^ν, vanishing on the 3-dimensional boundary; from which we obtain the coordinate conditions

$$\left[(-g)^{\frac{1}{2}}\frac{\partial L}{\partial g^{\mu\nu}}g^{\rho\mu}\right]_{|\rho}=0. \tag{2.4}$$

As an example, let us consider

$$W=\int d_4x(-g)^w\equiv\int d_4x(-g)^{\frac{1}{2}}(-g)^{w-\frac{1}{2}}. \tag{2.5}$$

For such a functional, (2.4) takes the form

$$(w-\tfrac{1}{2})[(-g)^w]_{|\nu}=0, \tag{2.6}$$

which, for any $w\neq\frac{1}{2}$, leads to $(-g)^{\frac{1}{2}}=\text{const}$, and in particular reduces to Einstein's condition $(-g)^{\frac{1}{2}}=1$.

Now, let us consider 10 functionals

$$\langle(-g)^wg^{\alpha\beta}\rangle=_{df}\int d_4x(-g)^wg^{\alpha\beta}$$

$$=\int d_4x(-g)^{\frac{1}{2}}(-g)^{w-\frac{1}{2}}g^{\alpha\beta}, \tag{2.7}$$

which could be interpreted as sort of "averages" of the tensor densities $(-g)^wg^{\alpha\beta}$ with the weight w. The variations of these quantities with respect to small coordinate transformations are [see (2.3)]

$$\delta\langle(-g)^wg^{\alpha\beta}\rangle=\int d_4x(-g)^w[(1-2w)g^{\alpha\beta}\delta x_{|\rho}{}^\rho$$

$$+\delta x^\alpha{}_{|\rho}g^{\rho\beta}+\delta x^\beta{}_{|\rho}g^{\alpha\rho}]. \tag{2.8}$$

Now, assuming that δx^α vanishes on the boundary and integrating by parts, one gets

$$\delta\langle(-g)^wg^{\alpha\beta}\rangle=-\int d_4x\{(1-2w)[(-g)^wg^{\alpha\beta}]_{|\rho}\delta x^\rho$$

$$+[(-g)^wg^{\alpha\rho}]_{|\rho}\delta x^\beta+[(-g)^wg^{\beta\rho}]_{|\rho}\delta x^\alpha\}. \tag{2.9}$$

In particular, for $w=\frac{1}{2}$, one gets

$$\delta\langle(-g)^{\frac{1}{2}}g^{\alpha\beta}\rangle=-\int d_4x\{[(-g)^{\frac{1}{2}}g^{\alpha\rho}]_{|\rho}\delta x^\beta$$

$$+[(-g)^{\frac{1}{2}}g^{\beta\rho}]_{|\rho}\delta x^\alpha\}. \tag{2.10}$$

[7] L. Infeld and J. Plebanski, *Motion and Relativity* (Pergamon Press, New York, 1960).

Therefore, one sees that the harmonic coordinates satisfying

$$[(-g)^{\frac{1}{2}}g^{\alpha\rho}]_{|\rho}=0,$$

are precisely these for which all ten functionals

$$\langle(-g)^{\frac{1}{2}}g^{\alpha\beta}\rangle$$

are stationary for small changes of coordinates. Inversely, it is trivial to see that, from conditions $\delta\langle(-g)^{\frac{1}{2}}\times g^{\alpha\beta}\rangle=0$ for arbitrary δx^{α}, de Donder conditions follow.

Now, coming back to the general case of an arbitrary w, one sees that the super-harmonic coordinates, i.e., those satisfying

$$[(-g)^{w}g^{\alpha\rho}]_{|\rho}=0,$$

are these for which all ten functionals $\langle(-g)^{w}g^{\alpha\beta}\rangle$ are stationary, if one restricts oneself to the volume-preserving coordinate variations [i.e., such that $\delta\partial(x')/\partial(x)=\delta x^{\rho}{}_{|\rho}=0$].

Now let us look for variational principles for the Einstein-Infeld conditions

$$[(-g)^{\frac{1}{2}}g^{0\rho}]_{|\rho}=0,\quad [(-g)^{\frac{1}{2}}g^{ab}]_{|b}=0. \tag{2.11}$$

The first condition (2.11) makes the functional $\langle(-g)^{\frac{1}{2}}\times g^{ab}\rangle$ stationary. Indeed, from (2.10) for $\alpha=\beta=0$ it follows that

$$\delta\langle(-g)^{\frac{1}{2}}g^{00}\rangle=-2\int d_4x[(-g)^{\frac{1}{2}}g^{0\rho}]_{|\rho}\delta x^{0}. \tag{2.12}$$

We now define six functionals

$$[(-g)^{\frac{1}{2}}g^{ab}]=\int_{df} d_3x(-g)^{\frac{1}{2}}g^{ab}, \tag{2.13}$$

which could be interpreted as our "averages" $\langle(-g)^{\frac{1}{2}}g^{ab}\rangle$ for a fixed but arbitrary time x^0. If we restrict ourselves to the coordinate transformations $x'^{\alpha}-x^{\alpha}=\delta x^{\alpha}(x^{\beta})$ such that $x'^{0}=x^{0}$ and $\delta x^{a}{}_{|0}=0$, we obviously have

$$\delta[(-g)^{\frac{1}{2}}g^{ab}]=\int d_3x(-g)^{\frac{1}{2}}[\delta x^{a}{}_{|r}g^{br}+\delta x^{b}{}_{|r}g^{ar}], \tag{2.14}$$

so assuming δx^{a} vanishing at spatial infinity and integrating by parts we get

$$\delta[(-g)^{\frac{1}{2}}g^{ab}]=-\int d_3x\{[(-g)^{\frac{1}{2}}g^{rb}]_{|r}\delta x^{a}+[(-g)^{\frac{1}{2}}g^{ra}]_{|r}\delta x^{b}\}; \tag{2.15}$$

from which it is seen that three of the Einstein-Infeld conditions $[(-g)^{\frac{1}{2}}g^{ab}]_{|b}=0$ make the functionals (2.13) stationary in the class of variations $\delta x^{a}{}_{|0}=0$, $\delta x^{0}=0$.

Thus we were able to show the existence of some variational principles connected with the most often used formal coordinate conditions. It is remarkable that, as we have seen, 4 formal coordinate conditions make many functionals simultaneously stationary.

3. SUPERHARMONIC COORDINATES AND SCHWARZCHILD'S SOLUTION

Let us denote the super-harmonic coordinates by $\bar{x}^{\alpha}$. In this notation the condition (1.4) must be rewritten in the form

$$[(-\bar{g})^{w}\bar{g}^{\alpha\beta}(\bar{x})]_{|\beta}=0. \tag{3.1}$$

After simple calculations, using the C. Neymann identities (see Appendix), one gets four equations for $\bar{x}^{\alpha}(x^{\beta})$

$$\bar{x}^{\alpha}{}_{|\nu\mu}g^{\nu\mu}(-g)^{w}+\bar{x}^{\alpha}{}_{|\nu}[g^{\nu\mu}(-g)^{w}]_{|\mu}+(2w-1)\bar{x}^{\sigma}{}_{|\rho\mu}\bar{x}^{\alpha}{}_{|\nu}g^{\nu\mu}x^{\rho}{}_{|\sigma}(-g)^{w}=0. \tag{3.2}$$

In the case of de Donder ($w=\frac{1}{2}$) conditions, these equations reduce to linear ones; but, in the general case of arbitrary w, Eq. (3.2) are highly nonlinear. It is remarkable that the standard study of characteristic surfaces associated with these equations leads to the equation

$$2(1-w)[g^{\alpha\beta}\varphi_{|\alpha}\varphi_{|\beta}]^{4}=0, \tag{3.3}$$

where φ is the characteristic surface. Therefore, for $w=1$, any surface is here characteristic.

In order to say something more about the super-harmonic coordinates, it seems advisable to study a solution of Einstein's equations in some concrete case. We shall concentrate on the case of Schwarzschild's solution. Assume therefore

$$ds^{2}=\left(1-\frac{2m}{r}\right)dx^{02}-\frac{dr^{2}}{1-(2m/r)}-r^{2}(d\vartheta^{2}+\sin^{2}\vartheta d\varphi^{2}). \tag{3.4}$$

We pass now to quasi-cartesian coordinates x^{α} and then transform them $x^{\alpha}\rightarrow\bar{x}^{\alpha}(x^{\beta})$, where $\bar{x}^{0}=x^{0}$, $\bar{x}^{k}=[\bar{r}(r)/r]x^{k}$. For such a case, as may be easily checked, the four equations (3.2) reduce to one equation for the function $\bar{r}(r)$:

$$(1-w)\frac{r^{2}}{\bar{r}}\left(1-\frac{2m}{r}\right)\frac{d^{2}\bar{r}}{dr^{2}}+(1-2w)\frac{r^{2}}{\bar{r}^{2}}\left(1-\frac{2m}{r}\right)\times\left(\frac{d\bar{r}}{dr}\right)^{2}+\frac{1}{\bar{r}}(m-4wm+2wr)\frac{d\bar{r}}{dr}-1=0. \tag{3.5}$$

For certain values of w, this nonlinear equation can be simplified. For $w=\frac{1}{2}$ we get, as we expected, a linear equation. For $w=1$, the term with second derivative vanishes and we obtain a nonlinear equation of the first order, which can be easily solved. The value of $w=\frac{1}{4}$ also seems to be distinguished, as for it the coefficient of the first derivative becomes simpler. In order to get the solution of (3.5) in the general case of arbitrary w, we shall try the form

$$\bar{r}(r)=f^{a}(r),\quad a=(1-w)/(2-3w). \tag{3.6}$$

(We assume here for the moment $w\neq 1$, $w\neq\frac{2}{3}$; $w=1$ and $w=\frac{2}{3}$ will be studied separately.) After inserting (3.6) into (3.5), we get for the function $f(r)$ the hypergeometric equation which, written in a standard form, is

$$(r'^2-r')\frac{d^2f}{dr'^2}+[(1+\alpha+\beta)r'-\gamma]\frac{df}{dr'}+\alpha\beta f=0, \qquad r'=\frac{r}{2m}, \tag{3.7}$$

where the parameters α, β, and γ are

$$\alpha=\frac{3w-2}{1-w}, \quad \beta=\frac{1}{1-w}=\alpha+3, \quad \gamma=\frac{4w-1}{2(1-w)}. \tag{3.8}$$

This equation has three singular points $r'=0$, 1, ∞, and one can find the solution around any singularity. All these solutions can then be connected so that the solution may be obtained for all r. Here, however, we shall restrict ourselves only to the study of solutions about $r=\infty$, i.e., for $r>2m$.

Two independent solutions of (3.7) about $r'=$ are

$$f_1(r')=r'^{1/(w-1)}F\left(\frac{1}{1-w}, 3-\frac{1}{2(1-w)}; 4; \frac{1}{r'}\right),$$

$$f_2(r')=r'^{(3w-2)/(w-1)}\left[1+\frac{3w-2}{4(w-1)^2}r'^{-1}+\frac{(2w-1)^2(3w-2)}{16(w-1)^4}r'^{-2}\right]$$

$$+\tfrac{1}{2}r'^{1/(w-1)}\sum_{n=0}^{\infty}\left\{\frac{\left(\frac{3w-2}{1-w}\right)_{n+3}\left(\frac{1}{2(w-1)}\right)_{n+3}}{n!(n+3)!}\right.$$

$$\times\left[-\ln r'-\frac{1}{3}+\sum_{k=0}^{n+2}\left(\frac{1}{\frac{3w-2}{1-w}+k}+\frac{1}{\frac{1}{2(w-1)}+k}\right)\right.$$

$$\left.\left.-\sum_{k=1}^{n}\left(\frac{1}{k+3}+\frac{1}{k}\right)\right]r'^{-n}\right\}. \tag{3.9}$$

The general solution of Eq. (3.5) is

$$\bar{r}(r)=2m\left[Cf_1\left(\frac{r}{2m}\right)+f_2\left(\frac{r}{2m}\right)\right]^{(1-w)/(2-3w)}. \tag{3.10}$$

This solution depends on one arbitrary constant C. The second constant has been fixed by imposing on $\bar{r}$ the condition

$$\lim_{r\to\infty}\frac{\bar{r}(r)}{r}=1. \tag{3.11}$$

Some special cases of the solution (3.10) are interesting. First, let us observe that for $w=(k+1)/k$, $(k\geqslant 1)$ or $w=(2l-1)/(2l)$, $(l\geqslant 3)$; both series in f_1 and in f_2 become finite. Secondly, if one of the numbers $(3w-2)/(1-w)$, $1/[2(w-1)]$ equals -1 or -2, the sum $\sum_{n=0}^{\infty}$ in the f_2 vanishes; such is the case for $w=0$, $\frac{1}{2}$, $\frac{3}{4}$; and for these w, the solutions (3.10) is extremely simple. For $w=0$ we have

$$\bar{r}(r)=\left[r^{\frac{1}{2}}(r-2m)^{\frac{1}{2}}-C_1\left(\frac{r}{2m}\right)^{\frac{1}{2}}\left(\frac{r}{2m}-1\right)^{\frac{1}{2}}+C_1\left(\frac{r^2}{4m^2}-\frac{r}{4m}-\frac{1}{8}\right)\right]^{\frac{1}{2}}, \tag{3.12}$$

for $w=\frac{1}{2}$,

$$\bar{r}(r)=(r-m)+C_2\left[1+\frac{r}{2m}\ln\left|1-\frac{2m}{r}\right|-\frac{1}{2}\ln\left|1-\frac{2m}{r}\right|\right], \tag{3.13}$$

for $w=\frac{3}{4}$,

$$\bar{r}(r)=r^3(r-2m)/(r^3+C_3), \tag{3.14}$$

where C_1, C_2, C_3 are arbitrary constants. The two latter solutions are especially interesting. As is immediately seen for $C_2=C_3=0$, they become, respectively,

$$\bar{r}(r)=r-m, \tag{3.13a}$$

$$\bar{r}(r)=r-2m. \tag{3.14a}$$

For $w=\frac{1}{2}$, this result has been known for a long time. Here, however, we see that one gets super-harmonic coordinates in the case $w=\frac{3}{4}$ also simply by the shift of Schwarzschildian r. [It may be shown that these two cases are the only ones for which such a shift is a solution of (3.5).] The shift is here through $2m$ and if we write Schwarzschild's line element (3.4) in these super-harmonic coordinates we get

$$ds^2=\left(1+\frac{2m}{\bar{r}}\right)^{-1}dt^2-\left(1+\frac{2m}{\bar{r}}\right)d\bar{r}^2-(\bar{r}+2m)^2(d\vartheta^2+\sin^2\vartheta d\varphi^2), \tag{3.15}$$

so in these coordinates we would cut out the Schwarzschild singularity.

Now we have to discuss two remaining solutions of (3.5), namely those for $w=\frac{2}{3}$ and $w=1$. For $w=\frac{2}{3}$ we have

$$\bar{r}(r)=\tfrac{1}{4}\exp\left\{-\frac{m}{2r}-\frac{[(m/r)+1]^2-3}{[1-(2m/r)]^{\frac{1}{2}}}\times[\ln(r^{\frac{1}{2}}+(r-2m)^{\frac{1}{2}})+A]-2A\right\}, \tag{3.16}$$

where A is an arbitrary constant. This solution holds for $r>2m$. For $w=1$ we get two solutions (the equation

is then a quadratic equation of the first order):

$$\bar{r}(r)=A_1[3m+(9m^2-4mr)^{\frac{1}{2}}]^{\frac{1}{2}} \times[m-(9m^2-4mr)^{\frac{1}{2}}]^{\frac{1}{2}}, \quad (3.17)$$

$$\bar{r}(r)=A_2[3m-(9m^2-4mr)^{\frac{1}{2}}]^{\frac{1}{2}} \times[m+(9m^2-4mr)^{\frac{1}{2}}]^{\frac{1}{2}}. \quad (3.18)$$

The first one (3.17) is real for $2m<r<(9/4)m$ (for $r<2m$ it is purely imaginary); the second one (3.18) is real for $r<(9/4)m$. In the whole remaining region, i.e., for $r>(9/4)m$, both solutions are complex and cannot be made real; so for $r>(9/4)m$, the transition to super-harmonic coordinates with $w=1$ is impossible.

ACKNOWLEDGMENTS

We wish to thank Professor L. Infeld and Professor P. Havas for their kind interest in this work, and J. Stachel and B. Mielnik for stimulating discussions.

APPENDIX

Consider an n-dimensional domain of coordinates x^A called Ω, with $(n-1)$-dimensional boundary $F\Omega$. The volume which one attributes to Ω is

$$V_\Omega=\int_\Omega d_n x. \quad (A.1)$$

Changing coordinates, i.e., variables of integration, from x^A to x'^A one gets

$$V_\Omega=\int_{\Omega'} d_n x' \frac{\partial(x)}{\partial(x')}. \quad (A.2)$$

For small changes of coordinates such that they preserve the shape of $F\Omega$, i.e., $\delta x^A=x^A(x')-x'^A=0$ for $x'\in F\Omega'$, and considering V_Ω as functional of $x^A(x'^B)$; we get

$$0=\delta V_\Omega=\int_{\Omega'} d_n x' \frac{\partial(x)}{\partial(x')}\frac{\partial x'^B}{\partial x^A}\frac{\partial \delta x^A}{\partial x'^B} =-\int_{\Omega'} d_n x' \frac{\partial}{\partial x'^B}\left(\frac{\partial(x)}{\partial(x')}\frac{\partial x'^B}{\partial x^A}\right)\delta x^A. \quad (A.3)$$

From the arbitrariness of Ω' and δx^A it follows identically that

$$\frac{\partial}{\partial x'^B}\left(\frac{\partial(x)}{\partial(x')}\frac{\partial x'^B}{\partial x^A}\right)=0; \quad (A.4)$$

these are the C. Neymann identities which could be checked by direct computation.

Exact Statistical Mechanics of a One-Dimensional System with Coulomb Forces

A. LENARD
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received March 27, 1961)

A system consisting of an equal number of positively and negatively charged "sheets" is considered in thermal equilibrium, with motion restricted to one dimension. The configurational part of the partition function can be represented as a sum of terms, each a simple algebraic expression. The summation is performed with the technique of generating functions. The asymptotic form in the limit of an infinite system is obtained from the pole of the generating function closest to the origin. This pole is the solution of a certain transcendental equation for which an explicit analytic representation in terms of an infinite continued fraction is available. It is shown that this equation is identical with the characteristic equation associated with the even Mathieu functions of even order.

In the limit, when the ratio of interparticle force to pressure is small, the system behaves as an ideal gas, the deviations from this state being expandable in powers of the square root of this ratio. In the opposite limit of large ratio, the particles associate in pairs of opposite charge, thus behaving like an ideal gas of neutral "molecules" which have an internal vibrational degree of freedom.

The analysis may be generalized to include the effect of a constant external electric field. For a given pressure there is a critical field which can never be surpassed without disrupting equilibrium.

1. INTRODUCTION

THE calculation of thermodynamic properties of systems from their atomic properties is generally frustrated by two difficulties. The first one is analytical: The computation or, at least, approximation of certain integrals ("cluster" integrals, contributions of "graphs," etc.) in terms of which the partition function is expressed algebraically. The other difficulty is algebraic, and is frequently associated with very complicated combinatorial problems, i.e., the systematic classification and eventual summing up of these contributions to the partition function. Usually these problems can be overcome only in the limit when some physical parameter is very small or large, and it is only in a few examples when an exact solution, encompassing the whole range of macroscopic parameters, is possible. This situation lends some interest to the study of simplified models which, while not realistic, still show some features of physical problems and, at the same time, are susceptible to exact analysis. It was in this spirit that one-dimensional systems have been investigated in the past.

The subject of the present paper is the statistical mechanical study of a one-dimensional system with a particularly simple long-range interaction law. It turns out that a complete discussion of the macroscopic behavior of this system is possible in spite of the fact that the statistical mechanics—at least in its combinatorial aspects—is not entirely trivial. We shall see that the absence of analytic problems follows from the elementary nature of all integrals into which the partition function can be decomposed. The combinatorics, in turn, can be mastered by special methods whose exposition is the principal purpose of this article.

There is added interest in this model on account of the long range of the interparticle force. The force law of the model to be discussed can be "physically" realized. It is the law of force between parallel, infinite, electrically charged plane condenser plates. Our system is then an assembly of a great number of such plates or sheets (referred to in the following as "particles") all restricted in their motion to the direction perpendicular to their planes, freely able to cross each other, and carrying fixed electrical surface charges (Sec. 2). It turns out that for the discussion of the statistical mechanics of this model the most useful ensemble is not the canonical ensemble of Gibbs, but rather a modified ensemble in which the pressure and temperature are given fixed values and the volume is derived as a phase average. While the general theory of such an ensemble has been given before,[1] it is probably not so familiar that a short summary should be superfluous (Sec. 3).

It will be shown how the partition function can be broken up into additive contributions of simple algebraic expressions each of which represents the phase integral over a certain region in phase space, a "configuration" (Sec. 4). The combinatorial problem is then the problem of summing these expressions over all configurations. With the aid of a simple pictorial representation of configurations we shall show that the natural technique for handling these sums is the method of generating functions (Sec. 5). After investigating the analytic nature of these generating functions we are led directly to the asymptotic form of the partition function for the limit of a very large number of particles. The limiting thermodynamic potential per particle appears to be simply expressible in terms of the solution of a certain transcendental equation whose left-hand side can be represented as an infinite continued fraction (Sec. 6). We shall then establish the correspondence between this equation and a classical boundary value problem of the Sturm-Liouville type which has been investigated in great detail (Sec. 7).

Making use of this mathematical correspondence, we shall then investigate the thermodynamics in two limiting cases (Secs. 8 and 9). The physical significance of certain "irreducible" configurations will be discussed

[1] W. B. Brown, Mol. Phys. 1, 68 (1958).

(Sec. 10). Finally, the generalization will be indicated to the case when an external electric field is present. This can be treated without new methods and with little extra complication (Sec. 11).

After this paper had been completed, the author learned that Professor S. Prager of the University of Minnesota had independently and simultaneously treated the present problem. This work is to be published in the 1961 volume of *Advances in Chemical Physics* edited by I. Prigogine.

2. DEFINITION OF THE MODEL

Consider a system of plane sheets (referred to in the following as "particles") freely movable in the x direction. Suppose that they all carry a surface mass density of magnitude unity, and that the ith one carries an electric surface charge density σ_i. The Newtonian equations of motion for this dynamical system are

$$d^2q_i/dt^2=2\pi\sigma_i\sum_j\sigma_j\,\mathrm{sg}(q_i-q_j), \tag{1}$$

where q_i denotes the position of the ith particle, and

$$\mathrm{sg}(x)=\begin{cases}1 & x>0\\ 0 & x=0\\ -1 & x<0.\end{cases}$$

The terms in the sum represent the electric force exerted by the particles on a selected one. The analytic simplicity of the model is a consequence of this force law; the force on a particle is constant during an interval in which no other particle passes over it. While this fact seems not to be of much help in comprehending the gross features of a truly dynamical problem, it is indicative of great simplifications in the treatment of thermal equilibrium.

Our model is related to that investigated most recently by Dawson.[2] The difference is that we deal with particles of both signs of charge, whereas Dawson considers particles of one sign of charge embedded in an oppositely charged, uniform, neutralizing background. While for many purposes this difference is relatively insignificant, it is quite crucial for the possibility of calculating the partition function. This has its roots in the fact that in our model the potential energy is linear in displacements, while in Dawson's model it is quadratic.

For the purposes of statistical mechanics it is necessary to have the Hamiltonian function which leads to Eq. (1). One may verify directly that this is

$$H(q_1,q_2,\cdots;p_1,p_2,\cdots)=\tfrac{1}{2}\sum_i p_i^2-2\pi\sum_{i<j}\sum\sigma_i\sigma_j|q_i-q_j|. \tag{2}$$

To complete the specification of the model we stipulate that $2N$ particles are given, N of which are charged positively and N negatively. All charges have the same magnitude $|\sigma_i|=\sigma$. The sums in Eqs. (1) and (2) run from 1 to $2N$.

[2] J. M. Dawson, Phys. Rev. **113**, 383 (1959).

3. THE CONSTANT PRESSURE ENSEMBLE

The fundamental probability density in the phase space of the system will be taken as follows

$$\rho(q_1,\cdots,p_{2N})=\exp\{\theta^{-1}[G-Pq_{\max}-H(q_1,\cdots,p_{2N})]\}, \tag{3}$$

when $q_i\geq 0 (1\leq i\leq 2N)$; $\rho=0$ otherwise. $q_{\max}$ is the largest among the q_i; it is to be regarded as a phase function. θ, P, and G are numerical parameters which are identified with the temperature, the pressure, and the Gibbs potential, respectively. If $f=f(q_1,\cdots,p_{2N})$ is any phase function, its average is

$$\langle f\rangle=(N!)^{-2}\int_{-\infty}^{\infty}d^{2N}p\int_0^{\infty}d^{2N}q f\rho. \tag{4}$$

In particular $\langle 1\rangle=1$, and this normalization relation defines G as a function of P and θ. Further important quantities are the average volume[3] $V=\langle q_{\max}\rangle$, the average energy $U=\langle H\rangle$ and the entropy $S=\langle\ln\rho\rangle$. Between these quantities the following relations hold

$$S=-(\partial G/\partial\theta)_P, \tag{5}$$

$$V=(\partial G/\partial P)_\theta, \tag{6}$$

and

$$U=G-PV+\theta S. \tag{7}$$

The reader will notice that we do not limit the system to a finite region (the "box") in the space of the coordinates as it is done in the standard canonical ensemble of Gibbs. Our probability distribution corresponds to a physical situation where the system is bounded on one side by a rigid wall and on the other by a freely movable "piston," that is to say, a device which exerts a constant pressure on the particle with largest coordinate. Thus the "volume" $q_{\max}$ is not fixed but fluctuates, and only its average has thermodynamic significance. The relation with respect to volume of the constant pressure and the constant volume ensembles is closely analogous to the relation with respect to energy of the canonical and microcanonical ensembles.

The aim of the following calculations is to obtain G as a function of θ and P, so that Eqs. (5)–(7) can be used to determine the other thermodynamic functions. An important simplification can be achieved immediately. *G depends in a nontrivial way only on a single variable.* The reason for this is the fact that the potential energy is a homogeneous function of first degree in the coordinates. Let $\epsilon_i=\mathrm{sg}(\sigma_i)$ be dimensionless charge

[3] Extensive thermodynamic quantities refer to unit surface area of the sheets. Thus V has the dimension of length. θ has the dimension of energy per unit area.

numbers, and let $\gamma=P/2\pi\sigma^2$. Also introduce dimensionless coordinates $x_i=2\pi\sigma^2 q_i/\theta$. In terms of these we obtain

$$G=-\theta\ln\{[(2\pi\theta)^{\frac{1}{2}}]^{2N}(\theta/2\pi\sigma^2)^{2N}Q(\gamma)\}, \tag{8}$$

where

$$Q(\gamma)=(N!)^{-2}\int_0^\infty d^{2N}x \times\exp\{-\gamma x_{\max}-\Phi(x_1,\cdots,x_{2N})\} \tag{9}$$

and

$$\Phi(x_1,\cdots,x_{2N})=-\sum_{1\le i<j\le 2N}\sum\epsilon_i\epsilon_j|x_i-x_j|. \tag{10}$$

The dependence on the pressure is entirely contained in the function $Q(\gamma)$. The dimensionless ratio γ plays a very important role, as we shall see. The temperature dependence is given explicitly by Eq. (8). It is important to remember that we are only interested in the asymptotic form of the thermodynamic functions in the limit of a very large number of particles. In this limit, the quantity G as well as all other extensive quantities become proportional to N. Thus it is expedient to introduce the Gibbs potential per particle

$$\mu=\lim_{N\to\infty}(G/2N). \tag{11}$$

Ignoring irrelevant constants, we have then

$$\mu=\mu(\theta,P)=\tfrac{1}{2}\theta\ln[\alpha(\gamma)/\theta^3], \tag{12}$$

where the function $\alpha(\gamma)$ is defined by

$$\ln\alpha(\gamma)=-\lim_{N\to\infty}\frac{\ln Q}{N}. \tag{13}$$

The existence of this limit will be demonstrated later (Sec. 6). In terms of the function $\alpha(\gamma)$, the equation of state is then

$$v=\lim_{N\to\infty}(V/2N)=\frac{\theta}{P}\left(\frac{\gamma}{2}\frac{d\ln\alpha}{d\gamma}\right). \tag{14}$$

The internal energy per particle is

$$u=\lim_{N\to\infty}(U/2N)=\theta\left(\frac{3}{2}-\frac{\gamma}{2}\frac{d\ln\alpha}{d\gamma}\right). \tag{15}$$

These follow directly from Eqs. (5)–(7) and (12). We also have the exact relation

$$u+Pv=\tfrac{3}{2}\theta. \tag{16}$$

4. INTEGRATION OVER CONFIGURATIONS

Consider the potential energy function Φ defined by Eq. (10). In any region of coordinate space where the relative order according to magnitudes of the coordinates $x_1, x_2, \cdots, x_{2N}$ is fixed, Φ is a homogeneous linear function of these coordinates. The coefficients of this linear function depend on the order of the coordinates. A simple consideration shows that the integral (9) is elementary. This is so because the class of functions consisting of exponentials of linear functions multiplied by polynomials is closed under the operation of indefinite integration. That is, however, not the same thing to say that the task of evaluating Q is trivial, since it is necessary to develop a systematic algorithm to deal with the very great number of integrations involved. Fortunately this can be done in a way which we now proceed to explain.

Imagine coordinate space broken up into $(2N)!$ regions, each corresponding to a definite linear succession of the $2N$ particles. These regions can be united into groups, each group containing $(N!)^2$ regions differing from each other only in permutation of the positively charged particles among themselves and the negatively charged particles among themselves. Such a group of regions will be called a "configuration." There are

$$\frac{(2N)!}{(N!)^2}=(-4)^N\binom{-\frac{1}{2}}{N} \tag{17}$$

configurations. Each is characterized by a definite linear succession of plus and minus signs. The potential energy remains unchanged when positively or negatively charged particles are permuted among themselves. Hence, if we conventionally relabel the particles in each region so that

$$0\le x_1\le x_2\le\cdots\le x_{2N}, \tag{18}$$

the potential energy becomes a linear function of the x_i with coefficients that depend *on the configuration only*. We may then rewrite the formula (9) in the following way

$$Q=\sum_C\int_0^\infty dx_{2N}\int_0^{x_{2N}}dx_{2N-1}\cdots\int_0^{x_2}dx_1 \times\exp\{-\gamma x_{2N}-\Phi_C(x_1,x_2,\cdots,x_{2N})\}. \tag{19}$$

Here, the summation goes over all configurations. Φ_C is the potential energy function for a definite configuration. We shall now show how an extraordinarily simple form for Φ_C may be obtained.

A configuration is uniquely characterized by the sequence $\epsilon_1, \epsilon_2, \cdots, \epsilon_{2N}$, where N of the ϵ_i are $+1$, N are -1. An alternative sequence of integers which can be used equally well to characterize a configuration is defined by

$$\begin{cases}\nu_0=0\\ \nu_i=\sum_{k=1}^{i}\epsilon_k. \quad (1\le i\le 2N),\end{cases} \tag{20}$$

ν_i is the total charge, in units of σ, up to and including the ith particle. We may think of the ν_i as a sequence

of integers subject to the sole conditions

$$\begin{cases} \nu_0=\nu_{2N}=0, \\ |\nu_i-\nu_{i-1}|=1, \quad (1\leq i\leq 2N). \end{cases} \tag{21}$$

The utility of the ν_i lies in the result of the following simple transformation

$$\begin{aligned} \Phi_C &= -\sum_{1\leq i<j\leq 2N}\sum \epsilon_i\epsilon_j|x_i-x_j| \\ &= -\sum_{1\leq i\leq 2N} x_i\epsilon_i\Big(\sum_{1\leq j<i}\epsilon_j-\sum_{i<j\leq 2N}\epsilon_j\Big) \\ &= -\sum_{1\leq i\leq 2N} x_i(\nu_i-\nu_{i-1})(\nu_{i-1}+\nu_i) \\ &= \sum_{1\leq i\leq 2N} x_i(\nu_{i-1}^2-\nu_i^2). \end{aligned} \tag{22}$$

The potential energy appears as a sum of terms each depending *on the coordinate of one particle only*. That this must be so is indeed obvious physically; within a configuration a particle moves quite independently of the others, the force on it depending only on how many positive or negative particles are on either side of it.

It is convenient to write Φ_C with a slightly different arrangement of terms

$$\Phi_C=\sum_{1\leq i\leq 2N}\nu_{i-1}^2(x_i-x_{i-1}), \tag{23}$$

where $x_0=0$. This way of writing exhibits the non-negative nature of Φ, a fact which is not immediately visible from the definition (10). Now substitute (23) into (19); the integral then becomes the Laplace transform of the convolution of the functions $\exp(-\nu_i^2x)$ $(i=0, 1, 2, \cdots, 2N-1)$, and this is equal[4] to the product of the Laplace transforms of these functions

$$\prod_{0\leq i\leq 2N-1}\int_0^\infty dx\,\exp[-(\gamma+\nu_i^2)x] = \prod_{0\leq i\leq 2N-1}\frac{1}{\gamma+\nu_i^2}. \tag{24}$$

Therefore we have the explicit algebraic formula

$$Q=\sum_C \frac{1}{(\gamma+\nu_0^2)(\gamma+\nu_1^2)\cdots(\gamma+\nu_{2N-1}^2)}. \tag{25}$$

The sum over all configurations C is a sum over all sequences ν_i satisfying conditions (21).

Before considering the problem of this summation, we shall only remark that the motivation for the use of the constant-pressure ensemble was, of course, precisely the possibility of doing the integrals in this simple manner. The appropriateness of the method of Laplace transformation and its interpretation in terms of the constant-pressure ensemble for one-dimensional systems has been noted before.[5]

5. THE SUM OVER CONFIGURATIONS

In this section the dependence of Q on N, hitherto treated as a constant, becomes important. Therefore, in the following we shall indicate this by a subscript.

In order to investigate the nature of the sum (25), it is convenient to make use of a simple pictorial representation. Each configuration can be associated uniquely with a certain graph.

A graph is simply the plot of ν_i versus i. For the sake of easier visualization we draw this as a continuous function, but only the "vertices," i.e., points with integer coordinates, have a significance. A graph is thus a broken line consisting of sections of slope ± 1. Each factor in a product $\prod_i(\gamma+\nu_i^2)^{-1}$ is assigned to a definite vertex and vice versa. An example of a graph is shown on Fig. 1. This graph represents a product in the sum (25) for $N=9$. The first and the last vertices are fixed at $\nu=0$. Note that there is no factor associated with the the last vertex.

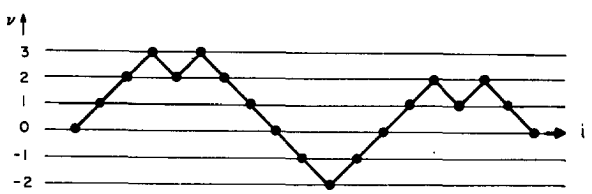

FIG. 1. The graph of the configuration $++{+}-{+}-{-}-{-}-{-}{+}{+}{+}{+}-{+}-{-}$.

The fundamental idea for grasping the nature of this summation is the decomposition of graphs into simpler subgraphs with a corresponding decomposition of the associated products.

We start with the concept of an "irreducible" configuration. A configuration is called irreducible if $\nu_i\neq 0$ for all $i\neq 0, 2N$, i.e., if its graph has only its endpoints in common with the i axis. Let $\bar{Q}_n$ stand for *the sum total of products belonging to all irreducible configurations with 2n vertices*. Consider now an arbitrary, in general reducible, configuration. Its graph can be regarded as a succession of irreducible graphs, each starting at one of the vertices with $\nu=0$. The expression associated with the whole graph is the product of the expressions associated with its irreducible components. We now carry out the summation over all graphs of a given length in three steps. First we hold the k (≥ 1) vertices with $\nu=0$ fixed and let the others vary in all possible ways without falling on the i axis. This gives a product

$$\bar{Q}_{n_1}\bar{Q}_{n_2}\cdots\bar{Q}_{n_k},$$

where $2n_1$ is the number of vertices in the first irre-

[4] D. V. Widder, *Advanced Calculus* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1947), Chap. XIII.

[5] Z. W. Salsburg, R. W. Zwanzig, and J. G. Kirkwood, J. Chem. Phys. 21, 1098 (1953), especially the end of this paper.

ducible component, $2n_2$ in the second, etc. Next, still holding k fixed, we sum over all possible *positive* integers $n_1, n_2, \cdots, n_k$, subject to the condition

$$n_1+n_2+\cdots+n_k=N. \tag{26}$$

Finally, we sum over k from 1 to N. It is clear that all configurations have been counted, and counted only once. Thus we get the following important combinatorial formula

$$Q_N=\sum_{k=1}^{N}\sum_{(n_1+n_2+\cdots+n_k=N)}\bar{Q}_{n_1}\bar{Q}_{n_2}\cdots\bar{Q}_{n_k}. \tag{27}$$

The physical significance of the quantities $\bar{Q}_n$ will be explored later (Sec. 8). The first few of them are

$$\begin{aligned}\bar{Q}_1&=\frac{2}{\gamma(\gamma+1)}\\ \bar{Q}_2&=\frac{2}{\gamma(\gamma+1)^2(\gamma+4)}\\ \bar{Q}_3&=\frac{2}{\gamma(\gamma+1)^2(\gamma+4)^2(\gamma+9)}+\frac{2}{\gamma(\gamma+1)^3(\gamma+4)^2}.\end{aligned} \tag{28}$$

We shall now show how the relationship (27) between the sequences Q_n and $\bar{Q}_n$ may be given a very simple form.

Let z be a complex variable, and let us introduce the *generating functions*

$$Q(z)\equiv 1+\sum_{n=1}^{\infty}Q_n z^n \tag{29}$$

and

$$\bar{Q}(z)\equiv\sum_{n=1}^{\infty}\bar{Q}_n z^n. \tag{30}$$

It is easy to see that, for sufficiently small $|z|$, these converge. We have, indeed, the following inequalities

$$\begin{aligned}\bar{Q}_n<Q_n<\frac{1}{(n!)^2}\int_0^\infty d^{2n}x e^{-\gamma x_{\max}}\\ =\frac{(2n)!}{(n!)^2}\frac{1}{\gamma^{2n}}=\left(-\frac{4}{\gamma^2}\right)^n\binom{-\frac{1}{2}}{n}.\end{aligned} \tag{31}$$

The integral is just Eq. (9) with Φ deleted, and the inequality follows from $\Phi\geq 0$. Therefore, the series for

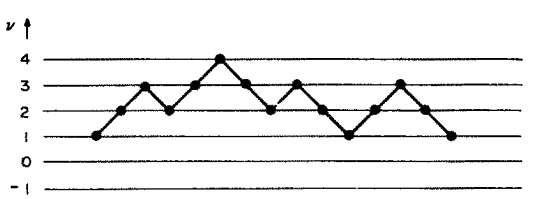

FIG. 2. A graph belonging to $G_8^{(1)}$.

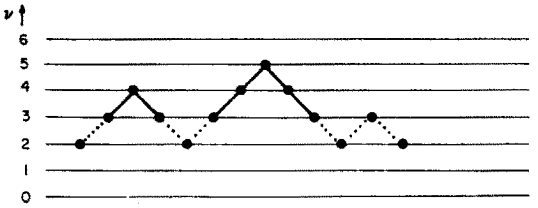

FIG. 3. A term in $G_7^{(2)}$ is equal to a term in the product $G_2^{(3)}G_3^{(3)}G_1^{(3)}/(\gamma+4)^4$.

$\bar{Q}(z)$ is majorized by that of $Q(z)$, and the latter is majorized by

$$\sum_{n=0}^{\infty}\left(-\frac{4z}{\gamma^2}\right)^n\binom{-\frac{1}{2}}{n}=\frac{1}{(1-4z/\gamma^2)^{\frac{1}{2}}}, \tag{32}$$

which converges for $|z|<\gamma^2/4$. Now multiply Eq. (27) by z^N and sum over N. A rearrangement of the order of summations is justified by the convergence of the series involved, and one obtains

$$Q(z)=[1-\bar{Q}(z)]^{-1}. \tag{33}$$

This simple analytic relation replaces the combinatorial formula (27).

It is convenient at this point to introduce the function $G^{(1)}(z)$ by the relation

$$\bar{Q}(z)=2\gamma^{-1}G^{(1)}(z) \tag{34}$$

and the corresponding sequence $G_n^{(1)}$ $(n=1, 2, \cdots)$ generated by it

$$G^{(1)}(z)=\sum_{n=1}^{\infty}G_n^{(1)}z^n. \tag{35}$$

In terms of graphs, $G_n^{(1)}$ is the sum total of the products from all graphs with $2n-1$ vertices such that *the minimal ν value $\nu=1$ is assumed at the first as well as at the last vertex (not excluding the possibility of $\nu=1$ in between)*. An example of such a graph is shown in Fig. 2. The factor γ^{-1} in Eq. (34) is just the single omitted vertex $\nu=0$ which would make such a graph represent an irreducible configuration. The factor 2 arises from the fact that such configurations occur in pairs related to each other by reflection of their graphs on the $\nu=0$ axis (or interchange of positive and negative charges).

In a similar way we may define generating functions $G^{(2)}(z)$, $G^{(3)}(z)$, $\cdots$ in terms of sequences $G_n^{(2)}$, $G_n^{(3)}$, $\cdots$ $(n=1, 2, \cdots)$. We define $G_n^{(r)}(r\geq 1)$ as the sum total of products from all graphs with $2n-1$ vertices such that *the minimal ν value $\nu=r$ is assumed at the first as well as at the last vertex (not excluding the possibility $\nu=r$ in between)*. We shall now show how the successive functions $G^{(1)}(z)$, $G^{(2)}(z)$, $\cdots$ are related to each other. If $n=1$, we have $G_n^{(r)}=1/(\gamma+r^2)$. Let then $n\geq 2$, and consider an arbitrary graph belonging to $G_n^{(r)}$ (see Fig. 3). Such a graph can be regarded as an alternation of vertices $\nu=r$ with smaller graphs that belong to some $G_m^{(r+1)}(1\leq m\leq n-1)$. We may now sum over all

graphs belonging to $G_n^{(r)}$ by first keeping the vertices with $\nu=r$ fixed and letting the other vertices vary in all possible ways so that for *them* $\nu\geq r+1$, and then letting the $\nu=r$ points vary in all possible ways. The reader will easily convince himself that the following combinatorial formula holds for $n\geq 2$,

$$G_n^{(r)}=\sum_{k=1}^{n-1}\frac{1}{(\gamma+r^2)^{k+1}}\times\sum_{(n_1+n_2+\cdots+n_k=n-1)}G_{n_1}^{(r+1)}G_{n_2}^{(r+1)}\cdots G_{n_k}^{(r+1)}. \tag{36}$$

Now, just as we have reduced Eq. (27) to the simple analytic relation (33) between the generating functions, so we may show by precisely the same sort of calculation that Eq. (36) implies

$$G^{(r)}(z)=\frac{z}{\gamma+r^2-G^{(r+1)}(z)}. \tag{37}$$

The interchange in orders of summation which is necessary for the derivation of (37) is justified by the convergence of the series involved. This follows from the obvious inequality

$$G_n^{(r+1)}<G_n^{(r)}, \tag{38}$$

which allows the series $G^{(r)}(z)$ to majorize $G^{(r+1)}(z)$. There is, therefore, a circle at least as large as $|z|=\gamma^2/4$, within which all series $G^{(r)}(z)$ converge.

The equation (37) which is of the nature of a recursion formula may be used to develop $G^{(1)}(z)$ into an infinite continued fraction. We have, in fact, proved the following

Theorem 1. *The configurational integral Q_N is the coefficient of z^N in the Maclaurin series of the function*

$$Q(z)=[1-\bar{Q}(z)]^{-1},$$

where $\bar{Q}(z)$ is the infinite continued fraction

$$\bar{Q}(z)=\cfrac{2}{\gamma}\,\cfrac{z}{\gamma+1-z-\cfrac{z}{\gamma+4-z-\cfrac{z}{\gamma+9-\cdots}}} \tag{39}$$

6. THE ANALYTIC NATURE OF $Q(z)$ THE ASYMPTOTIC LIMIT $N\to\infty$

It is now necessary to penetrate more deeply into the analytic nature of $Q(z)$. Cauchy's integral formula gives

$$Q_N=\frac{1}{2\pi i}\oint\frac{dz}{z^{N+1}}Q(z), \tag{40}$$

where the path of integration is a small closed curve surrounding the origin in a positive sense. We wish to study the asymptotic nature of Q_N as $N\to\infty$, and this requires some information on the singularities of $Q(z)$. It is well to point out here that the infinite continued fraction (39) is to be regarded as a formal expression. Its Maclaurin series defines an analytic function only wherever the series converges. Now, in the theory of continued fractions[6] such an expression is regarded as the limit of its finite approximating fractions which, in general, exists beyond the radius of convergence of the Maclaurin series and defines the analytic continuation of the latter. We shall, however, not need this theory, since all the desired results follow already from the more elementary treatment below. This consists of a series of simple propositions culminating in Theorem 2. Their proofs depend only on generally familiar facts of the theory of analytic functions.

Lemma 1. *Let $a_1, a_2, \cdots$ be a sequence of positive numbers, $a_i\geq a_1$, and let*

$$f(z)=\cfrac{z}{a_1-z-\cfrac{z}{a_2-z-\cfrac{z}{a_3-\cdots}}},$$

then the Maclaurin series of $f(z)$ converges at least inside the circle $|z|=\frac{1}{4}a_1^2$. For the proof let us compare the Maclaurin series of $f(z)$ with the series which is obtained when all a_i are replaced by a_1. From $a_i\geq a_1$ it follows that this comparison series has positive coefficients not smaller than those of $f(z)$. It is the Maclaurin series of

$$g(z)=\cfrac{z}{a_1-z-\cfrac{z}{a_1-\cdots}},$$

which is easily seen to be

$$g(z)=\tfrac{1}{2}[a_1-(a_1^2-4z)^{\frac{1}{2}}].$$

This series converges inside $|z|=\frac{1}{4}a_1^2$, hence so does $f(z)$.

Lemma 2. *Let $F(z,w)$ be a rational function of two variables, and let $f(z)$ be an analytic function, regular in a closed region R and not vanishing identically. Then the function $F(z,f(z))$ is an analytic function of z, regular in R with the possible exception of a finite number of poles.* $F(z,f(z))$ can have a singularity in R only where its denominator vanishes. But the latter is an analytic function in R, regular at all points, and not vanishing identically. Thus it can only have a finite number of zeros. These are the poles of $F[z,f(z)]$.

[6] H. S. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1948).

Lemma 3. *Let $f(z)$ be an analytic function vanishing at the origin and having a Maclaurin expansion with real positive coefficients. Furthermore let $f(z)$ have a pole on the circle of convergence of its Maclaurin series. Then the equation*

$$f(z)=1$$

has a real positive solution $z=z_0$ inside the circle of convergence, and any other solution of this equation has an absolute value greater than z_0. Let $z=a$ be the pole whose existence is assumed, and let $z\to a$ from the inside of the circle of convergence. Then $|f(z)|\to\infty$. But

$$f(|z|)\geq|f(z)|$$

because the coefficients of the series are positive. This shows that $f\to\infty$ as well when z moves to the radius of convergence on the positive real axis. Since $f(0)=0$, there must be a positive real value z_0 such that $f(z_0)=1$, and indeed only one, since f is monotone increasing on the positive real axis. Let $z=z_1$ be any other solution of the equation, which may be assumed to lie inside the circle of convergence, since otherwise the last statement of the lemma follows. But then z_1 cannot be real and positive, and we have the strict inequality

$$f(|z_1|)>|f(z_1)|=1=f(z_0),$$

showing that $|z_1|>z_0$. This proves Lemma 3.

Theorem 2. *$Q(z)$ is a meromorphic function. It has a simple pole $z=\alpha$ on the positive real axis such that all of its other poles have greater absolute value. Proof.* From Lemma 1 it follows that the function $G^{(r)}(z)$ has a Maclaurin expansion converging at least inside the circle $|z|=\frac{1}{4}(\gamma+r^2)^2$. $Q(z)$ may be represented as a rational function of z and $G^{(r)}(z)$ by means of Eqs. (33), (34), and the first $r-1$ of the Eqs. (37). Lemma 2 shows that $Q(z)$ is regular in the circle $|z|<\frac{1}{4}(\gamma+r^2)^2$ with the possible exception of a finite number of poles. But this circle can be chosen as large as desired by choosing r sufficiently large. Thus $Q(z)$ is regular in the whole open complex plane with the exception of poles,

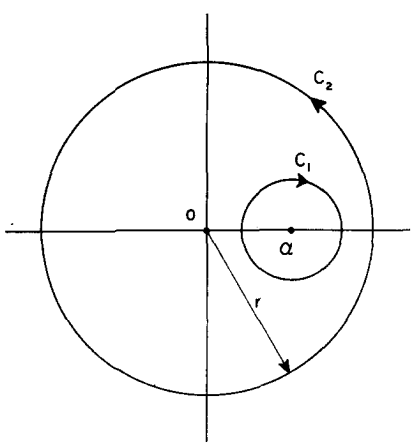

FIG. 4. Path of integration for Q_N.

i.e., it is meromorphic. Evidently the poles of $Q(z)$ are the solutions of the equation

$$\bar{Q}(z)=1. \tag{41}$$

But $\bar{Q}(z)$ is also a meromorphic function [and so are all the functions $G^{(r)}(z)$; for the same reason as $Q(z)$], so that it has a pole on the circle of convergence of its Maclaurin series. It satisfies the other conditions of Lemma 3 as well. The existence and properties of the number α of the theorem then follow from the conclusion of the lemma. That the pole of $Q(z)$ at $z=\alpha$ is simple follows from $\bar{Q}'(\alpha)>0$, which is quite evident as this quantity is a convergent series of positive terms. This completes the proof of Theorem 2.

After this preparation it is a simple matter to find the asymptotic behavior of Q_N as $N\to\infty$. In view of Theorem 2, the path of integration in Eq. (40) can be deformed into closed curves C_1 and C_2 with the following properties (Fig. 4). C_1 is any closed curve, traversed in the negative directions, and surrounding only one pole $z=\alpha$ of $Q(z)$. C_2 is a circle around the origin with a radius $r>\alpha$, such that it too surrounds only the pole $z=\alpha$. The integral over C_1 is just the residue at the pole, while the integral over C_2 is bounded in absolute value by $2\pi r$ multiplied by the maximum absolute value of the integrand. This leads immediately to the following asymptotic result[7]

$$Q_N\alpha^N=\frac{1}{\alpha\bar{Q}'(\alpha)}+O\{(\alpha/r)^N\},\quad(N\to\infty). \tag{42}$$

We actually need only a somewhat weaker statement, namely one obtained by taking the logarithm

$$\ln Q_N=-N\ln\alpha+O(1),\quad(N\to\infty). \tag{43}$$

This establishes the limit (13), and thus completes the computation of the thermodynamic potential (Sec. 3). We summarize the result in:

Theorem 3. *The thermodynamic properties of the system are given by the Gibbs potential per particle (12) where the quantity α is the smallest real positive root of Eq. (41) whose left-hand member is the infinite continued fraction (39).*

7. THE CHARACTERISTIC VALUE PROBLEM ASSOCIATED WITH EQ. (41)

The further study of the model then depends on investigating the properties of the function $\alpha=\alpha(\gamma)$. This function is monotone increasing, because the left-hand side of Eq. (41) $\bar{Q}=\bar{Q}(z,\gamma)$ is an increasing function of z on the positive real axis within the radius of convergence of the series (30), and a decreasing function of γ. It is easy to see by glancing at (39) that in the

[7] We use the symbol O in the mathematically precise sense of indicating a quantity less in absolute value than the argument multiplied by some number independent of the variable which goes to the limit.

limit $\gamma \to 0$ we must have $\alpha(\gamma) \sim \gamma/2$. The behavior in the opposite limit $\gamma \to \infty$ is a more sophisticated question. One is tempted to say, in a heuristic manner, that the denominators $\gamma+1$, $\gamma+4$, $\cdots$ should all be replaced by γ in the limit $\gamma \to \infty$. This leads to replacing $\bar{Q}$ by the periodic infinite fraction

$$\cfrac{2}{\gamma}\,\cfrac{z}{\gamma-\cfrac{z}{\gamma-\cdot}\,\cdot}\,\cdot = 1-(1-4z/\gamma^2)^{\frac{1}{2}}, \tag{44}$$

so that solving Eq. (41) one gets $\alpha(\gamma) \sim \frac{1}{4}\gamma^2$ $(\gamma \to \infty)$. However, it is clear that for *some* denominators this replacement is not justified, no matter how large (but finite) γ is, so that this simple argument is certainly incorrect.[8] It happens that the result just "derived" is correct, but a more refined asymptotic analysis is needed to establish it rigorously. There is, of course, the further problem of obtaining not just the dominant term but correction terms in this limiting case.

In this respect, it is fortunate that Eq. (41) plays an important role in a classical characteristic value problem of the Sturm-Liouville type which has been investigated very thoroughly in the mathematical literature. Thus it turns out that—from our point of view—it is only necessary to establish this connection, and then the known properties of the equation and its solution may be taken over and used to find the thermodynamic implications. For reasons that will become apparent in a moment it is not $\alpha=\alpha(\gamma)$ but the inverse function $\gamma=\gamma(\alpha)$ that is usually studied. Not only are the asymptotic series of this function known for the limits $\alpha \to 0$ and $\alpha \to \infty$, but it has even been tabulated numerically.

Let, temporarily, $\sqrt{\alpha}$ and γ be two independent positive numbers, and consider the problem of fixing them so that the differential equation

$$\left\{\frac{d^2}{dt^2}+2\alpha^{\frac{1}{2}}\cos t-\gamma\right\}y(t)=0 \tag{45}$$

shall have an even solution of period 2π not vanishing identically. We shall show that the condition for this is precisely

$$\bar{Q}(\alpha,\gamma)=1. \tag{41}$$

Indeed, let us expand $y(t)$ in a Fourier series

$$y(t)=\tfrac{1}{2}A_0+\sum_{r=1}^{\infty} A_r \cos rt, \tag{46}$$

and substitute into the differential equation. A simple calculation shows that the coefficients must be related by

$$\begin{cases} \tfrac{1}{2}\gamma A_0=\alpha^{\frac{1}{2}}A_1 \\ (\gamma+r^2)A_r=\alpha^{\frac{1}{2}}(A_{r-1}+A_{r+1}), \quad (r\geq 1). \end{cases} \tag{47}$$

On account of the homogeneous nature of the problem, only the ratios of the coefficients need be determined; the successive ratios are then related by

$$\frac{\alpha^{\frac{1}{2}}A_r}{A_{r-1}}=\cfrac{\alpha}{\gamma+r^2-\alpha^{\frac{1}{2}}A_{r+1}/A_r}, \quad (r\geq 1). \tag{48}$$

We now note that this sequence of equations is identical with (37) provided there we set $z=\alpha$; hence we get

$$\alpha^{\frac{1}{2}}A_r/A_{r-1}=G^{(r)}(\alpha),$$

$$A_r=A_0\frac{G^{(1)}(\alpha)G^{(2)}(\alpha)\cdots G^{(r)}(\alpha)}{\alpha^{\frac{1}{2}r}}. \tag{49}$$

If we now substitute A_1 thus obtained into the first of Eq. (47) we have precisely Eq. (41), as was to be shown. To justify the validity of these formal manipulations, it is necessary to show that the Fourier series (46) converges, together with at least its first two derivatives. For this purpose we may use the inequality

$$G^{(r)}(z)<\tfrac{1}{2}\{\gamma+r^2-[(\gamma+r^2)^2-4z]^{\frac{1}{2}}\}, \tag{50}$$

derived in the way indicated by the proof of Lemma 1. It is valid for all z for which the right-hand side has a convergent Maclaurin expansion; in particular, for $z=\alpha$ if r is sufficiently large. Furthermore, $1-[1-x]^{\frac{1}{2}}<x$ for all positive $x<1$; so that

$$G^{(r)}(\alpha)<\frac{2\alpha}{\gamma+r^2}<\frac{2\alpha}{r^2}, \quad (r \text{ sufficiently large}).$$

This gives the asymptotic upper bound

$$A_r=0\{(2\sqrt{\alpha})^r/(r!)^2\}, \quad (r\to\infty), \tag{51}$$

which shows that the Fourier series (46) converges fast enough so that it can be differentiated term by term any number of times.

The differential equation (45) is known as the Mathieu equation.[9] Its even periodic solutions are the functions ce_0, ce_2, ce_4, $\cdots$ in the notation of reference 9. We are interested in that solution for which the two parameters in the equation γ and $2\sqrt{\alpha}$ tend simultaneously to zero, i.e., the function $y(t)$ degenerates into a constant. This property identifies uniquely the function ce_0. The situation is summarized as follows:

Theorem 4. *The relation between the quantities α and γ (see Theorem 3) is the same as the relation between*

[8] An indication of trouble is that (44) has a branch point and not a pole on its circle of convergence. Furthermore, the solution of Eq. (41) appears *on* the circle of convergence and not inside of it, as it must for any γ no matter how large.

[9] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, New York, 1927), Chap. XIX.

the two parameters of the Mathieu equation (45) when the solution is the even Mathieu function of zero order.[10]

8. THE LIMIT $\gamma\to\infty$

We shall first consider the limit in which the effect of the interparticle force is small. This is the limit in which the mean kinetic energy is large compared to the mean potential energy. Another way of expressing the same state of affairs is to say that the pressure P is large compared to the "pressure" $2\pi\sigma^2$ exerted by one particle on another. This is the limit $\gamma=P/2\pi\sigma^2\to\infty$.

We have already remarked (cf. Sec. 7 above) that this limit presents peculiar difficulties when looked at purely from the point of view of solving Eq. (41). The correspondence with the Mathieu equation (Theorem 4) throws some further light on why this is so. The reader will observe that since we expect $\gamma\sim 2\sqrt{\alpha}$ as $\alpha\to\infty$, the problem of Eq. (45) becomes one of finding the solution of a differential equation whose highest derivative term is multiplied by a small parameter. What is involved is an asymptotic analysis of the way in which the nature of the equation and its solution changes in the limit.[11]

This has first been done by Ince[12]; and we shall not attempt to reproduce his somewhat complicated analysis, but merely quote his result[10]:

$$\gamma=2\alpha^{\frac{1}{2}}-\alpha^{\frac{1}{4}}+\frac{1}{16}+\frac{1}{256\alpha^{\frac{1}{4}}}+\cdots,\quad(\gamma,\alpha\to\infty).\tag{52}$$

This formula can be inverted

$$\alpha=\frac{\gamma^2}{4}\left\{1+(8\gamma)^{-\frac{1}{2}}+(32\gamma)^{-1}-\frac{1}{256\sqrt{2}\gamma^{\frac{3}{2}}}+\cdots\right\}^4.\tag{53}$$

Now substitute into (12), this gives

$$\mu=\frac{\theta}{2}\left\{-3\ln\theta+2\ln\gamma+\left(\frac{2}{\gamma}\right)^{\frac{1}{2}}-\frac{1}{8\gamma}+\frac{1}{192\sqrt{2}\gamma^{\frac{3}{2}}}+\cdots\right\}.\tag{54}$$

We shall recognize the meaning of the dominant term by calculating the equation of state (14)

$$\frac{Pv}{\theta}=1-\frac{1}{(8\gamma)^{\frac{1}{2}}}+\frac{1}{16\gamma}-\frac{1}{256\sqrt{2}\gamma^{\frac{3}{2}}}+\cdots,\tag{55}$$

which shows that the system with increasing γ tends more and more to an ideal gas state. It is characteristic that corrections are expandable in inverse powers of $\sqrt{\gamma}$. The internal energy per particle is given by Eq. (16)

$$u=\frac{\theta}{2}\left\{1+\frac{1}{(2\gamma)^{\frac{1}{2}}}-\frac{1}{8\gamma}+\frac{1}{128\sqrt{2}\gamma^{\frac{3}{2}}}+\cdots\right\}.\tag{56}$$

The first term is just the mean kinetic energy, the rest the mean potential energy. If we write the latter as $\sigma\phi$ where ϕ is the mean electric potential at the location of a particle, we get for the dominant term

$$\phi=\tfrac{1}{2}\theta(\pi/P)^{\frac{1}{2}}+\cdots.\tag{57}$$

9. THE LIMIT $\gamma\to 0$

In some ways the investigation of the opposite limit $\gamma\to 0$ is simpler; $\alpha=\alpha(\gamma)$ can be expanded in an asymptotic series in powers of γ. Making use of the theory of Mathieu functions we could, of course, quote the pertinent results; but in view of the simplicity, it is worth while to give here the derivation of at least the first two terms in this series.

Consider two functions

$$X(z)=\frac{2}{\gamma}\cfrac{z}{\gamma+1-[z/(\gamma+4)]}$$

and

$$Y(z)=\frac{2}{\gamma}\cfrac{z}{\gamma+1-z}\cfrac{}{\gamma+4-z}\cfrac{}{\gamma+1-z}\cfrac{}{\gamma+4-\cdots}$$

$$=\frac{1}{\gamma}\{\gamma+4-[(\gamma+4)^2-4z/(\gamma+1)]^{\frac{1}{2}}\}.$$

For real positive z, small enough that the Maclaurin series of $Y(z)$ converges, we have the inequalities [cf. Eq. (39)]

$$X(z)<\bar{Q}(z)<Y(z).\tag{58}$$

Thus α, which is the root of the equation $\bar{Q}(z)=1$, is bounded below by the root of $Y(z)=1$, and above by the root of $X(z)=1$. The latter two equations can be solved explicitly, and one obtains the inequalities

$$\frac{\gamma(\gamma+1)(\gamma+8)}{4(\gamma+4)}<\alpha(\gamma)<\frac{\gamma(\gamma+1)(\gamma+4)}{3\gamma+8}.\tag{59}$$

These hold for all γ. Now, the Maclaurin series in powers of γ of the two extreme members of (59) have the same first two terms; hence, by letting $\gamma\to 0$, we see that these are also the first two terms in the development of α

$$\alpha(\gamma)=\tfrac{1}{2}\gamma+\tfrac{7}{16}\gamma^2+0(\gamma^3).\tag{60}$$

The thermodynamic functions are then given by the

[10] The reader should be warned that there is no standard normalization for the Mathieu functions. Therefore, we have refrained from writing out the relation $\gamma=\gamma(\alpha)$ explicitly. This is easily done by comparing our Eq. (45) with the standardized form adopted by any particular author.

[11] See also remarks of footnote 8.

[12] E. L. Ince, Proc. Roy. Soc. Edinburgh **46**, 316 (1926).

formulas of Sec. 3: The Gibbs potential per particle

$$\mu=\tfrac{1}{2}\theta\{-3\ln\theta+\ln\tfrac{1}{2}\gamma+\tfrac{7}{8}\gamma+0(\gamma^2)\}, \tag{61}$$

the equation of state

$$Pv/\theta=\tfrac{1}{2}+\tfrac{7}{16}\gamma+\cdots, \tag{62}$$

and the internal energy per particle

$$\mu=\theta\{1-\tfrac{7}{16}\gamma+\cdots\}. \tag{63}$$

In the first approximation, the system appears as an ideal gas of N (and not $2N$) particles. This at the first sight surprising result can be explained as follows. For small values of the parameter γ, those parts of coordinate space have a relatively high probability for which the potential energy is close to its minimum value. Formula (23) shows that this is the case, in the first instance, for configurations for which the numbers ν_i^2 assume their minimum value. The restrictions (21) imply that these are the configurations for which

$$\nu_i=\begin{cases}0 & (i \text{ even})\\ \pm 1 & (i \text{ odd}).\end{cases} \tag{64}$$

An example is shown on Fig. 5. These are the configurations in which *particles are arranged in successive pairs, each pair having zero total charge.* We may say that the

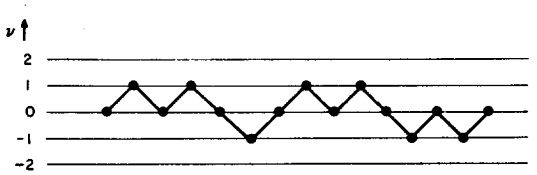

FIG. 5. A configuration of minimal potential energy.

system is an assembly of neutral "molecules" made up of two oppositely charged ions. In the configurations of the type (64), these molecules do not overlap, so that there is no force between them. Mathematically, restriction to these configurations means omitting all terms of the sum (27) save those with $n_1=n_2=\cdots=1$.

This interpretation in terms of neutral molecules is helpful even beyond the dominant terms of the limit $\gamma\to 0$. For instance, if we take the next term into account we have the effect of "binary interactions." The reader will have no difficulty in checking that, in this approximation, irreducible configurations contributing to $\bar{Q}_1$ and $\bar{Q}_2$ must be taken into account, the latter being the contribution from overlap-configurations of two molecules. In still higher approximations, triple interactions etc., come into play. This train of thought leads to a better understanding of the opposite limit $\gamma\to\infty$ too (Sec. 8). There *the number of particles that are in simultaneous interaction is large.* We shall come back to this point further below (Sec. 10).

We conclude with a remark on the internal energy. The dominant term $2N\theta$ shows that, on the average, we have the same amount $2N(\theta/2)$ of kinetic energy as potential energy. However, only half of this, $N\theta/2$, is the kinetic energy of the translational degrees of freedom of the molecules. The fact that, for the internal motion, the mean potential energy is twice the mean kinetic energy is the consequence of the proportionality of the potential energy to the absolute value of the separation.

10. SIGNIFICANCE OF IRREDUCIBLE CONFIGURATIONS

We have already seen that the concept of an irreducible configuration plays an important role both in the mathematics (Sec. 5) and the physical interpretation (Sec. 9). Here we shall consider some further points that illuminate the statistical structure of the system.

The reason for the importance of the irreducible configurations is that they are the *minimal electrically neutral parts* of a general configuration. They cannot be further subdivided into successive neutral parts. They are the noninteracting building blocks of the system. In contrast to other systems that can somehow be decomposed into noninteracting entities, one must remember that this decomposition for our system is not determined locally. It is unique only if the *whole* system is so decomposed. An extra positive particle on one end and an extra negative one on the other end will change totally the decomposition throughout the system. The physical reason for this is, of course, that the interparticle force is independent of distance. Nevertheless, the irreducible configurations should be thought of as having definite physical significance. Their average length is a measure of the efficacy of electrostatic shielding.

Let us introduce a random variable n_1 equal to half the number of particles in the irreducible configuration furthest to the left, say. n_1 is a definite phase function, constant over configurations (cf. Sec. 4), and indeed over all those configurations for which the first irreducible component has the same number of particles. It has thus a definite probability distribution which we shall now consider. *The probability that* $n_1=n$ $(n=1, 2, \cdots, N)$ *is that fraction of the sum* (27) *for which* $n_1=n$. In other words

$$\text{Prob}\{n_1=n\}=\bar{Q}_n Q_{N-n}/Q_N, \quad (1\leq n\leq N), \tag{65}$$

where $Q_0=1$ by convention. The reader will easily verify that the sum of these probabilities is 1 in virtue of Eq. (33). We are actually interested in the asymptotic form of this distribution as $N\to\infty$. This is quite simply obtained from the fundamental result (42)

$$P_n\equiv\lim_{N\to\infty}\text{Prob}\{n_1=n\}=\bar{Q}_n\alpha^n. \tag{66}$$

The normalization is now just the characteristic

equation (41). We can now form various moments

$$\langle n_1\rangle = \sum_{n=1}^{\infty} nP_n = \alpha\bar{Q}'(\alpha), \tag{67}$$

etc. All moments of this probability distribution are finite, because α is inside the radius of convergence of the series (30) and so term by term differentiation is justified any number of times. In fact, using an analysis quite analogous to that of Sec. 6, it is easy to show that P_n decreases exponentially with n in the limit $n\to\infty$. P_n is, of course, a function of γ, and when we want to call attention to this we will write $P_n(\gamma)$.

It is of interest now to ask about the limits of P_n as $\gamma\to\infty$ and $\gamma\to 0$. $P_n(\infty)$ can be derived from (66) and (53) together with the definition of $\bar{Q}_n$ as a sum over irreducible configurations (Sec. 5).

$$P_n(\infty) = \sum_{\substack{C \text{ irreducible} \\ \text{with } 2n \text{ vertices}}} \lim_{\gamma\to\infty} \left(\frac{\gamma^2}{4}\right)^n \times \frac{1}{(\gamma+\nu_0^2)(\gamma+\nu_1^2)\cdots(\gamma+\nu_{2n-1}^2)}$$

$$= \left(\frac{1}{4}\right)^n \times \begin{pmatrix}\text{number of irreducible configurations} \\ \text{with } 2n \text{ vertices}\end{pmatrix}.$$

The computation of the number in the parentheses is accomplished by the remark that its generating function is related to the generating function of the number (17) of all configurations with $2N$ vertices precisely as $\bar{Q}(z)$ is related to $Q(z)$, Eq. (33). These generating functions can be summed because they are binomial series, and one obtains after a simple calculation

$$P_n(\infty) = (-1)^{n-1}\binom{\frac{1}{2}}{n} = \frac{1}{2n-1}\,\frac{1\cdot 3\cdot 5\cdots(2n-1)}{2\cdot 4\cdot 6\cdots(2n)}. \tag{68}$$

This probability distribution can be interpreted very simply. $\gamma\to\infty$ is the limit in which the effect of the interparticle force vanishes (Sec. 8). Thus, all configurations become equally likely. In the limit, $N\to\infty$ when our system is infinitely large; this amounts to playing "heads and tails" with the successive signs of charge indefinitely. The question about the probability $P_n(\infty)$ of the first irreducible configuration having $2n$ vertices is then equivalent to asking: In a game of "heads and tails," what is the probability that the accumulated net gain of one side falls to zero for the first time after $2n$ tosses? This problem has been well discussed in elementary probability theory.[13]

[13] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, Inc., New York, 1957), Chap. III. Equation (4.7) of this chapter is identical with our (68).

This limiting probability distribution has a property that is important for a deeper understanding of the limit $\gamma\to\infty$ for our model. *All moments of (68) diverge.* The reason is that

$$P_n(\infty)\sim[2n(\pi n)^{\frac{1}{2}}]^{-1}, \quad (n\to\infty), \tag{69}$$

so that the decrease with n is not fast enough for the series expressing the moments to converge. But we have just seen that all moments of $P_n(\gamma)$ are finite for any finite γ. Thus we conclude

$$\begin{cases}\langle n_1\rangle, \quad \langle n_1^2\rangle, \cdots \to \infty \\ \quad \text{as} \quad \gamma\to\infty.\end{cases} \tag{70}$$

Now, the first moment $\langle n_1\rangle$ may be regarded as a measure of electrostatic shielding in the sense that it is the average number of particles within a distance over which electrostatic forces are neutralized (cf. discussion above). Hence, *in the limit $\gamma\to\infty$, the number of particles within the shielding distance tends to infinity.* This is one customary definition of this so-called "plasma limit."

The fact that the reach of the electrostatic forces becomes larger in the same limit as their effect becomes smaller is only seemingly paradoxical. It is related to the fact that in the coin tossing game the probability of long leads is relatively high.[14] Our model with γ finite corresponds to a game in which the probability of winning on a toss of coin is larger for the party that has accumulated a net loss. In such a game the persistence of long leads[14] would not occur. The electrostatic forces, by pulling in neutralizing charges, prevent their propagation through large distances.

Let us briefly look at the limit $\gamma\to 0$. For all n, the quantity $\gamma\bar{Q}_n(\gamma)$ remains finite in the limit. Since $\alpha\sim\gamma/2$, we therefore get

$$\begin{aligned} P_1(\gamma) &= 1-0(\gamma), \\ P_n(\gamma) &= 0(\gamma^{n-1}) \quad (n\geq 2), \end{aligned} \tag{71}$$

as $\gamma\to 0$. This result confirms the picture of the system in this limit, as discussed above (Sec. 9). It indicates that an approximation including terms of order γ^{n-1} corresponds to neglecting irreducible configuration of more than $2n$ particles.

11. EXTERNAL ELECTRIC FIELD

In closing we shall consider the generalization of the foregoing to the situation when the system is immersed in a constant, homogeneous electric field $\mathcal{E}$. The only change in the basic equation is that we add a term

$$-\mathcal{E}\sum_i \sigma_i q_i \tag{72}$$

to the Hamiltonian (2) representing the potential energy of the charged particles in the external field. The reader will verify without difficulty that the

[14] Reference 13, Sec. III.5.

important formula (23) is replaced by

$$\Phi_C = \sum_{1\le i\le 2N} (\nu_{i-1}^2 + 2\eta\nu_{i-1})(x_i - x_{i-1}), \qquad (73)$$

where $\eta = \mathcal{E}/(4\pi\sigma)$. It is clear that the whole of Secs. 4, 5, and 6 can be taken over with minor changes only. The factor belonging to a vertex is now

$$1/(\gamma + 2\eta\nu + \nu^2). \qquad (74)$$

Since this is not invariant under a change of sign of ν, we must treat the graphs below the $\nu=0$ axis separately from the graphs above the axis. It is not difficult to see that the formula which replaces (34) is now

$$\bar{Q}(z;\gamma,\eta) = \gamma^{-1}\{G^{(1)}(z;\gamma,\eta) + G^{(1)}(z;\gamma,-\eta)\}, \qquad (75)$$

where

$$G^{(1)}(z;\gamma,\eta) = \cfrac{z}{\gamma+2\eta+1-\cfrac{z}{\gamma+4\eta+4-\cfrac{z}{\gamma+6\eta+9-\ddots}}} \qquad (76)$$

As before, the Gibbs potential per particle is given by Eq. (12), where now $\alpha = \alpha(\gamma,\eta)$ is the smallest positive solution of $\bar{Q}(\alpha;\gamma,\eta) = 1$. The thermodynamic variable conjugate to the electric field is the electric dipole moment $M = \langle \sum_i \sigma_i q_i \rangle$. Its value per particle is

$$\lim_{N\to\infty} \frac{M}{2N} = -\frac{\theta}{8\pi\sigma}\left(\frac{\partial \ln\alpha}{\partial\eta}\right)_\gamma. \qquad (77)$$

There is an important restriction on the allowable values of γ and η. The convergence of all integrals in (19) requires that all factors (74) shall be positive. Thus γ and η must be such that *the quadratic form*

$$f(x) = x^2 + 2\eta x + \gamma \qquad (78)$$

shall be positive for all integer values of the variable x. An elementary consideration shows that the necessary and sufficient condition for this is

$$\gamma > \eta_0(2\eta - \eta_0), \qquad (79)$$

where $\eta_0 = \eta_0(\eta)$ is the nearest integer to η. The region allowed in the (γ,η) plane can be described in simple geometrical terms: It is bounded below by those tangents of the parabola $\gamma = \eta^2$ which touch at points with integer η values (Fig. 6). The physical explanation for this restriction is that for too large an electric field the external pressure cannot overcome the force of the field on a suitable charge fluctuation near the "piston."

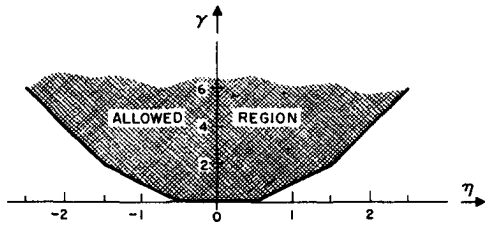

FIG. 6. Allowed region in the (γ,η) plane.

In closing we may mention that, here too, it is possible to convert the equation $\bar{Q}(\alpha) = 1$ into a characteristic value problem using a slight generalization of the technique of Sec. 7. A study of this problem would result in information on the thermodynamics of the system in the presence of an electric field.

General Spherical Harmonics Formulation of Plasma Boltzmann Equation

CHESTER P. CARPENTER,*
Avco Corporation, Electronics and Ordnance Division, Cincinnati, Ohio

AND

FRITZ W. MEZGER
General Electric Aircraft Nuclear Propulsion Department, Cincinnati 15, Ohio

(Received April 25, 1961)

The Boltzmann equation for the phase space distribution of electrons in the presence of ions is reduced to an infinite set of differential equations which do not involve angle variables. The usual method of expanding the electron phase space distribution function in terms of spherical harmonics is employed and it is assumed, in analyzing the scattering process, that the ion velocities can be neglected in comparison with the electron velocities. The expansion includes both polar and azimuthal angles obviating the assumption of symmetry about a polar axis made in previous work. The differential equation for the general component of the spherical harmonics expansion is derived and explicit equations for the first few components are presented. The component equations are seen to be considerably more tractable for cases which involve electric and/or magnetic fields along a single axis.

1. INTRODUCTION

THE spherical harmonics method of solving particle transport problems which are expressed in terms of a Boltzmann equation enjoys the important advantage of generality and suffers under the associated disadvantage of complexity. This inherent complexity can often be mitigated in the study of basic transport process by assuming symmetry about a polar axis. Thus, Allis,[1] in his study of the motion of ions and electrons, expands the electron phase space distribution function in spherical harmonics involving a polar angle only. In addition, he assumes the ion velocities to be negligible in comparison with the electron velocities in order to make the scattering integrals tractable.

Nevertheless, the assumption of symmetry abut a polar axis applies rigorously only to the cases of slab or spherical symmetry. For any other geometry, both polar and azimuthal angles must be considered and, as soon as they are, the formulation is valid for all geometries. The general spherical harmonics formulation of the Boltzmann equation in the absence of accelerating forces has been given by Weinberg and Wigner[2] in their discussion of the neutron transport problem. The general treatment of the plasma physics problem, for which the electrons are accelerated between collisions by a Lortenz force, has not heretofore been presented.

Previous developments of the spherical harmonics formulation of the plasma Boltzmann equation have been limited either in the sense of considering only a polar angle or of considering only the first few terms rather than the general term.[1,3–5] Recently, terms up to order three have been calculated by the use of a cartesian tensor scalar product equivalent to the spherical harmonics expansion.[6] The difficulty in the general program has involved the absence of appropriate recursion relations for the spherical harmonics in expressions involving their derivatives in combination with trigonometric functions of the polar and azimuthal angles. In this paper, recursion relations for some of the individual terms have been derived and, wherever this has not been done, it has proven possible to find the recursion relation for the actual combination of terms appearing in the spherical harmonics expansion of the Boltzmann equation.

This paper begins with the usual Boltzmann equation for electrons in the presence of ions which includes only elastic binary collisions between electrons and ions and which assumes the plasma is sufficiently diffuse so that individual electron-ion collisions can be distinguished from the interaction of the electron with the average local electromagnetic field. The equation is simplified by assuming the velocity of the ions can be neglected in calculating the scattering integrals. The distribution function of the electrons and the scattering cross section are expanded in spherical harmonics. By using recursion relations and the addition theorem for the spherical harmonics, the resulting equation is recast into a form for which all angularly dependent terms are expressed as linear combinations of spherical harmonics. From this point, the orthogonality of the spherical harmonics makes it a straightforward matter to derive the general expression for the component equations of the spherical harmonics expansion of the electron phase space distribution function. The characteristics of the component equations are discussed briefly.

2. BOLTZMANN AND SPHERICAL HARMONIC EQUATIONS

The Boltzmann equation for the phase space distribution function for electrons $f(\mathbf{x},\mathbf{v},t)$ interacting through

* Graduate student at the University of Cincinnati under a grant given by the Avco Corporation.

[1] W. P. Allis, "Motion of ions and electrons," *Handbuch der Physik* (Springer-Verlag, Berlin, Germany, 1956), Vol. XXI.

[2] A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chain Reactions* (University of Chicago Press, Chicago, Illinois, 1958).

[3] H. Margenau, Phys. Rev. **69**, 508 (1946).

[4] P. Rosen, Phys. Rev. **103**, 390 (1956).

[5] D. C. Kelly, H. Margenau, and S. C. Brown, Phys. Rev. **108**, 1367 (1957).

[6] T. W. Johnson, Phys. Rev. **120**, 1103 (1960).

binary elastic collisions with ions with distribution function $F(\mathbf{x},\mathbf{V},t)$ in the presence of an electric field $\mathbf{E}$ and a magnetic induction $\mathbf{B}$ is

$$\left[\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f+\frac{e}{m_e}\left[\mathbf{E}+\frac{1}{c}(\mathbf{v}\times\mathbf{B})\right]\cdot\nabla_v f\right]d^3v$$

$$=\int F(\mathbf{V}')f(\mathbf{v}')|\mathbf{v}'-\mathbf{V}'|\sigma(|\mathbf{v}'-\mathbf{V}'|,\chi)d^2\omega' d^3V' d^3v'$$

$$-\int F(\mathbf{v})f(\mathbf{v})|\mathbf{v}-\mathbf{V}|\sigma(|\mathbf{v}-\mathbf{V}|,\chi)d^2\omega' d^3V d^3v, \quad (1)$$

where the primed and unprimed velocities are connected by the usual expressions for an elastic collision, e/m_e is the ratio of charge to mass of an electron, c is the speed of light, σ is the differential scattering cross section, $d^2\omega'$ is an element of solid angle where

$$\omega'=(\mathbf{v}'-\mathbf{V}')/|\mathbf{v}'-\mathbf{V}'|, \quad (2)$$

and χ is the scattering angle

$$\chi=\cos^{-1}(\boldsymbol{\omega}\cdot\boldsymbol{\omega}'). \quad (3)$$

As shown by Allis,[1]

$$d^3v'=(|\mathbf{v}'-\mathbf{V}'|/|\mathbf{v}-\mathbf{V}'|)^3 d^3v. \quad (4)$$

Thus, Eq. (1) can be rewritten

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f+\frac{e}{m_e}\left[\mathbf{E}+\frac{1}{c}(\mathbf{v}\times\mathbf{B})\right]\cdot\nabla_v f$$

$$=\int F(\mathbf{V}')f(\mathbf{v}')\frac{|\mathbf{v}'-\mathbf{V}'|^4}{|\mathbf{v}-\mathbf{V}'|^3}\sigma(|\mathbf{v}'-\mathbf{V}'|,\chi)d^2\omega' d^3V'$$

$$-\int F(\mathbf{V})f(\mathbf{v})|\mathbf{v}-\mathbf{V}|\sigma(\mathbf{v}-\mathbf{V},\chi)d^2\omega' d^3V. \quad (5)$$

Making the assumption that the plasma is reasonably close to thermal equilibrium and consequently that the ratio of electron to ion velocities approximates the square root of the ratio of ion to electron masses, it is seen that $\mathbf{V}\ll\mathbf{v}$ and the calculation of the scattering integrals is simplified by writing,

$$F(\mathbf{V})\equiv F(\mathbf{x},\mathbf{V},t)=N(\mathbf{x},t)\delta(\mathbf{V}), \quad (6)$$

with a similar expression for $F(\mathbf{V}')$. Since for an elastic collision $|\mathbf{v}-\mathbf{V}|=|\mathbf{v}'-\mathbf{V}'|$ and since substituting Eq. (6) into Eq. (5) has the effect of setting $\mathbf{V}=\mathbf{V}'=0$, it is seen that the Boltzmann equation which serves as the basis for this spherical harmonics calculation is

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f+\frac{e}{m_e}\left[\mathbf{E}+\frac{1}{c}(\mathbf{v}\times\mathbf{B})\right]\cdot\nabla_v f$$

$$=Nv\int f(\mathbf{v}')\sigma(v',\chi)d^2\Omega'-Nvf\int\sigma(v,\chi)d^2\Omega', \quad (7)$$

where the vector $\boldsymbol{\omega}'$ has become the vector

$$\boldsymbol{\Omega}'=\mathbf{v}'/|\mathbf{v}'|=\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta), \quad (8)$$

which is simply the vector determining the direction of electron velocity, and the scattering angle, Eq. (3) becomes $\chi=\cos^{-1}(\boldsymbol{\Omega}\cdot\boldsymbol{\Omega}')$.

The same spherical harmonics notation used by Weinberg and Wigner[2] is adopted in order to facilitate comparison with their results for the Lorentz force-free case. Thus, the distribution function expansion is written

$$f(\mathbf{x},\mathbf{v},t)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}f^{l,m}(\mathbf{x},v,t)P_{l,m}(\boldsymbol{\Omega}), \quad (9)$$

where

$$P_{l,m}(\boldsymbol{\Omega})=\frac{e^{im\phi}(-\sin\theta)^m}{l!2^l}\left[\frac{(l-m)!}{(l+m)!}\right]^{\frac{1}{2}}\frac{d^{l+m}(\cos^2\theta-1)^l}{(d\cos\theta)^{l+m}}, \quad (10)$$

$$P_{l,-m}=(-)^m P_{l,m}{}^*, \quad (11)$$

$$\int P_{l',m'}P_{l,m}{}^*d^2\Omega=\frac{4\pi}{2l+1}\delta_{ll'}\delta_{mm'}, \quad (12)$$

where the asterisk refers to complex conjugate. Similarly, the scattering cross section is written as,

$$\sigma(v,\chi)=\sum_{l=0}^{\infty}\sigma_l(v)P_l(\cos\chi), \quad (13)$$

where P_l is the usual Legendre polynomial and

$$P_l(\cos\theta)=P_{l,0}(\boldsymbol{\Omega}) \quad (14)$$

and is related to the $P_{l,m}$ by the addition theorem

$$P_l(\cos\chi)=\sum_{m=-1}^{+1}P_{l,m}(\boldsymbol{\Omega})P_{l,m}{}^*(\boldsymbol{\Omega}'). \quad (15)$$

Since the scattering cross section is regarded as known, one can write

$$\sigma_l(v)=\frac{2l+1}{4\pi}\int P_l(\cos\chi)\sigma(v,\chi)d^2\Omega \quad (16)$$

by applying the orthogonality relationship Eq. (12) to Eq. (13) after multiplying by $P_l(\cos\chi)$.

The spherical harmonics expansion of the Boltzmann equation Eq. (7) is carried out by considering the time and configuration derivative terms, the velocity derivative terms, the electric force term, the magnetic force term, and the scattering terms in order.

3. TIME CONFIGURATION DERIVATIVES

Substituting Eq. (9) in the expression for the time and configuration derivatives, one obtains

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f=\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}P_{l,m}\left[\frac{\partial f^{l,m}}{\partial t}+v\sin\theta\cos\phi\frac{\partial f^{l,m}}{\partial x}+v\sin\theta\sin\phi\frac{\partial f^{l,m}}{\partial y}+v\cos\theta\frac{\partial f^{l,m}}{\partial z}\right]. \quad (17)$$

Noting that[7]

$$\cos\theta P_{l,m}=\frac{1}{2l+1}[A_l{}^mP_{l+1,m}+B_l{}^mP_{l-1,m}], \quad (18)$$

$$\sin\theta e^{i\phi}P_{l,m}=\frac{1}{2l+1}[-C_l{}^mP_{l+1,m+1}+D_l{}^mP_{l-1,m+1}], \quad (19)$$

$$\sin\theta e^{-i\phi}P_{l,m}=\frac{1}{2l+1}[E_l{}^mP_{l+1,m-1}+F_l{}^mP_{l-1,m-1}], \quad (20)$$

where

$$A_l{}^m=[(l-m+1)(l+m+1)]^{\frac{1}{2}} \quad (21)$$

$$B_l{}^m=[(l+m)(l-m)]^{\frac{1}{2}} \quad (22)$$

$$C_l{}^m=[(l+m+2)(l+m+1)]^{\frac{1}{2}} \quad (23)$$

$$D_l{}^m=[(l-m)(l-m-1)]^{\frac{1}{2}} \quad (24)$$

$$E_l{}^m=[(l-m+2)(l-m+1)]^{\frac{1}{2}} \quad (25)$$

$$F_l{}^m=[(l+m)(l+m-1)]^{\frac{1}{2}}, \quad (26)$$

one can combine Eqs. (19) and (20) to show

$$\sin\theta\cdot\cos\phi P_{l,m}=\frac{1}{2(2l+1)}[-C_l{}^mP_{l+1,m+1}+D_l{}^mP_{l-1,m+1}+E_l{}^mP_{l+1,m-1}-F_l{}^mP_{l-1,m-1}], \quad (27)$$

and

$$\sin\theta\cdot\sin\phi P_{l,m}=\frac{i}{2(2l+1)}[C_l{}^mP_{l+1,m+1}-D_l{}^mP_{l-1,m+1}+E_l{}^mP_{l+1,m-1}-F_l{}^mP_{l-1,m-1}]. \quad (28)$$

Substituting Eqs. (18), (27), and (28), one obtains the desired expression in which angular dependence is expressed in terms of a linear combination of associated spherical harmonics,

$$\frac{\partial f}{\partial t}+\mathbf{v}\cdot\nabla f=-\frac{v}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m+2)(l+m+1)]^{\frac{1}{2}}}{2l+1}\left(-\frac{\partial f^{l,m}}{\partial x}+i\frac{\partial f^{l,m}}{\partial y}\right)P_{l+1,m+1}+v\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m+1)(l-m+1)]^{\frac{1}{2}}}{2l+1}\times\frac{\partial f^{l,m}}{\partial z}P_{l+1,m}+\frac{v}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l-m+2)(l-m+1)]^{\frac{1}{2}}}{2l+1}\left(\frac{\partial f^{l,m}}{\partial x}+i\frac{\partial f^{l,m}}{\partial y}\right)P_{l+1,m-1}+\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{\partial f^{l,m}}{\partial t}P_{l,m}+\frac{v}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l-m)(l-m-1)]^{\frac{1}{2}}}{2l+1}\left(\frac{\partial f^{l,m}}{\partial x}-i\frac{\partial f^{l,m}}{\partial y}\right)P_{l-1,m+1}+v\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m)(l-m)]^{\frac{1}{2}}}{2l+1}\frac{\partial f^{l,m}}{\partial z}P_{l-1,m}-\frac{v}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m)(l+m-1)]^{\frac{1}{2}}}{2l+1}\left(\frac{\partial f^{l,m}}{\partial x}+i\frac{\partial f^{l,m}}{\partial y}\right)P_{l-1,m-1}. \quad (29)$$

Multiplying Eq. (29) by $P_{l,m}{}^*$ and integrating over all solid angles, Eq. (29) can be shown to be equivalent to the Weinberg and Wigner expression for the "streaming" terms.[8]

4. ELECTRIC FIELD TERMS

Noting that

$$\mathbf{E}\cdot\nabla_v f=E_v\frac{\partial f}{\partial v}+\frac{E_\theta}{v}\frac{\partial f}{\partial\theta}+\frac{E_\phi}{v\sin\theta}\frac{\partial f}{\partial\phi}, \quad (30)$$

where

$$E_v=E_z\cos\theta+E_x\sin\theta\cdot\cos\phi+E_y\sin\theta\cdot\sin\phi, \quad (31)$$

$$E_\theta=-E_z\sin\theta+E_x\cos\theta\cdot\cos\phi+E_y\cos\theta\cdot\sin\phi, \quad (32)$$

$$E_\phi=E_y\cos\phi=E_x\sin\phi, \quad (33)$$

the expansion of the electric force term can be written

$$\mathbf{E}\cdot\nabla_v f=(E_x\sin\theta\cos\phi+E_y\sin\theta\sin\phi+E_z\cos\theta)\cdot\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{\partial f^{l,m}}{\partial v}P_{l,m}+\frac{1}{v}(E_x\cos\theta\cos\phi+E_y\cos\theta\sin\phi-E_z\sin\theta)\times\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}f^{l,m}\frac{\partial}{\partial\theta}P_{l,m}+\frac{i}{v\sin\theta}\left[\frac{E_y}{2}(e^{i\phi}+e^{-i\phi})+\frac{iE_x}{2}(e^{i\phi}-e^{-i\phi})\right]\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}mf^{l,m}P_{l,m}, \quad (34)$$

[7] V. Rojansky, *Introductory Quantum Mechanics* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1938), p. 532.
[8] See p. 226 of reference 2.

where, from Eq. (10),

$$\frac{\partial}{\partial\phi}P_{l,m}=imP_{l,m}. \tag{35}$$

Defining,

$$G_l{}^m=[(l+m+1)(l-m)]^{\frac{1}{2}} \tag{36}$$

$$H_l{}^m=[(l-m+1)(l+m)]^{\frac{1}{2}}, \tag{37}$$

one can show[7]

$$m\cot\theta P_{l,m}=-\frac{e^{-i\phi}}{2}G_l{}^mP_{l,m+1}-\frac{e^{i\phi}}{2}H_l{}^mP_{l,m-1}, \tag{38}$$

$$\frac{d}{d\theta}P_{l,m}=\frac{e^{-i\phi}}{2}G_l{}^mP_{l,m+1}-\frac{e^{i\phi}}{2}H_l{}^mP_{l,m-1}. \tag{39}$$

Combining Eqs. (19), (20), and (39) one finds

$$\sin\theta\frac{d}{d\theta}P_{l,m}=\frac{1}{2l+1}[lA_l{}^mP_{l+1,m}-(l+1)B_l{}^mP_{l-1,m}]. \tag{40}$$

Multiplying Eq. (39) by $\cos\theta(e^{i\phi}+e^{-i\phi})$, and then by $\cos\theta(e^{i\phi}-e^{-i\phi})$, and using Eq. (18), one obtains the two expressions

$$2\cos\phi\cdot\cos\theta\frac{d}{d\theta}P_{l,m}=\frac{(1+e^{-2i\phi})}{2}\frac{G_l{}^m}{2l+1}[A_l{}^{m+1}P_{l+1,m+1}+B_l{}^{m+1}P_{l-1,m+1}]$$
$$-\frac{H_l{}^m}{2}\frac{(1+e^{2i\phi})}{2l+1}[A_l{}^{m-1}P_{l+1,m-1}+B_l{}^{m-1}P_{l-1,m-1}], \tag{41}$$

$$2i\sin\phi\cos\theta\frac{d}{d\theta}P_{l,m}=\frac{(1-e^{-2i\phi})}{2(2l+1)}G_l{}^m[A_l{}^{m+1}P_{l+1,m+1}+B_l{}^{m+1}P_{l-1,m+1}]$$
$$-\frac{H_l{}^m}{2(2l+1)}(e^{2i\phi}-1)[A_l{}^{m-1}P_{l+1,m-1}+B_l{}^{m-1}P_{l-1,m-1}]. \tag{42}$$

Replacing m by $m-1$ in Eq. (19) and m by $m+1$ in Eq. (20) and eliminating $P_{l-1,m}$ from the modified equations, one finds

$$\frac{P_{l+1,m}}{\sin\theta}=-\frac{1}{2m}[e^{-i\phi}D_l{}^{m-1}P_{l,m+1}+e^{i\phi}F_1{}^{m+1}P_{l,m-1}], \tag{43}$$

where use is made of the identity

$$D_l{}^{m-1}E_l{}^{m+1}-F_l{}^{m+1}C_l{}^{m-1}=-2m(2l+1). \tag{44}$$

After substituting Eqs. (18), (27), (28), (41), (42), and (43) into Eq. (34) one obtains

$$\mathbf{E}\cdot\nabla_v f=\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{1}{2(2l+1)}\frac{\partial f^{l,m}}{\partial v}[-C_l{}^mE_xP_{l+1,m+1}+2E_zA_l{}^mP_{l+1,m}+E_xE_l{}^mP_{l+1,m-1}+E_xD_l{}^mP_{l-1,m+1}$$
$$+2E_xB_l{}^mP_{l-1,m}-E_xF_l{}^mP_{l-1,m-1}]+\sum_{=0}^{\infty}\sum_{m=-l}^{m=l}\frac{1}{2(2l+1)}\frac{\partial f^{l,m}}{\partial v}[iE_yC_l{}^mP_{l+1,m+1}+iE_yE_l{}^mP_{l+1,m-1}$$
$$-iE_yD_l{}^mP_{l-1,m+1}-iE_yF_l{}^mP_{l-1,m-1}]+\frac{1}{4v}(E_x-iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{A_l{}^{m+1}G_l{}^m}{2l+1}f^{l,m}P_{l+1,m+1}-\frac{1}{v}E_z$$
$$\cdot\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{1}{2l+1}A_l{}^mf^{l,m}P_{l+1,m}-\frac{1}{4v}(E_x+iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{A_l{}^{m-1}H_l{}^m}{2l+1}f^{l,m}P_{l+1,m-1}$$

[*Equation (45) continued on next page.*]

$$+\frac{1}{4v}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}[B_l{}^{m+1}G_l{}^m(E_x+iE_y)+D_{l-1}{}^{m-1}(E_x-iE_y)]f^{l,m}P_{l-1,m+1}+\frac{1}{v}E_z\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{l+1}{2l+1}B_l{}^m f^{l,m}P_{l-1,m}$$

$$-\frac{1}{4v}(E_x+iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\left[\frac{B_l{}^{m-1}H_l{}^m}{2l+1}+F_{l-1}{}^{m+1}\right]f^{l,m}P_{l-1,m-1}$$

$$+\frac{1}{4v}(E_x+iE_y)e^{-2i\phi}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\left[\frac{A_l{}^{m+1}G_l{}^m}{2l+1}P_{l+1,m+1}+\left(\frac{B_l{}^{m+1}G_l{}^m}{2l+1}-D_{l-1}{}^{m-1}\right)P_{l-1,m+1}\right]f^{l,m}$$

$$-\frac{1}{4v}(E_x-iE_y)e^{2i\phi}\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\left[\frac{A_l{}^{m-1}H_l{}^m}{2l+1}P_{l+1,m-1}+\left(\frac{B^{m-1}H_l{}^m}{2l+1}-F_{l-1}{}^{m+1}\right)P_{l-1,m-1}\right]f^{l,m}, \quad (45)$$

where use has been made of the relation

$$e^{2i\phi}[F_l{}^mP_{l-1,m-1}-E_l{}^mP_{l+1,m-1}]=C_l{}^mP_{l+1,m+1}-D_l{}^mP_{l-1,m+1} \quad (46)$$

which follows by combining Eqs. (19) and (20), and the relation

$$e^{-2i\phi}[E_l{}^{m+2}P_{l-1,m+1}-F_l{}^{m+2}P_{l+1,m+1}]=E_l{}^mP_{l+1,m-1}-F_l{}^mP_{l-1,m-1} \quad (47)$$

which follows directly from Eq. (46) by noting that $C_l{}^m=F_l{}^{m+2}$ and $D_l{}^m=E_l{}^{m+2}$. Using the relation

$$\frac{A_l{}^{m+1}G_l{}^m}{2l+1}P_{l+1,m+1}+\left(\frac{B_l{}^{m+1}G_l{}^m}{2l+1}-D_{l-1}{}^{m-1}\right)P_{l-1,m+1}=\frac{(l-m)}{2l+1}e^{2i\phi}[F_l{}^mP_{l-1,m-1}-E_l{}^mP_{l+1,m-1}] \quad (48)$$

which follows from Eqs. (46) and from

$$\frac{B_l{}^{m+1}G_l{}^m}{2l+1}-D_{l-1}{}^{m-1}=\frac{(m-l)}{2l+1}D_l{}^m, \quad (49)$$

$$A_l{}^{m+1}G_l{}^m=(l-m)C_l{}^m, \quad (50)$$

and the relation

$$\frac{A_l{}^{m-1}H_l{}^m}{2l+1}P_{l+1,m-1}+\left(\frac{B_l{}^{m-1}H_l{}^m}{2l+1}-F_{l-1}{}^{m+1}\right)P_{l-1,m-1}=\frac{(l+m)}{2l+1}e^{-2i\phi}[E_l{}^{m+2}P_{l-1,m+1}-F_l{}^{m+2}P_{l+1,m+1}], \quad (51)$$

which follows from Eq. (47) and from

$$A_l{}^{m-1}H_l{}^m=(l+m)E_l{}^m, \quad (52)$$

$$\frac{-B_l{}^{m-1}H_l{}^m}{2l+1}+F_{l-1}{}^{m+1}=\frac{(l+m)}{2l+1}F_l{}^m. \quad (53)$$

Equation (45) can be rewritten in a form which involves the desired linear combination of spherical harmonics

$$\mathbf{E}\cdot\nabla_v f=(E_x-iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m+1)(l+m+2)]^{\frac{1}{2}}}{2(2l+1)}\left[\frac{lf^{l,m}}{v}-\frac{\partial f^{l,m}}{\partial v}\right]P_{l+1,m+1}+E_z\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m+1)(l-m+1)]^{\frac{1}{2}}}{2l+1}$$

$$\times\left[\frac{\partial f^{l,m}}{\partial v}-\frac{lf^{l,m}}{v}\right]P_{l+1,m}+(E_x+iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l-m+1)(l-m+2)]^{\frac{1}{2}}}{2(2l+1)}\left[\frac{\partial f^{l,m}}{\partial v}-\frac{lf^{l,m}}{v}\right]P_{l+1,m-1}$$

$$+(E_x-iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l-m)(l-m-1)]^{\frac{1}{2}}}{2(2l+1)}\left[\frac{\partial f^{l,m}}{\partial v}+l\frac{f^{l,m}}{v}\right]P_{l-1,m+1}+E_z\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m)(l-m)]^{\frac{1}{2}}}{2l+1}$$

$$\times\left[\frac{\partial f^{l,m}}{\partial v}+(l+1)\frac{f^{l,m}}{v}\right]P_{l-1,m}-(E_x+iE_y)\sum_{l=0}^{\infty}\sum_{m=-l}^{m=l}\frac{[(l+m)(l+m-1)]^{\frac{1}{2}}}{2(2l+1)}\left[\frac{\partial f^{l,m}}{\partial v}+(l+1)\frac{f^{l,m}}{v}\right]P_{l-1,m-1}. \quad (54)$$

5. MAGNETIC FIELD TERMS

Noting that

$$\nabla_v f = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} \left[\mathbf{i}_v P_{l,m} \frac{\partial f^{l,m}}{\partial v} + \mathbf{i}_\theta \frac{f^{l,m}}{v} \frac{\partial}{\partial \theta} P_{l,m} + \mathbf{i}_\phi \frac{f^{l,m}}{v \sin\theta} \frac{\partial}{\partial \phi} P_{l,m} \right], \tag{55}$$

where $\mathbf{i}_v$, $\mathbf{i}_\theta$, $\mathbf{i}_\phi$ are unit vectors in the direction of increasing v, θ, ϕ, one finds

$$(\mathbf{v}\times\mathbf{B})\cdot\nabla_v f = -B_\phi \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} f^{l,m} \frac{\partial}{\partial\theta} P_{l,m} + \frac{B_\theta}{\sin\theta} \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} f^{l,m} \frac{\partial}{\partial\phi} P_{l,m}$$

$$= [B_x \sin\phi - B_y \cos\phi] \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} f^{l,m} \frac{\partial}{\partial\theta} P_{l,m} + [B_x \cos\phi \cot\theta + B_y \sin\phi \cot\theta - B_z] \sum_{l=0}^{\infty} \sum_{m=-l}^{l} im f^{l,m} P_{l,m}, \tag{56}$$

where use has been made of Eq. (35) and B_θ and B_ϕ have been expressed in terms of B_x, B_y, B_z. By combining Eqs. (38) and (39), one obtains

$$-\cos\phi \frac{\partial}{\partial\theta} P_{l,m} + im \sin\phi \cot\theta P_{l,m} = -\tfrac{1}{2} G_l{}^m P_{l,m+1} + \tfrac{1}{2} H_l{}^m P_{l,m-1}. \tag{57}$$

Substituting Eqs. (38), (39), and (57) into Eq. (56), the desired expression for the magnetic force term results:

$$(\mathbf{v}\times\mathbf{B})\cdot\nabla_v f = -\frac{i}{2}(B_x - iB_y) \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} [(l-m)(l+m+1)]^{\frac{1}{2}} f^{l,m} P_{l,m+1}$$

$$-\frac{i}{2}(B_x + iB_y) \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} [(l+m)(l-m+1)]^{\frac{1}{2}} f^{l,m} P_{l,m-1} - iB_z \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} m f^{l,m} P_{l,m}. \tag{58}$$

6. SCATTERING INTEGRALS

Comparing Eqs. (13) and (15) it is seen that

$$\sigma(v,\chi) = \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \sigma_{l'}(v) P_{l',m'}(\boldsymbol{\Omega}) P_{l',m'}{}^*(\boldsymbol{\Omega}'), \tag{59}$$

and hence the scattering integrals of Eq. (7) can be written

$$Nv \int \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f^{l,m}(v') P_{l'm}(\boldsymbol{\Omega}') \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} \sigma_{l'}(v) P_{l'm'}(\boldsymbol{\Omega}) P_{l',m'}{}^*(\boldsymbol{\Omega}') d^2\Omega' - 4\pi N v \sigma_0 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f^{l,m} P_{l,m}, \tag{60}$$

where σ_0 is defined by Eq. (16) when one notes that $P_0 \equiv 1$. Using the orthogonality relation, Eq. (12), this expression reduces to the desired expression

$$4\pi N v \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{\sigma_l}{2l+1} f^{l,m} P_{l,m} - 4\pi N v \sigma_0 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f^{l,m} P_{l,m}, \tag{61}$$

where it is noted that $f_{l,m}(v') = f_{l,m}(v)$ since $v = v'$ in an elastic collision with an ion of very large mass.

7. THE COMPONENT EQUATIONS

The general equation relating the components $f^{l,m}(\mathbf{x},v,t)$ in the spherical harmonics expansion of $f(\mathbf{x},\mathbf{v},t)$ can now be calculated by substituting into Eq. (7) the terms as given by Eqs. (29), (54), (58), and (61), multiplying by

$P_{l,m}^*$, integrating over $d^2\Omega$, and making use of the orthogonality relation, Eq. (12),

$$\frac{\partial f^{l,m}}{\partial t}=-\frac{[(l+m-1)(l+m)]^{\frac{1}{2}}}{2(2l-1)}\left[\frac{e}{m_e}(E_x-iE_y)\left(\frac{[l-1]f^{l-1,m-1}}{v}-\frac{\partial f^{l-1,m-1}}{\partial v}\right)-v\left(\frac{\partial f^{l-1,m-1}}{\partial x}-i\frac{\partial f^{l-1,m-1}}{\partial y}\right)\right]$$

$$-\frac{[(l+m)(l-m)]^{\frac{1}{2}}}{2l-1}\left[\frac{e}{m_e}E_z\left(\frac{\partial f^{l-1,m}}{\partial v}-\frac{l-1}{v}f^{l-1,m}\right)+v\frac{\partial f^{l-1,m}}{\partial z}\right]+\frac{[(l-m)(l-m-1)]^{\frac{1}{2}}}{2(2l-1)}$$

$$\times\left[\frac{e}{m_e}(E_x+iE_y)\left(\frac{l-1}{v}f^{l-1,m+1}-\frac{\partial f^{l-1,m+1}}{\partial v}\right)-v\left(\frac{\partial f^{l-1,m+1}}{\partial x}+i\frac{\partial f^{l-1,m+1}}{\partial y}\right)\right]+\frac{ieB_z}{m_ec}f^{l,m}+\frac{ie}{2m_ec}(B_x+iB_y)f^{l,m+1}$$

$$+\frac{ie}{2m_ec}[(l-m+1)(l+m)]^{\frac{1}{2}}(B_x-iB_y)f^{l,m-1}-\frac{[(l-m+2)(l-m+1)]^{\frac{1}{2}}}{2(2l+3)}\left[\frac{e}{m_e}(E_x-iE_y)\right.$$

$$\times\left(\frac{\partial f^{l+1,m+1}}{\partial v}+(l+1)\frac{f^{l+1,m-1}}{v}\right)+v\left(\frac{\partial f^{l+1,m-1}}{\partial x}-i\frac{\partial f^{l+1,m-1}}{\partial y}\right)\Big]-\frac{[(l+m+1)(l-m+1)]^{\frac{1}{2}}}{2l+3}$$

$$\times\left[\frac{e}{m_e}E_z\left(\frac{\partial f^{l+1,m}}{\partial v}+(l+2)\frac{f^{l+1,m}}{v}\right)+v\frac{\partial f^{l+1,m}}{\partial z}\right]+\frac{[(l+m+2)(l+m+1)]^{\frac{1}{2}}}{2(2l+3)}\left[\frac{e}{m_e}(E_x+iE_y)\right.$$

$$\left.\times\left(\frac{\partial f^{l+1,m+1}}{\partial v}+(l+2)\frac{f^{l+1,m+1}}{v}\right)+v\left(\frac{\partial f^{l+1,m+1}}{\partial x}+i\frac{\partial f^{l+1,m+1}}{\partial y}\right)\right]+\frac{4\pi N}{2l+1}v\sigma_1 f^{l,m}-4\pi Nv\sigma_0 f^{l,m},\quad(62)$$

where it is noted that $f^{l,m}$ vanishes for all negative l and whenever m does not lie between l and $-l$.

The first few component equations are listed as follows. For $l=0$ and $m=0$,

$$\frac{\partial f^{0,0}}{\partial t}=\frac{ie}{m_ec}B_zf^{0,0}-\frac{\sqrt{2}}{6}\frac{e}{m_e}(E_x-iE_y)\left(\frac{\partial f^{1,-1}}{\partial v}+\frac{f^{1,-1}}{v}\right)-\frac{\sqrt{2}}{6}v\left(\frac{\partial f^{1,-1}}{\partial x}-i\frac{\partial f^{1,-1}}{\partial y}\right)-\frac{e}{3m_e}E_z\left(\frac{\partial f^{1,0}}{\partial v}+\frac{2f^{1,0}}{v}\right)$$

$$-\frac{1}{3}v\frac{\partial f^{1,0}}{\partial z}+\frac{\sqrt{2}}{6}\frac{e}{m_e}(E_x+iE_y)\left(\frac{\partial f^{1,1}}{\partial v}+\frac{2}{v}f^{1,1}\right)+\frac{\sqrt{2}}{6}v\left(\frac{\partial f^{1,1}}{\partial x}+i\frac{\partial f^{1,1}}{\partial y}\right)+4\pi Nv\sigma_0f^{0,0}-4\pi Nv\sigma_0f^{0,0}.\quad(63)$$

For $l=1$ and $m=0$,

$$\frac{\partial f^{1,0}}{\partial t}=-\frac{e}{m_e}E_z\frac{\partial f^{0,0}}{\partial v}-v\frac{\partial f^{0,0}}{\partial z}+\frac{ieB_z}{m_ec}f^{1,0}+\frac{ie}{2m_ec}(B_x+iB_y)f^{1,1}+\frac{ie}{\sqrt{2}m_ec}(B_x-iB_y)f^{1,-1}-\frac{\sqrt{6}}{10}\left[\frac{e}{m_e}(E_x-iE_y)\right.$$

$$\times\left(\frac{\partial f^{2,-1}}{\partial v}+2\frac{f^{2,-1}}{v}\right)+v\left(\frac{\partial f^{2,-1}}{\partial x}-i\frac{\partial f^{2,-1}}{\partial y}\right)\Big]-\frac{2}{5}\left[\frac{e}{m_e}E_z\left(\frac{\partial f^{2,0}}{\partial v}+3\frac{f^{2,0}}{v}\right)+v\frac{\partial f^{2,0}}{\partial z}\right]$$

$$+\frac{\sqrt{6}}{10}\left[\frac{e}{m_e}(E_x+iE_y)\left(\frac{\partial f^{2,1}}{\partial v}+3\frac{f^{2,1}}{v}\right)+v\left(\frac{\partial f^{2,1}}{\partial x}+i\frac{\partial f^{2,1}}{\partial y}\right)\right]+\frac{4\pi}{3}Nv\sigma_1f^{1,0}-4\pi Nv\sigma_0f^{1,0}.\quad(64)$$

For $l=1$ and $m=1$,

$$\frac{\partial f^{1,1}}{\partial t}=\frac{1}{\sqrt{2}}\frac{e}{m_e}(E_x-iE_y)\frac{\partial f^{0,0}}{\partial v}+\frac{1}{\sqrt{2}}v\left(\frac{\partial f^{0,0}}{\partial x}-i\frac{\partial f^{0,0}}{\partial y}\right)+\frac{ieB_z}{m_ec}f^{1,1}+\frac{ie}{\sqrt{2}m_ec}(B_x-iB_y)f^{1,0}-\frac{\sqrt{2}}{10}\frac{e}{m_e}(E_x-iE_y)$$

$$\times\left(\frac{\partial f^{2,0}}{\partial v}+\frac{\partial f^{2,0}}{v}\right)-\frac{\sqrt{2}}{10}v\left(\frac{\partial f^{2,0}}{\partial x}-i\frac{\partial f^{2,0}}{\partial y}\right)-\frac{\sqrt{3}}{5}\frac{e}{m_e}E_z\left(\frac{\partial f^{2,1}}{\partial v}+\frac{3f^{2,1}}{v}\right)-\frac{\sqrt{3}}{5}v\frac{\partial f^{2,1}}{\partial z}+\frac{\sqrt{3}}{5}\frac{e}{m_e}(E_x+iE_y)$$

$$\times\left(\frac{\partial f^{2,2}}{\partial v}+\frac{3f^{2,2}}{v}\right)+\frac{\sqrt{3}}{5}v\left(\frac{\partial f^{2,2}}{\partial x}+i\frac{\partial f^{2,2}}{\partial y}\right)+\frac{4\pi}{3}Nv\sigma_1f^{1,1}-4\pi Nv\sigma_0f^{1,1}.\quad(65)$$

Note that, in view of Eq. (11) and the fact that f must be real,

$$f^{l,-m}=(-)^m f^{l,m*}, \tag{66}$$

so that the solution of Eq. (65) for $f^{1,1}$ implies that one also has $f^{1,-1}$.

In general, almost complete cross coupling between different $f^{l,m}$ is present. However, the z component of the magnetic induction $\mathbf{B}$ produces no cross coupling between the different components of $f^{l,m}$. The presence of the electric and magnetic fields greatly complicates the equations as compared with the field free case. However, for the special case of single component fields consisting only of B_z or E_z, these equations are only slightly more complex and the same method of solution as for the field free equations or for the case of symmetry about a polar axis[1] should be applicable. For a reasonably isotropic case with arbitrary fields, one would assume that all the $f^{l,m}$ except $f^{0,0}$, $f^{1,-1}$, $f^{1,0}$, $f^{1,1}$ vanish and simultaneously solve Eqs. (63), (64), and (65) with all the other $f^{l,m}$ which appear in these equations set equal to zero. The resulting equations, corresponding to a $P_{1,1}$ approximation, are considerably simplified and should yield to the same types of analytic and numerical techniques that have been developed for spherical harmonics calculations of neutron physics problems.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

Construction of Symmetry-Adapted Functions in the Many-Particle Problem

R. K. NESBET*
Department of Physics, Boston University, Boston, Massachusetts
(Received March 7, 1961)

A new method is presented for obtaining many-particle angular momentum eigenfunctions and matrix elements of an invariant Hamiltonian. The same technique can be used to construct symmetry-adapted functions for any group of operators that commute with the Hamiltonian, and to simplify the evaluation of matrix elements in the symmetry-adapted basis. Applied to an arbitrary configuration, the method produces orthonormal functions identical with those that would be obtained by Schmidt orthogonalization of the projections of the original basis functions of the configuration. Because of this relationship, matrix elements of the Hamiltonian are greatly simplified, but the functions are obtained without explicitly constructing the projection operators or their matrix representations. To illustrate the method, it is applied to the spin coupling of configurations with three, four, and five particles outside closed shells, and to the two ^2D functions of the atomic configuration d^3, in Russell-Saunders coupling. Tables of the coefficients needed to evaluate all independent matrix elements are obtained for these examples, and typical matrix elements are calculated.

I. INTRODUCTION

THE purpose of this paper is to describe a general technique for dealing with one of the central practical problems in many-particle quantum mechanics, the evaluation of matrix elements of a Hamiltonian with nontrivial symmetry properties, in a basis of functions constructed so as to take as much advantage as possible of this symmetry in simplifying the calculations. The examples given in this paper are all concerned with systems of identical Fermions with a spherically symmetrical Hamiltonian. In such a case, the basis functions are angular momentum eigenfunctions constructed as linear combinations of Slater determinants. Even with this restriction, there are several different situations which have led historically to the development of special methods: coupling of several elementary spins (the vector model, valence bond theory), coupling of orbital angular momenta as well as spin (Russell-Saunders coupling), coupling of total angular momenta of individual particles ($j-j$ coupling), inclusion of isotopic spin (Russell-Saunders coupling or $j-j$ coupling in nuclei). Specific examples are worked out here only for the first two cases, but the same formalism can be used equally well for the others.

In the general situation to which the present method is applicable, there is some group of linear operators that commute with a given Hamiltonian. The irreducible matrix representations of this group define the transformation properties of standard symmetry-adapted functions, such that the matrix elements of the group operators in the basis of these functions are just the irreducible representation matrices of the group.[1] The transformation properties of a symmetry-adapted function for a given group are specified by two

* National Cancer Institute Special Research Fellow at the Institut Pasteur, Paris, France, 1960–1961.

[1] An excellent introduction to group representation theory, from a physicist's point of view, is given in L. D. Landau and E. M. Lifschitz, *Quantum Mechanics, Non-Relativistic Theory* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1958), Chap. XII. For the theory of angular momentum with many applications, see M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957); and A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957).

Note that, in view of Eq. (11) and the fact that f must be real,

$$f^{l,-m}=(-)^m f^{l,m*}, \tag{66}$$

so that the solution of Eq. (65) for $f^{1,1}$ implies that one also has $f^{1,-1}$.

In general, almost complete cross coupling between different $f^{l,m}$ is present. However, the z component of the magnetic induction $\mathbf{B}$ produces no cross coupling between the different components of $f^{l,m}$. The presence of the electric and magnetic fields greatly complicates the equations as compared with the field free case. However, for the special case of single component fields consisting only of B_z or E_z, these equations are only slightly more complex and the same method of solution as for the field free equations or for the case of symmetry about a polar axis[1] should be applicable. For a reasonably isotropic case with arbitrary fields, one would assume that all the $f^{l,m}$ except $f^{0,0}$, $f^{1,-1}$, $f^{1,0}$, $f^{1,1}$ vanish and simultaneously solve Eqs. (63), (64), and (65) with all the other $f^{l,m}$ which appear in these equations set equal to zero. The resulting equations, corresponding to a $P_{1,1}$ approximation, are considerably simplified and should yield to the same types of analytic and numerical techniques that have been developed for spherical harmonics calculations of neutron physics problems.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

Construction of Symmetry-Adapted Functions in the Many-Particle Problem

R. K. NESBET*
Department of Physics, Boston University, Boston, Massachusetts
(Received March 7, 1961)

A new method is presented for obtaining many-particle angular momentum eigenfunctions and matrix elements of an invariant Hamiltonian. The same technique can be used to construct symmetry-adapted functions for any group of operators that commute with the Hamiltonian, and to simplify the evaluation of matrix elements in the symmetry-adapted basis. Applied to an arbitrary configuration, the method produces orthonormal functions identical with those that would be obtained by Schmidt orthogonalization of the projections of the original basis functions of the configuration. Because of this relationship, matrix elements of the Hamiltonian are greatly simplified, but the functions are obtained without explicitly constructing the projection operators or their matrix representations. To illustrate the method, it is applied to the spin coupling of configurations with three, four, and five particles outside closed shells, and to the two ^2D functions of the atomic configuration d^3, in Russell-Saunders coupling. Tables of the coefficients needed to evaluate all independent matrix elements are obtained for these examples, and typical matrix elements are calculated.

I. INTRODUCTION

THE purpose of this paper is to describe a general technique for dealing with one of the central practical problems in many-particle quantum mechanics, the evaluation of matrix elements of a Hamiltonian with nontrivial symmetry properties, in a basis of functions constructed so as to take as much advantage as possible of this symmetry in simplifying the calculations. The examples given in this paper are all concerned with systems of identical Fermions with a spherically symmetrical Hamiltonian. In such a case, the basis functions are angular momentum eigenfunctions constructed as linear combinations of Slater determinants. Even with this restriction, there are several different situations which have led historically to the development of special methods: coupling of several elementary spins (the vector model, valence bond theory), coupling of orbital angular momenta as well as spin (Russell-Saunders coupling), coupling of total angular momenta of individual particles ($j-j$ coupling), inclusion of isotopic spin (Russell-Saunders coupling or $j-j$ coupling in nuclei). Specific examples are worked out here only for the first two cases, but the same formalism can be used equally well for the others.

In the general situation to which the present method is applicable, there is some group of linear operators that commute with a given Hamiltonian. The irreducible matrix representations of this group define the transformation properties of standard symmetry-adapted functions, such that the matrix elements of the group operators in the basis of these functions are just the irreducible representation matrices of the group.[1] The transformation properties of a symmetry-adapted function for a given group are specified by two

* National Cancer Institute Special Research Fellow at the Institut Pasteur, Paris, France, 1960–1961.

[1] An excellent introduction to group representation theory, from a physicist's point of view, is given in L. D. Landau and E. M. Lifschitz, *Quantum Mechanics, Non-Relativistic Theory* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1958), Chap. XII. For the theory of angular momentum with many applications, see M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957); and A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957).

indices, Λ to denote the irreducible representation, and a subspecies index M to denote a particular row of the representation matrix.

The proper generalization of a *configuration*, in the sense of atomic spectroscopy or nuclear theory, is the vector space spanned by an orthonormal set of functions that transform into linear combinations of each other under operations in the symmetry group. If symmetry-adapted functions are constructed by taking appropriate linear combinations of the basis functions for a configuration, matrix elements of the Hamiltonian are greatly simplified, since they vanish between symmetry-adapted functions that differ in either of the two indices Λ or M, and are independent of the subspecies index M. The original basis functions could equally well be expressed as linear combinations of the set of all linearly independent symmetry-adapted functions from the same configuration, since these are simply alternative choices of the basis of a vector space. If expanded this way, each original basis function appears as a sum of functions belonging to the different symmetry species (ΛM), each of which is the *projection* of the original basis function onto the corresponding symmetry species. The projection operator ${}^{M}_{\Lambda}\mathcal{O}$, which picks out the (ΛM) projection of an arbitrary function, can be written explicitly as a sum over all the operators in the symmetry group, weighted by certain elements of the irreducible representation matrices.[2] Because of the properties of the projection operator, matrix elements of the Hamiltonian between symmetry-adapted functions obtained as projections are particularly simple. Such matrix elements are reduced from quadratic to linear forms in the coefficients that define the symmetry-adapted functions as linear combinations of the original basis.[3]

In the case of Russell-Saunders coupling for atoms, the Hamiltonian is invariant under permutation of the particle coordinates, as well as under rotations of the frame of reference independently for space and spin coordinates. The standard projection operator of group theory for the totally antisymmetrical representation of the permutation group is just the operator that converts a product of orbitals into a Slater determinant, and as a result of the general rules for matrix elements between projected functions, matrix elements between normalized Slater determinants differ from those for simple products of orbitals only by the occurrence of exchange integrals for the two-particle operator (the Coulomb potential). Hence it is convenient to use Slater determinants as basis functions in constructing symmetry-adapted functions for the space and spin rotation groups. The standard projection operator is not at all convenient to use for the rotation group, since it requires summation over the group, hence integration over the parameters (e.g., Euler angles) that specify a continuous group; but fortunately, the matrix of the projection operator can be constructed quite easily by a method due to Löwdin[3] that uses the properties of the angular momentum operators, which are the infinitesimal generators of the rotation group.

In an earlier paper,[4] it was shown that it is possible to obtain the symmetry-adapted projections of the basis functions of a given configuration without explicitly constructing either the projection operator or its matrix elements. Whenever several symmetry-adapted functions of the same symmetry species occur in a given configuration, this method requires, as in the method of Löwdin, a final transformation by Schmidt orthogonalization to produce an orthonormal basis of symmetry-adapted functions from the projections of the original basis functions, since the projections are nonorthogonal and usually not linearly independent.

It will be shown in the present paper that by carrying out the calculations required by the earlier method[4] in a certain standard sequence, the Schmidt orthogonalized projected functions can be obtained directly, thus eliminating not only the special Schmidt orthogonalization, but also the construction of the individual projected functions, which are no longer needed. The over-all process now seems to be somewhat more efficient than the method of Löwdin, whenever the latter requires one or more matrix multiplications in the course of construction of the matrix of the projection operator. Either of these methods leads to simpler formulas for matrix elements of the Hamiltonian than do earlier methods that do not use the properties of the projection operator. As applied to evaluation of the matrix elements for the two independent doublets in the configuration *abc* (e.g., three electrons outside of closed shells), worked out in Sec. VI, the present method appears to be somewhat simpler than that of Trainor,[5] who has considerably simplified the older method of Yamanouchi[6] that has been extensively used for nuclear calculations.[7] Trainor obtains one of the two doublet functions for this configuration in the paper cited—it should be emphasized that the present method obtains all independent functions for each symmetry species as an orthonormal set, in a form from which matrix elements can be evaluated directly.

It is difficult to compare the present method with that of Boys,[8] which also leads to a complete basis of

[2] E. Wigner, *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren* (Friedrich Vieweg and Sohn, Braunschweig, Germany, 1931), pp. 120–133.

[3] P.-O. Löwdin, Phys. Rev. **97**, 1509 (1955); Colloq. intern. centre natl. recherche sci. (Paris) **82**, 23 (1958); Advances in Chem. Phys. **2**, pp. 285–294 (1959); Technical Notes, Quantum Chemistry Group, Uppsala, No. 2, 1957; Nos. 12 and 14, 1958; No. 27, 1959 (unpublished).

[4] R. K. Nesbet, Ann. Phys. **3**, 397 (1958).

[5] L. E. H. Trainor, Can. J. Phys. **35**, 555 (1957).

[6] T. Yamanouchi, Proc. Phys.-Math. Soc. Japan (3), **19**, 436 (1937).

[7] H. A. Jahn, Proc. Roy. Soc. (London) **A201**, 516 (1950); **A205**, 192 (1951); H. A. Jahn and H. Van Wieringen, *ibid* **A209**, 502 (1951); B. H. Flowers, *ibid* **A212**, 248 (1952); **A215**, 398 (1952); A. R. Edmonds and B. H. Flowers, *ibid* **A214**, 515 (1952).

[8] S. F. Boys, Proc. Roy. Soc. (London) **A201**, 125 (1950); **A217**, 136, 235 (1953); M. J. M. Bernal and S. F. Boys, Trans. Roy. Soc. (London) **A245**, 139 (1952); S. F. Boys and V. E. Price, Trans. Roy. Soc. (London) **A246**, 57 (1954).

orthonormal symmetry-adapted functions for the rotation group (angular momentum eigenfunctions), and to formulas for matrix elements for arbitrary configurations. The present method is certainly more efficient in deriving individual matrix elements from first principles, but the method of Boys provides for a systematic tabulation of coefficients that takes advantage of coefficients already calculated for N-particle configurations in evaluating those needed for $(N+1)$-particle configurations. The same remark applies to the methods developed by Racah,[9] and, in general, to the procedure of building up configurations by adding one or two particles at a time using fractional parentage coefficients.

For a finite group it is possible to construct the projection operators explicitly, and this is perhaps the most efficient procedure for projection operators that can be factorized.[10] In any case, the Schmidt orthogonalized projected functions are probably the most convenient form for symmetry-adapted functions, and tables of coefficients and formulas for matrix elements can be used in the same form as given here.

As examples of the present method, all of the nontrivial eigenfunctions of total spin with $M_S=S$ have been obtained for configurations with three, four, and five particles outside of closed shells, and some typical matrix elements have been worked out for these configurations. The two ^2D functions from the configuration d^3 and the corresponding two by two matrix of the electronic Hamiltonian have been calculated, as an example of the case of Russell-Saunders coupling. Results are given in Sec. VI.

II. NOTATION AND DEFINITIONS

It is necessary to distinguish between the Slater determinants of the original basis (orthonormal when the orbitals are taken to be orthonormal) and various functions derived from them. In general a normalized Slater determinant will be denoted by Φ_μ, the (ΛM) projection of this determinant by ${}^{M}{}_{\Lambda}\Theta_\mu$, and the normalized function proportional to this projection by ${}^{M}{}_{\Lambda}\Psi_\mu$. The normalized function obtained by Schmidt orthogonalization of ${}^{M}{}_{\Lambda}\Theta_\mu$ to the projections of other determinants in the same configuration will be denoted by ${}^{M}{}_{\Lambda}\Psi_\mu'$. Obviously the definition of these functions depends upon the order in which Schmidt orthogonalization is carried out. The normalized function ${}^{M}{}_{\Lambda}\Psi_\mu$ has been called a *projected determinant* in an earlier paper.[4] The indices Λ and M will be omitted whenever it is unnecessary to distinguish between different symmetry species.

Let $\mathcal{O}$ or ${}^{M}{}_{\Lambda}\mathcal{O}$ be the projection operator that picks out the (ΛM) component of an arbitrary function. If this operator is applied to Φ_μ, it will produce the projection Θ_μ. But it is clear that Θ_μ can contain components only from determinants belonging to the same configuration as Φ_μ. If the determinants of this configuration are enumerated as $\Phi_1\cdots\Phi_n$, where $\mu\leqslant n$, then

$$\Theta_\mu=\mathcal{O}\Phi_\mu=l_\mu^{-1}\sum_{i=1}^{n} y_{\mu i}\Phi_i, \tag{1}$$

where l_μ is chosen so that $y_{\mu\mu}=1$. Then it can be shown[4] that the normalized function Ψ_μ (projected determinant) is

$$\Psi_\mu=l_\mu^{\frac{1}{2}}\Theta_\mu, \tag{2}$$

and that

$$l_\mu=\sum_{i=1}^{n}|y_{\mu i}|^2. \tag{3}$$

Since all matrix elements of any Hermitian operator that commutes with the projection operator can be evaluated without computing l_μ,[4] it is sufficient to tabulate the coefficients $y_{\mu i}$, and to denote the projected determinant by

$$\Psi_\mu\sim\sum_{i=1}^{n} y_{\mu i}\Phi_i. \tag{4}$$

The fact that it is possible to evaluate matrix elements of the Hamiltonian without evaluating all of the coefficients $y_{\mu i}$ has an important application in the theory of the Heisenberg exchange operator.[11]

If a linearly independent subset of the projected determinants Ψ_μ from a single configuration are orthogonalized by the Schmidt process, this will give orthonormal functions

$$\Psi_\mu'=k_\mu^{-\frac{1}{2}}\sum_{i=1}^{n} x_{\mu i}\Phi_i, \tag{5}$$

where k_μ is chosen so that $x_{\mu\mu}=1$. Then

$$k_\mu=\sum_{i=1}^{n}|x_{\mu i}|^2, \tag{6}$$

and, as before, it will be shown that matrix elements of Hermitian operators that commute with the projection operator can be evaluated without evaluating k_μ. The orthogonalized functions Ψ_μ' are a complete orthonormal basis for symmetry-adapted functions of symmetry species (ΛM) from the given configuration. It is sufficient to denote these functions by

$$\Psi_\mu'\sim\sum_{i=1}^{n} x_{\mu i}\Phi_i, \tag{7}$$

where there is one Ψ_μ' for each linearly independent projection Θ_μ.

Suppose that there are $m\leqslant n$ linearly independent projections Θ_j in the given configuration. Then the orthonormal basis functions Ψ_μ' can also be expressed as linear combinations of these projections,

$$\Psi_\mu'\sim\sum_{j=1}^{m} a_{\mu j}\Theta_j, \tag{8}$$

where $a_{\mu\mu}=1$. It will be shown that the constant of proportionality is just $k_\mu^{\frac{1}{2}}$, for k_μ given by Eq. (6), so

$$\Psi_\mu'=k_\mu^{\frac{1}{2}}\sum_{j=1}^{m} a_{\mu j}\Theta_j. \tag{9}$$

[9] G. Racah, Phys. Rev. **61**, 186 (1942); **62**, 438 (1942); **63**, 367 (1943); **76**, 1352 (1949).

[10] M. A. Melvin, Revs. Modern Phys. **28**, 18 (1956); H. V. McIntosh, Technical Notes, Quantum Chemistry Group, Uppsala Nos. 19 and 20, 1958 (unpublished); J. Math. Phys. **1**, 453 (1960).

[11] R. K. Nesbet, Ann. Phys. **4**, 87 (1958); Phys. Rev. **119**, 658 (1960).

It will be shown that the coefficients a of Eq. (9) can be obtained by inspection, given the coefficients x of Eq. (5), and that both sets of coefficients can be arranged in a single rectangular table, from which matrix elements of any operator that commutes with the projection operator $\mathcal{O}$ can be evaluated for functions Ψ_μ' either from the same configuration or from different configurations.

In deriving formulas for the matrix elements, it is convenient to assume that the first m projected determinants from the given configuration are linearly independent; this can be arranged by suitably reordering the basis determinants. Then the summation of Eq. (9) is over the first m indices of the summation in Eq. (5). It will be shown later that this reordering is not really necessary, since redundant coefficients $a_{\mu j}$ of Eq. (9) can be set equal to zero.

III. EVALUATION OF MATRIX ELEMENTS

Formulas will be derived for matrix elements of any Hermitian operator H that is invariant under the group of transformations whose irreducible representations are denoted by (ΛM), with projection operators ${}^{\mathrm{M}}{}_{\Lambda}\mathcal{O}$. The operators H and ${}^{\mathrm{M}}{}_{\Lambda}\mathcal{O}$ commute.

The fundamental formula, the so-called "turnover rule,"[3] is

$$(\Theta_i, H\Theta_j) = (\Phi_i, H\Theta_j) = (\Theta_i, H\Phi_j), \tag{10}$$

where $\Theta_i = {}^{\mathrm{M}}{}_{\Lambda}\mathcal{O}\,\Phi_i$ and $\Theta_j = {}^{\mathrm{M}}{}_{\Lambda}\mathcal{O}\Phi_j$. Matrix elements between symmetry-adapted functions vanish unless both functions belong to the same symmetry species. This will be assumed and indices Λ, M will be omitted.

Consider the orthogonalized projected determinants ${}_{\mathrm{I}}\Psi_\mu'$ and ${}_{\mathrm{II}}\Psi_\nu'$, expressed as linear combinations of determinants or their projections from configurations I and II, respectively, with coefficients as in Eqs. (5) and (9). Then

$$\begin{aligned}({}_{\mathrm{I}}\Psi_\mu', H\,{}_{\mathrm{II}}\Psi_\nu') &= (k_\mu/k_\nu)^{\frac{1}{2}} \sum_{j=1}^{m_{\mathrm{I}}} \sum_{i=1}^{n_{\mathrm{II}}} a^{\mathrm{I}}{}_{\mu j}{}^{*} x_{\nu i}{}^{\mathrm{II}} (\Theta_j, H\Phi_i) \\ &= (k_\mu/k_\nu)^{\frac{1}{2}} \sum_{j=1}^{m_{\mathrm{I}}} \sum_{i=1}^{n_{\mathrm{II}}} a^{\mathrm{I}}{}_{\mu j}{}^{*} x_{\nu i}{}^{\mathrm{II}} (\Phi_j, H\Phi_i),\end{aligned} \tag{11}$$

using Eq. (10). Similarly,

$$\begin{aligned}({}_{\mathrm{I}}\Psi_\mu', H\,{}_{\mathrm{II}}\Psi_\nu') &= (k_\nu/k_\mu)^{\frac{1}{2}} \sum_{i=1}^{n_{\mathrm{I}}} \sum_{j=1}^{m_{\mathrm{II}}} x^{\mathrm{I}}{}_{\mu i}{}^{*} a_{\nu j}{}^{\mathrm{II}} (\Phi_i, H\Theta_j) \\ &= (k_\nu/k_\mu)^{\frac{1}{2}} \sum_{i=1}^{n_{\mathrm{I}}} \sum_{j=1}^{m_{\mathrm{II}}} x^{\mathrm{I}}{}_{\mu i}{}^{*} a_{\nu j}{}^{\mathrm{II}} (\Phi_i, H\Phi_j).\end{aligned} \tag{12}$$

The constants k_μ, k_ν can be eliminated by taking the geometrical mean of Eqs. (11) and (12), to give

$$\begin{aligned}({}_{\mathrm{I}}\Psi_\mu', H\,{}_{\mathrm{II}}\Psi_\nu') = \{&[\textstyle\sum_j{}^{\mathrm{I}} \sum_i{}^{\mathrm{II}} a^{\mathrm{I}}{}_{\mu j}{}^{*} x_{\nu i}{}^{\mathrm{II}} (\Phi_j, H\Phi_i)] \\ &\times [\textstyle\sum_i{}^{\mathrm{I}} \sum_j{}^{\mathrm{II}} x^{\mathrm{I}}{}_{\mu i}{}^{*} a_{\nu j}{}^{\mathrm{II}} (\Phi_i, H\Phi_j)]\}^{\frac{1}{2}}.\end{aligned} \tag{13}$$

Equation (13) is a generalization of the formula for projected determinants given in the earlier paper.[4] The relationship between the two terms in square brackets is simple, since they must be proportional to each other because of the identity between Eqs. (11) and (12), but nontrivial, as will be seen on working through detailed examples. It provides a very useful check of both the arithmetic and algebra in practical calculations. Since different pairs of determinants occur in Eqs. (11) and (12), this identity checks not only the coefficients a and x, but also the formulas by which the matrix elements between determinants are reduced to linear combinations of independent integrals over the one- and two-particle operators in the operator H. For example, if there were errors in the Gaunt coefficients tabulated by Condon and Shortley,[12] this could cause the identity to be violated, if applied to matrix elements for atomic wave functions.

In using Eq. (13), the procedure is to evaluate both terms in square brackets, reduce them to linear combinations of independent one- and two-particle integrals, verify that the coefficients of corresponding integrals are proportional with a constant factor, then write the final formula with each coefficient taken as the geometric mean of the two calculated coefficients, with due care to make a consistent choice of the sign of the square root {e.g., $[(-a)(-b)]^{\frac{1}{2}} = -(ab)^{\frac{1}{2}}$}. If the work were programmed for a digital computer, the check could be omitted, and Eq. (12) could be used directly with constants k_μ determined from Eq. (6).

IV. CONSTRUCTION OF AN ORTHONORMAL BASIS

Functions from different configurations are necessarily orthogonal, but the requirement of orthogonality imposes certain conditions on the coefficients a and x for functions Ψ_μ' from the same configuration. Since the identity operator is Hermitian and commutes with any other operator, it satisfies the conditions required for use of Eq. (12). Hence, by Eq. (12),

$$\begin{aligned}(\Psi_\mu', \Psi_\nu') &= (k_\nu/k_\mu)^{\frac{1}{2}} \textstyle\sum_i \sum_j x_{\mu i}{}^{*} a_{\nu j} (\Phi_i, \Phi_j) \\ &= (k_\nu/k_\mu)^{\frac{1}{2}} \textstyle\sum_i x_{\mu i}{}^{*} a_{\nu i}.\end{aligned} \tag{14}$$

Since the functions Ψ_μ' are to be orthonormal, the coefficients a must be chosen so that

$$\textstyle\sum_{i=1}^{m} x_{\mu i}{}^{*} a_{\nu i} = \delta_{\mu\nu}. \tag{15}$$

If the functions Ψ_μ' are obtained by Schmidt orthogonalization from the projections Θ_μ, the coefficients a of Eq. (9) will form a triangular matrix such that $a_{\mu j} = 0$ unless $j \leqslant \mu$. This follows from the Schmidt process: each successive Ψ_μ' is obtained by orthogonalizing Θ_μ to the set of projections Θ_j with $j < \mu$. Hence the matrix inverse to $a_{\mu j}$ is triangular, and by Eq. (15), the coefficients $x_{\mu i}$ must vanish unless $i \geqslant \mu$. Since the matrix of these coefficients is rectangular in general, the left-hand m by m square matrix will be triangular as a result of Schmidt orthogonalization.

Suppose that the coefficients x have been obtained in this trapezoidal form for an orthonormal set of functions Ψ_μ', a complete basis for linear expansion of functions

[12] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, New York, 1953).

TABLE I. Doublet functions from configuration *abc*. The first section of the table gives the matrix of the linear homogeneous equations (from matrix elements of the operator S^+) satisfied by the coefficients $x_{\mu i}$. The coefficients $x_{\mu i}$ are given in the second section of the table together with the auxiliary coefficients $a_{\mu i}$ and the normalization constants k_μ. In each row the $a_{\mu i}$ are listed from left to right up to and including the starred figure; the $x_{\mu i}$ are listed from left to right beginning with the starred figure, which is common to both sets of coefficients. The index μ for each row of coefficients is to be given the value of the index i for the column in which the asterisk occurs.

Equations $i=2$	3	4	
1	−1	1	
Coefficients			k_μ
$-\frac{1}{2}$	1*	1	2
1*	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$

of symmetry species (ΛM) from a given configuration. By suitably ordering the original determinants, their projections Θ_j have been chosen to be linearly independent for $j \leqslant m$, and are a complete but not orthogonal basis for the same functions. Either basis can be expanded in terms of the other. In particular, since Ψ_μ' can be expressed as a linear combination of projected functions, it follows from Eqs. (10) and (5) that

$$(\Psi_\mu', \Theta_i) = (\Psi_\mu', \Phi_i) = k_\mu^{-\frac{1}{2}} x_{\mu i}^*, \tag{16}$$

giving the expansion coefficients for Θ_i in the orthonormal basis of the functions Ψ_μ'. Then the coefficients $k_\mu^{\frac{1}{2}} a_{\mu j}$ of the inverse transformation, Eq. (9), are determined from Eq. (15), which can be written as

$$\sum_i (k_\mu^{-\frac{1}{2}} x_{\mu i}^*)(k_\nu^{\frac{1}{2}} a_{\nu i}) = \delta_{\mu\nu}. \tag{17}$$

When the coefficients x are in trapezoidal form, the coefficients a are necessarily in the triangular form characteristic of the Schmidt process. This fact has been used by Löwdin[3] to simplify the calculation of the orthonormalized projected functions Ψ_μ'. Given the nonorthogonal projections obtained by explicit use of the projection operator, it suffices to reduce the coefficients y of Eq. (1) to trapezoidal form by Gaussian elimination. If the original functions are projected functions, the resulting functions are orthogonal.

The present argument is the converse of that of Löwdin. Here it is shown that, if a set of mutually orthogonal functions is obtained that is a complete basis for functions of symmetry species (ΛM) from a given configuration, and if the coefficients are in trapezoidal form, then it follows that these basis functions are the same except for numerical factors as the set of orthonormalized projections. Thus, the coefficients x and a required for evaluation of matrix elements by Eqs. (12) or (13) can be obtained without making formal use of the projection operator. The trapezoidal form of the matrix of coefficients x makes it possible to solve Eq. (15) by inspection for the coefficients a, since this only requires inversion of a triangular matrix.

The form of the matrices x and a is indicated by Tables I through VI, which will be discussed in Sec. VI. Since $a_{\mu i}=0$ for $i>\mu$, $x_{\mu i}=0$ for $i<\mu$, and $a_{\mu\mu}=x_{\mu\mu}=1$, the two sets of coefficients can be entered in a single rectangular table for each configuration. The elements $a_{\mu\mu}=x_{\mu\mu}=1$ are indicated by an asterisk. Hence the coefficients $a_{\mu i}$ are read from left to right, up to and including the 1 marked by an asterisk, and the coefficients $x_{\mu i}$ begin with this 1 and continue to the right. The index μ for the coefficients in a given row of these tables is equal to the index i of the column marked by the asterisk in that row. With these conventions, the occurrence of a determinant whose projection is not linearly independent of the projections of those enumerated before it in the original basis, but which is followed by others whose projections are linearly independent, does not require any change in the

TABLE II. Triplet functions from configuration *abcd*. See remarks in caption to Table I.

Equations $i=2$	3	4	5	
1	−1	1	−1	
Coefficients				k_μ
$\frac{1}{2}$	$-\frac{1}{2}$	1*	1	2
$-\frac{1}{3}$	1*	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$
1*	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{4}{3}$

TABLE III. Singlet functions from configuration *abcd*. See remarks in caption to Table I.

Equations $i=6$	7	8	9	10	11	
1	−1	0	1	0	0	
−1	0	−1	0	1	0	
0	−1	1	0	0	1	
0	0	0	−1	1	−1	
Coefficients						k_μ
$-\frac{1}{2}$	1*	1	1	1	0	4
1*	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	1	3

TABLE IV. Quartet functions from configuration *abcde*. See remarks in caption to Table I.

Equations $i=2$	3	4	5	6	
1	−1	1	−1	1	
Coefficients					k_μ
$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	1*	1	2
$\frac{1}{3}$	$-\frac{1}{3}$	1*	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$
$-\frac{1}{4}$	1*	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{4}{3}$
1*	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	5/4

TABLE V. Doublet functions from configuration *abcde*. See remarks in caption to Table I.

Equations $i=$ 7	8	9	10	11	12	13	14	15	16	
1	−1	0	1	0	0	−1	0	0	0	
−1	0	−1	0	1	0	0	−1	0	0	
0	−1	1	0	0	1	0	0	−1	0	
0	0	0	−1	1	−1	0	0	0	−1	
0	0	0	0	0	0	−1	1	−1	1	
Coefficients										k_μ
$-\frac{3}{4}$	$\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{1}{2}$	1*	1	0	1	1	0	4
$\frac{1}{2}$	$-\frac{1}{2}$	0	1*	$\frac{1}{2}$	$-\frac{1}{2}$	1	$\frac{1}{2}$	$-\frac{1}{2}$	0	3
$\frac{1}{2}$	$-\frac{1}{2}$	1*	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	3
$-\frac{1}{3}$	1*	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	9/4
1*	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{3}$	2

formalism, if all coefficients $a_{\mu i}$ referring to the projection Θ_i of this determinant are set equal to zero. This has the effect of shifting the element $a_{\mu\mu}=1$ one or more places to the right in the table, but does not require reordering the basis functions. An example of this occurs in Table VI.

Ordinarily, one of Eqs. (11) and (12) will involve considerably more terms than the other. For matrix elements between functions in the same configuration, it is convenient to choose $\mu \geqslant \nu$ and to evaluate Eq. (12). Then the triangular form of the coefficient matrices simplifies this expression as much as possible, so that matrix elements $(\Phi_i, H\Phi_j)$ occur only for indices such that $i \geqslant \mu \geqslant \nu \geqslant j$. Presumably this would be the appropriate formula to use for a digital computer program, which would not have to depend on checking Eq. (11) against Eq. (12). It is an immediate result of Eq. (12) that diagonal elements of H in the original basis of determinants are absent from the nondiagonal elements of H in the basis of orthonormalized projected functions.[3]

V. EVALUATION OF COEFFICIENTS

It has been shown in the previous section that the coefficients needed to evaluate matrix elements by Eq. (13), characteristic of the projection operator

TABLE VI. ^2D functions from configuration d^3. The first four equations come from matrix elements of L^+ while the fifth equation comes from matrix elements of S^+. See remarks in caption to Table I.

Equations $i=$ 6	5	4	3	2	1	
0	0	0	0	1	1	
0	$(2/3)^{\frac{1}{2}}$	0	1	1	0	
1	0	1	1	0	0	
1	$(2/3)^{\frac{1}{2}}$	0	0	0	0	
1	0	0	−1	0	1	
Coefficients						k_μ
5/4	0	1*	−1	1	−1	4
1*	$-(3/2)^{\frac{1}{2}}$	−5/4	$\frac{1}{4}$	$\frac{3}{4}$	$-\frac{3}{4}$	21/4

formalism, can be obtained without explicitly introducing the projection operator. It is necessary to find for each symmetry species an orthonormal set of functions Ψ_μ' that is a complete basis for the symmetry-adapted functions of a given configuration and for which the coefficients $x_{\mu i}$ of Eq. (5) form a trapezoidal array.

In the previous paper,[4] it was pointed out that these coefficients will always be determined by a set of homogeneous linear equations obtained from the matrix elements of the group generators. These equations are especially simple in the case of the rotation group or of groups with similar algebraic properties, when there is a step-up operator

$$J^+=J_x+iJ_y. \tag{18}$$

By a suitable choice of the basis orbitals, each Slater determinant of the basis set can be chosen to be an eigenfunction of the operator J_z, with eigenvalue M in appropriate units. A necessary and sufficient condition that a function

$$\Psi_\mu=\sum_i x_{\mu i}\Phi_i, \tag{19}$$

where $\{J_z-\mathrm{M}\}\Phi_i=0$ for all i should be an eigenfunction of $\mathbf{J}\cdot\mathbf{J}$ as well as of J_z, with total angular quantum number $J=M$, is

$$J^+\Psi_\mu\equiv 0. \tag{20}$$

Now let Φ_j be some determinant of the same configuration as the Φ_i, but satisfying

$$\{J_z-\mathrm{M}-1\}\Phi_j=0. \tag{21}$$

Then Eq. (20) implies the linear homogeneous equations

$$\sum_i (\Phi_j, J^+\Phi_i)x_{\mu i}=0 \text{ for all } j. \tag{22}$$

If there are several commuting angular momenta, or operators such as isotopic spin, then equations similar to Eq. (22) arise from each operator, and are all to be solved simultaneously.

In the previous paper,[4] Eq. (22) has been used to derive nonorthogonal projected determinants, which were then orthogonalized by the Schmidt process. As has been shown in the present paper, this orthogonalization can be accomplished more directly, if orthogonal solutions to the homogeneous equations are obtained in trapezoidal form. This can be done by the standard method of Gaussian elimination as follows:

First, the given homogeneous equations are reduced to triangular form by Gaussian elimination (subtracting multiples of one equation from the others). This process automatically removes any equations linearly dependent on the rest. Suppose that there are n basis functions with the same axial angular momentum quantum numbers, and that there are $m \leqslant n$ independent eigenfunctions of the total angular momenta, with quantum numbers J equal to the corresponding axial quantum numbers M. Then there will be $n-m$ equations

remaining after Gaussian elimination. If this has been carried out working from right to left, each of these $n-m$ equations will have its right-hand nonvanishing element in a different column. The m linearly independent solutions are determined by choosing arbitrary values for the coefficients $x_{\mu i}$ for each index i that corresponds to one of the m columns which do not contain the far right-hand element of any of the $n-m$ reduced equations. When these free coefficients have been selected, the remaining coefficients are determined, and can be calculated by inspection.

Second, a solution is obtained by setting the first $m-1$ free coefficients equal to zero, the mth equal to unity. If μ_m is the index of the column containing the mth free coefficient, then the remaining coefficients $x_{\mu_m j}$ are determined for $j>\mu_m$, and all coefficients vanish for $j<\mu_m$.

Third, to find a solution orthogonal to the first, the set of coefficients $x_{\mu_m j}{}^*$ are treated as coefficients of an additional homogeneous equation, since the orthogonality condition is

$$\sum_j x_{\mu_m j}{}^* x_{\mu j}=0, \quad \mu\neq\mu_m. \tag{23}$$

One more cycle of Gaussian elimination is carried out, which will reduce the row of coefficients $x_{\mu_m j}{}^*$ to a row terminating on the right in the μ_mth column, reducing the number of solutions to $m-1$. Then the second step is repeated, but with the $(m-1)$th free-coefficient set equal to unity.

If this process is iterated, eventually m solutions will be obtained which are mutually orthogonal, and for which the coefficients are in trapezoidal form. The coefficients $a_{\nu i}$ can be obtained by inspection from Eq. (15), equivalent to inverting a triangular matrix. The coefficients $a_{\nu i}$ for which index i does not correspond to a free coefficient in the original Gaussian elimination process are to be set equal to zero. The indices μ corresponding to free coefficients denote the orthogonalized projected determinants $\Psi_\mu{}'$ obtained from the projected determinants Θ_μ by Schmidt orthogonalization carried out in the reverse order. Thus the last function obtained by the present method will be the normalized projection of the first Slater determinant in the basis set.

An alternative procedure is to obtain m linearly independent solutions by choosing successively one of the free coefficients to be unity and the others zero. Then the Schmidt process, applied to these solutions working from right to left, will produce the same orthonormal solutions as before. Note that the orthogonalization is still carried out in the reverse order to the equivalent orthogonalization of the linearly independent projected functions. There did not appear to be any advantage in applying this alternative procedure in the examples given below, but it might be preferable when the number of equations is large and the number of solutions small.

VI. EXAMPLES

The coefficients needed for evaluation of matrix elements for eigenfunctions of total spin in configurations *abc*, *abcd*, and *abcde* are given in Tables I–V. The tables have been constructed by the method described in Sec. V. The orbitals are assumed to be orthonormal.

The Slater determinants of the configuration *abc* are enumerated as follows, using the notation

$$(abc\cdots, de\cdots)\equiv \det a\alpha b\alpha c\alpha\cdots d\beta e\beta\cdots, \tag{24}$$

and listing all determinants with $M_S=\frac{3}{2}$ first, then all those with $M_S=\frac{1}{2}$:

$$\begin{aligned}\Phi_1&=(abc,)\\ \Phi_2&=(ab,c)\\ \Phi_3&=(ac,b)\\ \Phi_4&=(bc,a).\end{aligned} \tag{25}$$

This enumeration, for each set with equal M_S, lists orbitals with spin $\alpha(m_s=\frac{1}{2})$ in each determinant in dictionary order, orbitals with spin β $(m_s=-\frac{1}{2})$ also in dictionary order, then lists the symbols for the determinants in dictionary order. The same rule is used to enumerate the determinants in configurations *abcd* and *abcde*.

The determinant Φ_1 of Eq. (25) is already an eigenfunction of total spin with $S=\frac{3}{2}$, and there are no other functions with $M_S=\frac{3}{2}$ in the configuration. Hence there is only one coefficient, $x_{11}=1$, with $k_1=1$, so no special analysis is needed.

The three determinants Φ_2, Φ_3, Φ_4 of Eq. (25), with $M_S=\frac{1}{2}$, can be combined to give two independent spin eigenfunctions with $S=\frac{1}{2}$. The coefficients for these are listed in Table I. Matrix elements between the orthonormal doublet functions, $\Psi_2{}'$ and $\Psi_3{}'$, will be evaluated as an example of the application of Eqs. (11)–(13).

It is convenient to denote the one-electron and Coulomb integrals, which are the same for all diagonal matrix elements, simply by E_0, and to use the following notation for the remaining two-electron integrals of the Coulomb repulsion in appropriate units,

$$[ab\,|\,cd]\equiv\int\int a^*(1)b(1)r_{12}{}^{-1}c^*(2)d(2)d\tau_1 d\tau_2. \tag{26}$$

The matrix elements of the nonrelativistic electronic Hamiltonian for the determinants Φ_2, Φ_3, Φ_4 are

$$\begin{aligned}H_{22}&=E_0-[ab\,|\,ba]\\ H_{32}&=H_{23}=[bc\,|\,cb]\\ H_{33}&=E_0-[ac\,|\,ca]\\ H_{42}&=H_{24}=-[ac\,|\,ca]\\ H_{43}&=H_{34}=[ab\,|\,ba]\\ H_{44}&=E_0-[bc\,|\,cb].\end{aligned} \tag{27}$$

Then, from Table I and Eq. (13),

$$(\Psi_2',H\Psi_2')$$
$$=H_{22}+\tfrac{1}{2}H_{23}-\tfrac{1}{2}H_{24}=E_0-[ab|ba]+\tfrac{1}{2}[ac|ca]+\tfrac{1}{2}[bc|cb],$$
$$(\Psi_3',H\Psi_2')$$
$$=\left\{\begin{bmatrix}-\frac{1}{2}(H_{22}+\frac{1}{2}H_{23}-\frac{1}{2}H_{24})\\+1(H_{32}+\frac{1}{2}H_{33}-\frac{1}{2}H_{34})\end{bmatrix}\begin{bmatrix}(H_{23}+H_{24})\\\end{bmatrix}\right\}^{\frac{1}{2}}$$
$$=\{(-\tfrac{3}{4}[ac|ca]+\tfrac{3}{4}[bc|cb])(-[ac|ca]+[bc|cb])\}^{\frac{1}{2}}$$
$$=-(3/4)^{\frac{1}{2}}([ac|ca]-[bc|cb]),$$
$$(\Psi_3',H\Psi_3')$$
$$=-\tfrac{1}{2}(H_{23}+H_{24})+1(H_{33}+H_{34})$$
$$=E_0+[ab|ba]-\tfrac{1}{2}[ac|ca]-\tfrac{1}{2}[bc|cb]. \quad (28)$$

If the internal check, illustrated here by comparison of the two terms in square brackets for $(\Psi_3',H\Psi_2')$, is not wanted, it is simpler to use Eq. (12). Then

$$(\Psi_3',H\Psi_2')=\left(\frac{3/2}{2}\right)^{\frac{1}{2}}(H_{23}+H_{24})$$
$$=-(3/4)^{\frac{1}{2}}([ac|ca]-[bc|cb]). \quad (29)$$

Alternatively, by Eq. (11),

$$(\Psi_3',H\Psi_2')=\left(\frac{2}{3/2}\right)^{\frac{1}{2}}\begin{bmatrix}-\frac{1}{2}(H_{22}+\frac{1}{2}H_{23}-\frac{1}{2}H_{24})\\+1(H_{32}+\frac{1}{2}H_{33}-\frac{1}{2}H_{34})\end{bmatrix}$$
$$=-(3/4)^{\frac{1}{2}}([ac|ca]-[bc|cb]). \quad (30)$$

Coefficients for the configuration *abcd*, with four orbitals outside of closed shells, are given for the triplet and singlet spin eigenfunctions in Tables II and III. The basis determinants are enumerated by the rule used before, so in analogy to Eq. (25),

$$\begin{aligned}\Phi_1&=(abcd,),\\\Phi_2&=(abc,d),\\\Phi_3&=(abd,c),\\&\cdots\\\Phi_6&=(ab,cd),\\\Phi_7&=(ac,bd),\\&\cdots.\end{aligned} \quad (31)$$

The triplet eigenfunctions of this configuration have been calculated by Löwdin,[3] and are a good example for comparison of the two methods.

Coefficients for the configuration *abcde*, with five orbitals outside of closed shells, are listed for the quartet and doublet spin eigenfunctions in Tables IV and V. By the rule used here for enumerating the basis determinants,

$$\begin{aligned}\Phi_1&=(abcde,),\\\Phi_2&=(abcd,e),\\&\cdots\\\Phi_7&=(abc,de),\\&\cdots.\end{aligned} \quad (32)$$

As before, the determinant Φ_1 is an eigenfunction of total spin, with $S=\frac{5}{2}$ in this case, and corresponds to a table of coefficients with a single entry, $x_{11}=1$.

To illustrate the application of the present method to the interaction between different configurations, consider matrix elements of the nonrelativistic Hamiltonian between doublet spin eigenfunctions from the configurations $abcf^2$ and *abcde*. The coefficients needed are given in Tables I and V, respectively, noting that the doubly-occupied orbital f^2 does not affect the derivation of the coefficients of Table I. The basis determinants with $M_S=\frac{1}{2}$ of the first configuration are

$$\begin{aligned}\Phi_{2A}&=(abf,cf),\\\Phi_{3A}&=(acf,bf),\\\Phi_{4A}&=(bcf,af).\end{aligned} \quad (33)$$

The basis determinants with $M_S=\frac{1}{2}$ of the second configuration are $\Phi_7\cdots\Phi_{16}$, as in Eq. (32).

The only nonvanishing matrix elements between the basis determinants of the two different configurations are

$$H_{2A,8}=H_{2A,9}=H_{3A,10}=H_{3A,11}=H_{4A,13}=H_{4A,14}=[df|ef]. \quad (34)$$

Since the matrix elements are real numbers, and the Hamiltonian is Hermitian, $H_{2A,8}=H_{8,2A}$, etc., and in general the order of subscripts can be freely interchanged.

On applying Eq. (13) to a typical matrix element for the orthogonalized doublet functions,

$$(\Psi_{3A}',H\Psi_{10}')=\left\{\begin{bmatrix}-\frac{1}{2}(H_{2A,10}+\frac{1}{2}H_{2A,11}-\frac{1}{2}H_{2A,12}+H_{2A,13}+\frac{1}{2}H_{2A,14}-\frac{1}{2}H_{2A,15})\\+1(H_{3A,10}+\frac{1}{2}H_{3A,11}-\frac{1}{2}H_{3A,12}+H_{3A,13}+\frac{1}{2}H_{3A,14}-\frac{1}{2}H_{3A,15})\\\cdot[\frac{1}{2}(H_{3A,7}+H_{4A,7})-\frac{1}{2}(H_{3A,8}+H_{4A,8})+1(H_{3A,10}+H_{4A,10})]\end{bmatrix}\right\}^{\frac{1}{2}}$$
$$=\{(\tfrac{3}{2}[df|ef])([df|ef])\}^{\frac{1}{2}}$$
$$=(3/2)^{\frac{1}{2}}[df|ef], \quad (35)$$

or, by Eq. (12),

$$(\Psi_{3A}',H\Psi_{10}')=(3/2)^{\frac{1}{2}}[\tfrac{1}{2}(H_{3A,7}+H_{4A,7})-\tfrac{1}{2}(H_{3A,8}+H_{4A,8})+1(H_{3A,10}+H_{4A,10})]=(3/2)^{\frac{1}{2}}[df|ef]. \quad (36)$$

If all ten matrix elements between the doublet functions of the two configurations are evaluated, those that are not zero are found to be

$$(\Psi_{2A}',H\Psi_8')=(\Psi_{3A}',H\Psi_{10}')=(3/2)^{\frac{1}{2}}[df|ef],$$
$$(\Psi_{2A}',H\Psi_9')=(\Psi_{3A}',H\Psi_{11}')=(1/2)^{\frac{1}{2}}[df|ef]. \quad (37)$$

The final example is typical of Russell-Saunders coupling in the theory of atomic spectra. Coefficients for the ^2D functions in the configuration d^3 have been obtained by the method of Sec. V, and are listed in Table VI. This example has been considered in the earlier paper,[4] where projected determinants were constructed, then orthogonalized by the Schmidt process, and the basis determinants are enumerated to agree with that reference. The matrix elements are quoted by Condon and Shortley,[12] and were derived by traditional methods.[13] The basis determinants are listed in opposite order to their enumeration in order to produce the same ^2D functions as those in Condon and Shortley, where one of the two orthogonal functions chosen arbitrarily happens to be the projected determinant Ψ_6.

The basis determinants (with $M_L=2$, $M_S=\frac{1}{2}$) are

$$\Phi_1=(d_2d_1,d_{-1}),$$
$$\Phi_2=(d_2d_0,d_0),$$
$$\Phi_3=(d_2d_{-1},d_1),$$
$$\Phi_4=(d_2d_{-2},d_2),$$
$$\Phi_5=(d_1d_0,d_1),$$
$$\Phi_6=(d_1d_{-1},d_2). \quad (38)$$

The matrix elements of the electronic Hamiltonian for the two orthonormal ^2D functions of Table VI can be expressed as linear combinations of one-electron integrals K_d and the independent radial integrals F_k, using the Gaunt coefficients c^k tabulated by Condon and Shortley.[12] From Eq. (12) or (13) the matrix elements are calculated to be

$$(\Psi_4',H\Psi_4')=3K_d+3F_0+7F_2+63F_4,$$
$$(\Psi_4',H\Psi_6')=-3(21)^{\frac{1}{2}}(F_2-5F_4), \quad (39)$$
$$(\Psi_6',H\Psi_6')=3K_d+3F_0+3F_2-57F_4,$$

in agreement with the result quoted by Condon and Shortley,[12] and with the earlier paper,[4] where the orthogonalized function Ψ_a is equal to $-\Psi_4'$ of the present paper, and Ψ_b is equal to Ψ_6'. These functions have also been obtained by Löwdin,[3] by explicit construction of the matrix of the projection operator.

Note added in proof. In more complicated examples, it is convenient to list the basis determinants in an order that leads wherever possible to symmetry-adapted functions of definite parentage. This facilitates comparison with earlier work and simplifies the calculated coefficients.

[13] C. W. Ufford and G. H. Shortley, Phys. Rev. 42, 167(1932).

Schrödinger Scattering Amplitude. I*

ALEX GROSSMANN AND TAI TSUN WU†
Physics Department, Brandeis University, Waltham, Massachusetts, and Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts

(Received March 3, 1961)

The Schrödinger scattering amplitude for a fixed potential is studied as a function of the three components of the initial momentum, the three components of the final momentum, and the square root of the energy.

1. INTRODUCTION

THE purpose of this and of the accompanying paper is to study the amplitude for scattering from a fixed potential, which need not be spherically symmetric. The amplitude f is considered as a function of the components of initial and final momenta, and of the square root of the energy. All the variables are allowed to take complex values, and a domain in which f is holomorphic is determined. The method used consists, essentially, in writing f as an inner product of square integrable functions and in applying known theorems on Hilbert space elements and operators. No use is made of the Born expansion, the Fredholm formulas, or the partial wave expansion.[1]

2. SCATTERING AMPLITUDE AS AN INNER PRODUCT

Consider the scattering of a particle of mass m by a fixed potential $V(\mathbf{x})$. For the sake of simplicity, use units in which $2m=\hbar=1$. It is assumed that $V(\mathbf{x})$ is a real, Lebesgue measurable function which satisfies the following conditions:

(a) The integral

$$\int\int |V(\mathbf{x})V(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}d\mathbf{x}d\mathbf{y} \tag{1}$$

exists.

(b) The integral

$$I(\kappa)=\int |V(\mathbf{x})|\,e^{\kappa|\mathbf{x}|}d\mathbf{x} \tag{2}$$

exists for some $\kappa>0$.

Denote by α the l.u.b. of real numbers κ such that $I(\kappa)$ exists.

The above assumptions are made throughout this and the accompaning paper. They will not be repeated.

A simple evaluation[2] shows that the integral

$$J(\gamma)=\int\int |V(\mathbf{x})V(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\times\exp\{\gamma|\mathbf{x}-\mathbf{y}|\}d\mathbf{x}d\mathbf{y} \tag{3}$$

exists for all $\gamma<\alpha$ (γ real).

Furthermore: If $V(\mathbf{x})$ is locally square integrable and satisfies

$$V(\mathbf{x})=O(e^{-\alpha|\mathbf{x}|}),\quad (|\mathbf{x}|\to\infty), \tag{4}$$

then $V(\mathbf{x})$ satisfies (1) and (2).

Finally: If $V(\mathbf{x})$ satisfies conditions (1) and (2), then the "rotated and displaced" potential $V(R^{-1}\mathbf{x}-\mathbf{c})$ satisfies the same conditions.

To every potential that satisfies (1) and (2) there correspond certain square integrable functions which will be now introduced.

Let $L^{(2)}$ denote the set of functions of $\mathbf{x}$ that are Lebesgue square integrable over the entire three-dimensional Euclidean space; $L^{(2)}$ becomes a Hilbert space if the inner product and the norm are defined—as usual—by

$$(\varphi_1,\varphi_2)=\int \varphi_1^*(\mathbf{x})\varphi_2(\mathbf{x})d\mathbf{x}\quad (\varphi_1\in L^{(2)},\varphi_2\in L^{(2)})$$

and

$$\|\varphi\|=(\varphi,\varphi)^{\frac{1}{2}}\quad (\varphi\in L^{(2)}).$$

If T is a linear operator from $L^{(2)}$ into $L^{(2)}$, then its bound norm is defined by

$$\|T\|=\sup\|T\varphi\|,$$

where $\varphi\in L^{(2)}$ and $\|\varphi\|=1$.

If $T(\mathbf{x},\mathbf{y})$ is Lebesgue square integrable over the whole six-dimensional Euclidean space, then the linear operator T defined by

$$(T\varphi)(\mathbf{x})=\int T(\mathbf{x},\mathbf{y})\varphi(\mathbf{y})d\mathbf{y}$$

is completely continuous (i.e., transforms bounded sets into compact sets). Moreover,

$$\|T\|\leq\left\{\int\int |T(\mathbf{x},\mathbf{y})|^2d\mathbf{x}d\mathbf{y}\right\}^{\frac{1}{2}}.$$

* Supported in part by the U. S. Air Force.

† Alfred P. Sloan Foundation Fellow. Now on leave at the Institute for Advanced Study, Princeton, New Jersey.

[1] A small sample of publications on the scattering amplitude is: V. Bargmann, Revs. Modern Phys. **21**, 488 (1949); A. O. Barut and K. H. Ruei (Syracuse University preprint); R. Blanckenbeckler, M. L. Goldberger, N. N. Khuri, and S. B. Treiman, Ann. Phys. **10**, 62 (1960); T. Ikebe (University of Tokyo preprint); J. M. Jauch and I. I. Zinnes, Nuovo cimento **11**, 553 (1959); R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951); A. Klein and C. Zemach, Ann. Phys. **7**, 440 (1959); A. Martin, Nuovo cimento **14**, 403 (1959); R. E. Peierls, Proc. Roy. Soc. (London) **A253**, 16 (1959); A. Ya. Povzner, Mat. Sbornik **32**, (74) 109 (1953).

[2] All the details can be found in a report entitled "Potential Scattering" available from the Physics Department, Brandeis University. The results on Lebesgue integral and on Hilbert space, to be used in the proofs, can be found, e.g., in Chap. 2 and 4 of F. Riesz and B. Sz-Nagy, *Functional Analysis* (Frederic Ungar, 1955).

Consider the function

$$s(\mathbf{x})=\begin{cases} 1 & \text{if } V(\mathbf{x})>0 \\ 0 & \text{if } V(\mathbf{x})=0 \\ -1 & \text{if } V(\mathbf{x})<0. \end{cases} \tag{5}$$

Given any complex vector (that is, triplet of complex numbers) $\mathbf{q}$, define $a_\mathbf{q}$ and $b_\mathbf{q}$ by

$$a_\mathbf{q}(\mathbf{x})=(2\pi)^{-\frac{3}{2}}s(\mathbf{x})\,|V(\mathbf{x})|^{\frac{1}{2}}e^{i\mathbf{q}\cdot\mathbf{x}}, \tag{6}$$

$$b_\mathbf{q}(\mathbf{x})=(2\pi)^{-\frac{3}{2}}|V(\mathbf{x})|^{\frac{1}{2}}\exp\{-i\mathbf{q}^*\cdot\mathbf{x}\}, \tag{7}$$

where the star denotes complex conjugation.

If the real vector

$$\operatorname{Im}\mathbf{q}=-\tfrac{1}{2}i(\mathbf{q}-\mathbf{q}^*)$$

satisfies

$$|\operatorname{Im}\mathbf{q}|<\tfrac{1}{2}\alpha,$$

then $a_\mathbf{q}\in L^{(2)}$, $b_\mathbf{q}\in L^{(2)}$, and

$$\|a_\mathbf{q}\|^2=\|b_\mathbf{q}\|^2\leqslant(2\pi)^{-3}I(2|\operatorname{Im}\mathbf{q}|), \tag{9}$$

where I is defined by (2).

For any complex number k, let $A(k)$ be the linear operator

$$A(\mathbf{x},\mathbf{y};k)=-(4\pi)^{-1}s(\mathbf{x})\,|V(\mathbf{x})V(\mathbf{y})|^{\frac{1}{2}}|\mathbf{x}-\mathbf{y}|^{-1}\times\exp\{ik|\mathbf{x}-\mathbf{y}|\}. \tag{10}$$

If

$$\operatorname{Im}k>-\tfrac{1}{2}\alpha, \tag{11}$$

then $A(k)$ is a completely continuous operator from $L^{(2)}$ into $L^{(2)}$, and

$$\|A(k)\|^2\leqslant(4\pi)^{-2}J(-2\operatorname{Im}k), \tag{12}$$

where J is defined by (3).

In the remainder of this paper and in the accompanying paper, it is always assumed that (11) *holds, and that complex vectors* $\mathbf{p}$, $\mathbf{q}$ *satisfy* (8).

The function f, to be introduced now, is the main subject of study in this work. Define G_0 by

$$G_0(\mathbf{x},\mathbf{y};k)=-(4\pi)^{-1}|\mathbf{x}-\mathbf{y}|^{-1}\exp\{ik|\mathbf{x}-\mathbf{y}|\} \tag{13}$$

and define G as the solution of the integral equation

$$G(\mathbf{x},\mathbf{y};k)=G_0(\mathbf{x},\mathbf{y};k)+\int G_0(\mathbf{x},\mathbf{z};k)V(\mathbf{z})G(\mathbf{z},\mathbf{y};k)d\mathbf{z}, \tag{14}$$

provided that (14) has a unique solution. If $\mathbf{q}$ and $\mathbf{p}$ are two complex vectors, define f to be

$$f(\mathbf{q},\mathbf{p};k)=(2\pi)^{-3}\int e^{-i\mathbf{q}\cdot\mathbf{x}}V(\mathbf{x})e^{i\mathbf{p}\cdot\mathbf{x}}d\mathbf{x}+(2\pi)^{-3}\int\int e^{-i\mathbf{q}\cdot\mathbf{x}}V(\mathbf{x})G(\mathbf{x},\mathbf{y};k)V(\mathbf{y})e^{i\mathbf{p}\cdot\mathbf{y}}d\mathbf{x}d\mathbf{y}. \tag{15}$$

Notice that f is a function of seven independent complex variables.

Theorem: *If k is such that the number one is not an eigenvalue of $A(k)$, then*

$$f(\mathbf{q},\mathbf{p};k)=(b_{-\mathbf{q}},[1-A(k)]^{-1}a_\mathbf{p}). \tag{16}$$

Proof: Multiply (14) by $s(\mathbf{x})\,|V(\mathbf{x})V(\mathbf{y})|^{\frac{1}{2}}$ and take into account that

$$V(\mathbf{z})=|V(\mathbf{z})|^{\frac{1}{2}}s(\mathbf{z})\,|V(\mathbf{z})|^{\frac{1}{2}}.$$

If $N(\mathbf{x},\mathbf{y};k)$ is defined as

$$N(\mathbf{x},\mathbf{y};k)=s(\mathbf{x})\,|V(\mathbf{x})V(\mathbf{y})|^{\frac{1}{2}}G(\mathbf{x},\mathbf{y};k), \tag{17}$$

then (14) becomes

$$N(k)=A(k)+A(k)N(k). \tag{18}$$

Equation (18) has, among all linear operators in $L^{(2)}$, the unique solution

$$N(k)=-1+[1-A(k)]^{-1}. \tag{19}$$

This follows from the Fredholm alternative (which holds for all completely continuous operators) and from the assumption on k. Then (15) is

$$f(\mathbf{q},\mathbf{p};k)=(b_{-\mathbf{q}},a_\mathbf{p})+(b_{-\mathbf{q}},N(k)a_\mathbf{p})=(b_{-\mathbf{q}},[1+N(k)]a_\mathbf{p})=(b_{-\mathbf{q}},[1-A(k)]^{-1}a_\mathbf{p}), \tag{20}$$

Q.E.D.

If all seven variables are real and if $|\mathbf{q}|=|\mathbf{p}|=|k|$, then f is the scattering amplitude, multiplied by $(2\pi^2)^{-1}$.

3. SYMMETRIES

It is sometimes convenient to express f in terms of another set of variables. Introduce

$$\mathbf{P}=\tfrac{1}{2}(\mathbf{p}+\mathbf{q}), \tag{21}$$

$$\mathbf{t}=\tfrac{1}{2}(\mathbf{p}-\mathbf{q}), \tag{22}$$

and define g by

$$g(\mathbf{t},\mathbf{P};k)=f(\mathbf{q},\mathbf{p};k). \tag{23}$$

Theorem 2: *g satisfies*

$$g(\mathbf{t},\mathbf{P};k)=g(\mathbf{t},-\mathbf{P};k) \tag{24}$$

and

$$g^*(\mathbf{t},\mathbf{P};k)=g(-\mathbf{t}^*,-\mathbf{P}^*;-\mathbf{k}^*). \tag{25}$$

Proof: Notice first that $L^{(2)}$ has a richer structure than an abstract Hilbert space. Given any $\varphi\in L^{(2)}$, the complex conjugate φ^* is defined by $\varphi^*(\mathbf{x})=[\varphi(\mathbf{x})]^*$. Given an integral operator B, the complex conjugate B^* and the transposed B_T are defined by the kernels $B^*(\mathbf{x},\mathbf{y})$ and $B(\mathbf{y},\mathbf{x})$, respectively.

In

$$g(\mathbf{t},\mathbf{P};k)=(b_\mathbf{t},a_\mathbf{t})+(b_{\mathbf{t}-\mathbf{P}},N(k)a_{\mathbf{t}+\mathbf{P}}), \tag{26}$$

the first term does not contain $\mathbf{P}$. In order to study the second term, introduce the selfadjoint linear operator

S from $L^{(2)}$ into $L^{(2)}$, defined by

$$(S\varphi)(\mathbf{x})=s(\mathbf{x})\varphi(\mathbf{x}).$$

For every $\mathbf{q}$,

$$Sa_{\mathbf{q}}=b_{\mathbf{q}}^*, \quad S^2a_{\mathbf{q}}=a_{\mathbf{q}},$$
$$Sb_{\mathbf{q}}=a_{\mathbf{q}}^*, \quad S^2b_{\mathbf{q}}=b_{\mathbf{q}}.$$

Furthermore,

$$SA(k)S=A_T(k)$$

and

$$S^2A(k)=A(k)S^2=A(k).$$

It follows from the uniqueness of the solution to (18) that

$$SN(k)S=N_T(k)$$

whenever $N(k)$ exists. So

$$(b\mathbf{t}-\mathbf{p},N(k)a\mathbf{t}+\mathbf{p})=(b\mathbf{t}-\mathbf{p},S^2N(k)S^2a\mathbf{t}+\mathbf{p})$$
$$=(a^*\mathbf{t}-\mathbf{p},N_T(k)b^*\mathbf{t}+\mathbf{p})=(b\mathbf{t}+\mathbf{p},N(k)a\mathbf{t}-\mathbf{p})$$

which proves (24). The proof of (25) is completely straightforward.

Let R represent a real three-dimensional rotation. Consider the transformation

$$\mathbf{x}'=R\mathbf{x}+\mathbf{d}, \tag{27}$$

where $\mathbf{d}$ is a fixed real vector. Define a potential V' by

$$V'(\mathbf{x}')=V(\mathbf{x}). \tag{28}$$

Let g' be the function (23) that arises from the potential V'. Then a simple calculation yields the following result:

Theorem 3: The functions g and g' are related by

$$g'(\mathbf{t},\mathbf{P};k)=e^{2i\mathbf{t}\cdot\mathbf{d}}g(R^{-1}\mathbf{t},R^{-1}\mathbf{P};k). \tag{29}$$

In particular, if V is invariant with respect to the rigid motion (29), then

$$g(\mathbf{t},\mathbf{P};k)=e^{-2i\mathbf{t}\cdot\mathbf{d}}g(R\mathbf{t},R\mathbf{P};k).$$

4. ANALYTICITY PROPERTIES

Let $\mathcal{O}$ be an open set of the k plane. For every $k\in\mathcal{O}$, let $B(k)$ be a bounded operator from $L^{(2)}$ into $L^{(2)}$. If, for every $\varphi\in L^{(2)}$, $\psi\in L^{(2)}$, the function

$$(\varphi,B(k)\psi)$$

is holomorphic in $\mathcal{O}$, then $B(k)$ is said[3] to be holomorphic in $\mathcal{O}$. If $B(k)$ is holomorphic in an open neighborhood of a point k_0, then $B(k)$ is said to be holomorphic at k_0.

The operator family $A(k)$, defined by (10), is holomorphic in the open half-plane (11). This can be shown by applying Fubini's theorem and Morera's theorem to the integral

$$\oint(\varphi,A(k)\psi)dk=\oint\int\int\varphi^*(\mathbf{x})A(\mathbf{x},\mathbf{y};k)\psi(\mathbf{y})d\mathbf{x}d\mathbf{y}dk$$

taken over an arbitrary closed path in (11). The nth power $(A(k))^n$ is also holomorphic in (11), for $n=0, 1, 2, \cdots$.

About $N(k)$, defined by (19), the following assertion holds:

Theorem 4: In any compact set contained in the half-plane (11), *there can be only finitely many points at which $N(k)$ is not holomorphic. $N(k)$ is holomorphic at every point where it exists.*

It is convenient to prove first

Lemma 1: There exists a number γ_0 such that $N(k)$ exists and is holomorphic in the open half-plane

$$\mathrm{Im}k>\gamma_0. \tag{30}$$

Proof: By (3), there exists a number γ_0 such that

$$(4\pi)^{-2}J(-2\gamma_0)=\theta^2<1.$$

Then, by (12),

$$\|A(k)\|\leqslant\theta<1 \tag{31}$$

in the half-plane (30), and the series

$$A(k)+(A(k))^2+\cdots \quad (\mathrm{Im}\,k>\gamma_0) \tag{32}$$

converges in norm to $N(k)$. For arbitrary $\varphi\in L^{(2)}$ and $\psi\in L^{(2)}$, the series

$$\sum_{n=1}^{\infty}(\varphi,(A(k))^n\psi) \tag{33}$$

is majorized by

$$\sum_{n=1}^{\infty}\|\varphi\|\|\psi\|\theta^n, \tag{34}$$

and so converges uniformly in (30); its sum is consequently holomorphic in (30), which proves the lemma.

Lemma 2: If k' and k'' are any two points in (11), then

$$\|A(k')-A(k'')\|^2\leqslant(4\pi)^{-2}|k'-k''|^2J''(-2\kappa), \tag{35}$$

where κ is the smaller of the numbers Im k', Im k'', and

$$J''(\gamma)=\int\int|V(\mathbf{x})V(\mathbf{y})|\times\exp\{\gamma|\mathbf{x}-\mathbf{y}|\}d\mathbf{x}d\mathbf{y} \quad (\gamma<\alpha). \tag{36}$$

Proof: There exists a point k on the line joining k' and k'', such that

$$|e^{ik'|\mathbf{x}-\mathbf{y}|}-e^{ik''|\mathbf{x}-\mathbf{y}|}|\leqslant|k'-k''|\,|\mathbf{x}-\mathbf{y}|\,|e^{ik|\mathbf{x}-\mathbf{y}|}|.$$

Consequently,

$$\|A(k')-A(k'')\|^2\leqslant(4\pi)^{-2}|k'-k''|^2J''(-2\,\mathrm{Im}\,k),$$

and the assertion follows from the remark that J'' is nondecreasing.

[3] See H. G. Garnir, *Les Problèmes aux Limites de la Physique Mathématique* (Birkhäuser, Basel, 1958), pp. 39–55.

Lemma 3: Let $\mathfrak{D}$ be an open bounded subset of (11), with the property that

$$\|A(k')-A(k'')\| \leqslant \tfrac{1}{3} \tag{37}$$

for every $k' \in \mathfrak{D}$, $k'' \in \mathfrak{D}$. If $N(k)$ exists for at least one point $k_0 \in \mathfrak{D}$, then there can be only finitely many points in $\mathfrak{D}$ at which $N(k)$ is not holomorphic.

Proof: The operator $A(k_0)$ can be decomposed into a sum

$$A(k_0)=K_0+L \tag{38}$$

such that $\|L\| \leqslant \frac{1}{3}$ and that the range of K_0 is finite dimensional. Define

$$C(k)=A(k)-K_0 \quad (k \in \mathfrak{D}). \tag{39}$$

Then, for every $k \in \mathfrak{D}$,

$$\|C(k)\| \leqslant \|A(k)-A(k_0)\|+\|A(k_0)-K_0\| \leqslant \tfrac{2}{3}<1. \tag{40}$$

Since K_0 does not depend on k, $C(k)$ is holomorphic in $\mathfrak{D}$. By (40), $[1-C(k)]^{-1}$ exists for every $k \in \mathfrak{D}$ and is holomorphic in $\mathfrak{D}$.

Consider the linear operator

$$H(k)=[1-C(k)]^{-1}K_0. \tag{41}$$

The range of $H(k)$ is finite dimensional, and $H(k)$ is holomorphic in $\mathfrak{D}$. By (39),

$$[1-A(k)]=[1-C(k)][1-H(k)]. \tag{42}$$

Consequently, the operator $[1-A(k)]^{-1}=1+N(k)$ exists whenever the number one is not an eigenvalue of $H(k)$.

Since $H(k)$ is holomorphic in $\mathfrak{D}$ and the range of $H(k)$ is finite dimensional, the set $\mathfrak{S}$ of points $k \in \mathfrak{D}$ at which the number one is an eigenvalue of $H(k)$ consists of the zeros of a function holomorphic in $\mathfrak{D}$. Accordingly, either $\mathfrak{S}$ contains a finite number of points, or $\mathfrak{S}=\mathfrak{D}$. The last alternative contradicts the assumption that $N(k_0)$ exists.

At all the points of $\mathfrak{D}-\mathfrak{S}$, the operator $[1-H(k)]^{-1}$ is holomorphic, and so is the product[4]

$$[1-H(k)]^{-1}[1-C(k)]^{-1}=[1-A(k)]^{-1}.$$

This proves Lemma 3.

Let now $\mathfrak{C}$ be any compact subset of (11), and $\lambda > -\frac{1}{2}\alpha$ the smallest of numbers $\operatorname{Im} k (k \in \mathfrak{C})$. There exists a chain of open sets $\mathfrak{D}_1, \cdots \mathfrak{D}_r$, contained in (11) and such that

(a) The diameter of $\mathfrak{D}_i$ $(i=1, \cdots r)$ is less than $\frac{4}{3}\pi J''(-2\lambda)^{-\frac{1}{2}}$.

(b) The intersection $\mathfrak{D}_i \cap \mathfrak{D}_{i+1}$ $(i=1, \cdots r-1)$ is nonempty.

(c) $\mathfrak{D}_1$ is contained in the half-plane (30).

(d) The union $\mathfrak{D}_1 \cup \cdots \cup \mathfrak{D}_r$ covers $\mathfrak{C}$. By (35), every set $\mathfrak{D}_i$ has the property (37). Because of Lemma 1, $N(k)$ is holomorphic in $\mathfrak{D}_1$. The assertion of Theorem 4 follows by successive application of Lemma 3 to $\mathfrak{D}_2$, $\mathfrak{D}_3 \cdots \mathfrak{D}_r$.

Theorem 5: Assume that

$$\operatorname{Im} k > -\tfrac{1}{2}\alpha,$$
$$|\operatorname{Im} \mathbf{p}| < \tfrac{1}{2}\alpha,$$
$$|\operatorname{Im} \mathbf{q}| < \tfrac{1}{2}\alpha,$$

and that k is such that $N(k)$ exists. Then f is holomorphic at $(\mathbf{q},\mathbf{p}; k)$.

Proof: For every $\varphi \in L^{(2)}$ and $\psi \in L^{(2)}$, the function $(\varphi, a_{\mathbf{p}})$, [respectively $(b_{-\mathbf{q}}, \psi)$] is holomorphic at $\mathbf{p}$ (respectively $\mathbf{q}$). This is shown as above, with the help of Morera's theorem. The assertion follows then from Theorem 4.

ACKNOWLEDGMENTS

For very helpful discussions we would like to thank Professor V. Bargmann, Professor R. W. P. King, and Professor S. S. Schweber.

[4] See reference 3, p. 45.

Schrödinger Scattering Amplitude. II*

ALEX GROSSMANN
Physics Department, Brandeis University, Waltham, Massachusetts

(Received March 3, 1961)

The results of the preceding paper are used to indicate procedures of calculation of the scattering amplitude, obtain several expansions, find bounds on the variation of the amplitude under a change in the potential, and study multiple scattering.

1. INTRODUCTION

IN the preceding paper[1] the Schrödinger scattering amplitude f was written in a form that involves only elements of $L^{(2)}$ and bounded operators in that space. This form will now be used for a more detailed study of f. Whenever f is expanded into a series, an explicit estimate of the remainder will be given.

2. CALCULATION OF f

Let δ be a number such that $0<\delta<1$. Denote by $\mathfrak{D}$ an open subset of (I.11), such that

$$\|A(k')-A(k'')\| \leqslant \tfrac{1}{2}\delta \tag{1}$$

whenever $k'\in\mathfrak{D}$, $k''\in\mathfrak{D}$. Sets $\mathfrak{D}$ that have the property (1) can be found by (I.35) and (I.36).

Chose an arbitrary $k_0\in\mathfrak{D}$. Find an operator K_0 of finite-dimensional range, such that

$$\|A(k_0)-K_0\| \leqslant \tfrac{1}{2}\delta. \tag{2}$$

A suitable operator K_0 is obtained if $A(\mathbf{x},\mathbf{y};k_0)$ is approximated (in norm) by a step function $K_0(\mathbf{x},\mathbf{y})$.

Define

$$C(k)=A(k)-K_0. \tag{3}$$

For every $k\in\mathfrak{D}$, the series

$$\sum_{h=0}^{\infty}(C(k))^h=[1-C(k)]^{-1} \tag{4}$$

is convergent in norm, and

$$\|[1-C(k)]^{-1}-\sum_{h=0}^{m}(C(k))^h\| \leqslant \frac{\delta^m}{1-\delta}. \tag{5}$$

Denote by $\chi_1\cdots\chi_n$ an orthonormal basis of the range of K_0. Such a basis is immediately found if $K_0(\mathbf{x},\mathbf{y})$ is a step function.

Consider the $n\times n$ matrix $\mathbf{M}(k)$ defined by

$$M_{rs}(k)=(\chi_r, K_0[1-C(k)]^{-1}\chi_s) \quad (r,s=1,\cdots n) \tag{6}$$

and the matrix

$$\mathbf{R}(k)=[\mathbf{1}-\mathbf{M}(k)]^{-1}. \tag{7}$$

Theorem 1: *For every* $k\in\mathfrak{D}$,

$$f(\mathbf{q},\mathbf{p};k)=(b_{-\mathbf{q}},[1-C(k)]^{-1}a_{\mathbf{p}})+(b_{-\mathbf{q}},[1-C(k)]^{-1}\chi_s) \times R_{sm}(k)(\chi_m, K_0[1-C(k)]^{-1}a_{\mathbf{p}}). \tag{8}$$

Theorem 1 is proved by a simple calculation,[2] which also shows that

$$\|[1-A(k)]^{-1}\| \leqslant \frac{1}{1-\delta}+\frac{1}{(1-\delta)^2}\Big[\sum_{s=1}^{n}\sum_{m=1}^{n}|R_{sm}(k)|^2\Big]^{\frac{1}{2}}. \tag{9}$$

The possible singularities of f are the zeroes of $\det[1-\mathbf{M}(k)]$.

3. VARIATION OF THE POTENTIAL

Problem: Given two potentials V_1 and V_2, determine by how much amplitude f_1 (due to V_1) can differ from the amplitude f_2 (due to V_2).

The potentials V_1 and V_2 are assumed to satisfy conditions (1) and (2) of (I).

Let λ be real, and $\lambda>-\alpha$. Consider the number

$$d(V_1,V_2;\lambda)=(4\pi)^{-1}\Big\{\int\int \exp\{-\lambda|\mathbf{x}-\mathbf{y}|\}|\mathbf{x}-\mathbf{y}|^{-2} \times[s_1(\mathbf{x})|V_1(\mathbf{x})V_1(\mathbf{y})|^{\frac{1}{2}} -s_2(\mathbf{x})|V_2(\mathbf{x})V_2(\mathbf{y})|^{\frac{1}{2}}]^2 d\mathbf{x}d\mathbf{y}\Big\}^{\frac{1}{2}}. \tag{10}$$

The number $d(V_1,V_2;\lambda)$ is the norm of the difference—in the Hilbert space of all functions square integrable over the x-y space—between the function

$$(4\pi)^{-1}s_1(\mathbf{x})|V_1(\mathbf{x})V_1(\mathbf{y})|^{\frac{1}{2}}|\mathbf{x}-\mathbf{y}|^{-1}\exp\{-\tfrac{1}{2}\lambda|\mathbf{x}-\mathbf{y}|\}$$

and the function

$$(4\pi)^{-1}s_2(\mathbf{x})|V_2(\mathbf{x})V_2(\mathbf{y})|^{\frac{1}{2}}|\mathbf{x}-\mathbf{y}|^{-1}\exp\{-\tfrac{1}{2}\lambda|\mathbf{x}-\mathbf{y}|\}.$$

Consequently, $d(V_1,V_2;\lambda)$ has all the properties of a "distance" between V_1 and V_2.

Given a real vector $\mathbf{c}$ which satisfies

$$|\mathbf{c}|<\alpha,$$

consider the numbers

$$t(V_1,V_2;\mathbf{c}) =(2\pi)^{-\frac{3}{2}}\Big\{\int[s_1(\mathbf{x})|V_1(\mathbf{x})|^{\frac{1}{2}}-s_2(\mathbf{x})|V_2(\mathbf{x})|^{\frac{1}{2}}]^2 \times\exp[-\mathbf{c}\cdot\mathbf{x}]d\mathbf{x}\Big\}^{\frac{1}{2}} \tag{11}$$

* Supported by AF Contract.

[1] A. Grossmann and T. T. Wu, J. Math. Phys. 2, 710 (1961), hereafter referred to as (I). The assumptions and notations of that paper will be used without special mention.

[2] See reference 2 of (I) for details of all the proofs in this paper.

and

$$r(V_1,V_2;\mathbf{c})=(2\pi)^{-\frac{3}{2}}\left\{\int[|V_1(\mathbf{x})|^{\frac{1}{2}}-|V_2(\mathbf{x})|^{\frac{1}{2}}]^2 \times\exp[-\mathbf{c}\cdot\mathbf{x}]d\mathbf{x}\right\}^{\frac{1}{2}}, \quad (12)$$

which also have the properties of a distance.

Define $A_1(k)$, $A_2(k)$, $a_{\mathrm{p};1}$, $a_{\mathrm{p};2}$, $b_{\mathrm{q};1}$ and $b_{\mathrm{q};2}$ as in (I).

Theorem 2: *If k is such that*

$$\|[1-A_1(k)]^{-1}\|\leqslant\beta_1<\infty \quad (13)$$

and

$$\|[1-A_2(k)]^{-1}\|\leqslant\beta_2<\infty, \quad (14)$$

then

$$\begin{aligned}|f_1(\mathbf{q},\mathbf{p};k)-f_2(\mathbf{q},\mathbf{p};k)|&\leqslant\beta_1\|b_{\mathrm{q};1}\|t(V_1,V_2;2\,\mathrm{Im}\,\mathbf{p})\\&+\beta_1\beta_2\|b_{\mathrm{q};1}\|\|a_{\mathrm{p};2}\|d(V_1,V_2;2\,\mathrm{Im}\,k)\\&+\beta_2\|a_{\mathrm{p};2}\|r(V_1,V_2;2\,\mathrm{Im}\,\mathbf{q}).\end{aligned} \quad (15)$$

Proof: Notice that

$$\begin{aligned}\|A_1(k)-A_2(k)\|&\leqslant d(V_1,V_2;2\,\mathrm{Im}k),\\\|a_{\mathrm{p};1}-a_{\mathrm{p};2}\|&=t(V_1,V_2;2\,\mathrm{Im}\mathbf{p}),\\\|b_{-\mathrm{q};1}-b_{-\mathrm{q};2}\|&=r(V_1,V_2;2\,\mathrm{Im}\mathbf{q}),\end{aligned}$$

and that

$$\begin{aligned}&[1-A_1(k)]^{-1}-[1-A_2(k)]^{-1}\\&\quad=[1-A_1(k)]^{-1}[A_1(k)-A_2(k)][1-A_2(k)]^{-1},\end{aligned}$$

which gives

$$\|[1-A_1(k)]^{-1}-[1-A_2(k)]^{-1}\|\leqslant\beta_1\beta_2\|A_1(k)-A_2(k)\|.$$

The assertion of the theorem follows now from (I.16).

The bounds (13) and (14) can be found, e.g., through (9). The norms $\|a_{\mathrm{p}}\|$, etc., are explicitly known.

Similar considerations yield bounds on the variation (with V) of the partial derivatives of f, and on the variation of f with its arguments.

4. EFFECTIVE RANGE EXPANSION

The analyticity properties established in Sec. 4 of (I) make possible the study of power series expansions of f. It is convenient to use, in this section, the variables $\mathbf{P}$ and $\mathbf{t}$, defined by (I.21) and (I.22). [See also (I.23).]

Let $\mathbf{v}=(v_1,v_2,v_3)$ and $\mathbf{W}=(W_1,W_2,W_3)$ be triplets of nonnegative integers. Write

$$\begin{aligned}|\mathbf{v}|&=v_1+v_2+v_3 & |\mathbf{W}|&=W_1+W_2+W_3\\\mathbf{v}!&=v_1!v_2!v_3! & \mathbf{W}!&=W_1!W_2!W_3!\\\mathbf{t}^{\mathbf{v}}&=t_1^{v_1}t_2^{v_2}t_3^{v_3} & \mathbf{P}^{\mathbf{W}}&=P_1^{W_1}P_2^{W_2}P_3^{W_3}\end{aligned}$$

$$\frac{\partial^{|\mathbf{v}|+|\mathbf{W}|}}{\partial\mathbf{t}^{\mathbf{v}}\partial\mathbf{P}^{\mathbf{W}}}=\frac{\partial^{v_1+v_2+v_3+W_1+W_2+W_3}}{\partial t_1^{v_1}\partial t_2^{v_2}\partial t_3^{v_3}\partial P_1^{W_1}\partial P_2^{W_2}\partial P_3^{W_3}}.$$

The Taylor expansion of g around $\mathbf{t}=\mathbf{P}=0$ is

$$g(\mathbf{t},\mathbf{P};k)=\sum_{\mathbf{v}}\sum_{\mathbf{W}}\frac{1}{\mathbf{v}!\mathbf{W}!}\mathbf{t}^{\mathbf{v}}\mathbf{P}^{\mathbf{W}}\left(\frac{\partial^{|\mathbf{v}|+|\mathbf{W}|}g}{\partial t^{\mathbf{v}}\partial P^{\mathbf{W}}}\right)_{\mathbf{t}=\mathbf{P}=0}. \quad (16)$$

Theorem 3: *If k is such that $N(k)$ exists, then the series* (16) *is absolutely convergent in the domain*

$$\sum_{i=1}^{3}(|t_i|+|P_i|)<\tfrac{1}{2}\alpha. \quad (17)$$

Proof: By Theorem I.5, the function $g(\mathbf{t},\mathbf{P};k)$ is holomorphic in the domain

$$\begin{aligned}|\mathrm{Im}(\mathbf{t}-\mathbf{P})|&<\tfrac{1}{2}\alpha\\|\mathrm{Im}(\mathbf{t}+\mathbf{P})|&<\tfrac{1}{2}\alpha,\end{aligned}$$

which contains the domain (17). Furthermore, (17) is a Reinhardt domain. The assertion then follows from a standard theorem of the theory of functions of several complex variables.

Because of (I.24), all the terms with odd $|\mathbf{W}|$ vanish in (16).

In every circle $|k|\leqslant r$, contained in (I.11), there are, by Theorem I.4, only finitely many singularities of $N(k)$. Let r_0 be a number such that the circle $k\leqslant r_0$ is free of singularities. Assume that $r_0>0$.

If (17) holds and if $|k|<r_0$, then g is the sum of the absolutely convergent power series

$$g(\mathbf{t},\mathbf{P};k)=\sum_{\mathbf{v}}\sum_{\mathbf{W}}\sum_{m=0}^{\infty}\frac{k^m}{v!W!m!}\mathbf{t}^{\mathbf{v}}\mathbf{P}^{\mathbf{W}}\hat{g}(\mathbf{v},\mathbf{W};m), \quad (18)$$

where

$$\hat{g}(\mathbf{v},\mathbf{W};m)=\left(\frac{\partial^{|\mathbf{v}|+|\mathbf{W}|+m}g}{\partial\mathbf{t}^{\mathbf{v}}\partial\mathbf{P}^{\mathbf{W}}\partial k^m}\right)_{\mathbf{t}=\mathbf{P}=k=0}.$$

If $|\mathbf{W}|$ is odd, then

$$\hat{g}(\mathbf{v},\mathbf{W};m)=0.$$

For every $\mathbf{v}$, $\mathbf{W}$; m,

$$\hat{g}^*(\mathbf{v},\mathbf{W};m)=(-1)^{|\mathbf{v}|+|\mathbf{W}|+m}\hat{g}(\mathbf{v},\mathbf{W};m).$$

All the numbers $\hat{g}(\mathbf{v},\mathbf{W};m)$ can be determined by solving the single operator equation

$$N(0)=A(0)+A(0)N(0).$$

The methods of Sec. 2 yield "*a priori*" bounds on $g(\mathbf{t},\mathbf{P};k)$. These bounds make it possible to estimate, in a routine way, the error committed by truncating the series (18).

5. MULTIPOLE EXPANSION

In this section, the amplitude f will be expanded into a series, the terms of which transform in a simple way if the potential is rotated.

In the expressions

$$a_{\mathbf{p}}(\mathbf{x})=(2\pi)^{-\frac{3}{2}}s(\mathbf{x})|V(\mathbf{x})|^{\frac{1}{2}}e^{i\mathbf{p}\cdot\mathbf{x}}=a_0(\mathbf{x})e^{i\mathbf{p}\cdot\mathbf{x}} \quad (19)$$

$$b_{-\mathbf{q}}^*(\mathbf{x})=(2\pi)^{-\frac{3}{2}}|V(\mathbf{x})|^{\frac{1}{2}}e^{-i\mathbf{q}\cdot\mathbf{x}}=b_0^*(\mathbf{x})e^{-i\mathbf{q}\cdot\mathbf{x}}, \quad (20)$$

the exponentials can be expanded into spherical waves.

It is convenient to write this expansion as

$$e^{i\mathbf{q}\cdot\mathbf{x}}=\sum_{lm} h_l(q^2x^2)y_{lm}(\mathbf{x})y_{lm}^*(\mathbf{q}). \tag{21}$$

Here,

$$h_l(z)=4\pi i^l z^{-\frac{1}{2}l} j_l(z^{\frac{1}{2}}) = (\pi)^{\frac{1}{2}}2^{1-l}i^l\sum_{n=0}^{\infty}\frac{(-1)^n z^n}{4^n\Gamma(n+1)\Gamma(l+\frac{3}{2}+n)} \tag{22}$$

is an entire function of z, and y_{lm} is the solid harmonic.[3]

Define

$$a_{\mathbf{p}^2;\,lm}(\mathbf{x})=a_0(\mathbf{x})h_l(x^2p^2)y_{lm}(\mathbf{x}) \tag{23}$$

$$b_{\mathbf{q}^2;\,lm}(\mathbf{x})=b_0(\mathbf{x})h_l^*(x^2q^2)y_{lm}^*(\mathbf{x}) \tag{24}$$

$$(l=0,1,2,\cdots;\ m=-l,\ -l+1,\ \cdots l).$$

For every l, m,

$$a_{\mathbf{p}^2;\,lm}\in L^{(2)} \quad \text{and} \quad b_{\mathbf{q}^2;\,lm}\in L^{(2)}.$$

This follows from the standing assumption (I.8), the inequality

$$|\mathrm{Im}(q^2)^{\frac{1}{2}}|\leqslant|\mathrm{Im}\mathbf{q}|$$

and the bound

$$|h_l(z)|\leqslant\sqrt{2}(2\pi)^{\frac{1}{2}}2^{-l-1}[\Gamma(l+\tfrac{3}{2})]^{-1}\exp\{|\mathrm{Im}z^{\frac{1}{2}}|\}.$$

Theorem 4: If

$$\kappa\equiv 3(|\mathrm{Re}\mathbf{p}|+|\mathrm{Im}\mathbf{p}|)<\tfrac{1}{2}\alpha, \tag{25}$$

then the series

$$\sum_{lm} a_{p^2;\,lm}y_{lm}^*(\mathbf{p}) \tag{26}$$

converges strongly to $a_{\mathbf{p}}$, *and*

$$\Big\|a_{\mathbf{p}}-\sum_{l=0}^{L}\sum_{m=-l}^{l} a_{p^2;\,lm}y_{lm}^*(\mathbf{p})\Big\|^2 \leqslant(2\pi)^{-3}\left(\frac{\kappa^{L+1}}{(L+1)!}\right)^2 I_{2L+2}(2\kappa), \tag{27}$$

where I_l *is defined by*

$$I_l(\gamma)=\int|V(\mathbf{x})|\,|\mathbf{x}|^l e^{\gamma|\mathbf{x}|}d\mathbf{x}, \quad (\gamma<\alpha). \tag{28}$$

Proof: A straightforward evaluation shows that, for every real $\mathbf{x}$ and every complex $\mathbf{q}$,

$$\Big|e^{i\mathbf{q}\cdot\mathbf{x}}-\sum_{l=0}^{L}\sum_{m=-l}^{l} h_l(q^2x^2)y_{lm}(\mathbf{x})y_{lm}^*(\mathbf{q})\Big| \leqslant\frac{\kappa^{L+1}|\mathbf{x}|^{L+1}}{(L+1)!}e^{\kappa|\mathbf{x}|}.$$

Substituted into the left-hand side of (27), this proves the theorem.

[3] The definitions concerning harmonics, Clebsch-Gordan coefficients, and representations of the rotation group are the same as in A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1957).

The condition (25) can probably be relaxed.

The expansion of $b_{-\mathbf{q}}$ is

$$b_{-\mathbf{q}}=\sum_{lm} b_{q^2;\,lm}y_{lm}(-\mathbf{q}). \tag{29}$$

It is strongly convergent if

$$3(|\mathrm{Re}\mathbf{q}|+|\mathrm{Im}\mathbf{q}|)<\tfrac{1}{2}\alpha. \tag{30}$$

Define $f_{m_1m_2}^{(l_1l_2)}$ by

$$f_{m_1m_2}^{(l_1l_2)}(q^2,p^2;k)=(b_{q^2;\,l_1m_1},[1+N(k)]a_{p^2;\,l_2m_2}). \tag{31}$$

Assume that k is such that $N(k)$ exists. Let $\mathbf{p}$ and $\mathbf{q}$ be such that (25) and (30) hold. Then $f(\mathbf{q},\mathbf{p};k)$ can be expanded into the absolutely convergent series

$$f(\mathbf{q},\mathbf{p};k)=\sum_{l_1m_1}\sum_{l_2m_2} f_{m_1m_2}^{(l_1l_2)}(q^2,p^2;k) \times y_{l_1m_1}^*(-\mathbf{q})y_{l_2m_2}^*(\mathbf{p}), \tag{32}$$

which will be called the multipole expansion.

Let R be a real three-dimensional rotation. Define a potential $\bar{V}(\mathbf{x})$ by

$$\bar{V}(\mathbf{x})=V(R^{-1}\mathbf{x}). \tag{33}$$

A direct computation shows that the functions $\bar{f}_{m_1m_2}^{(l_1l_2)}$ corresponding to the "rotated" potential are related to the original functions by

$$\bar{f}_{\bar{m}_1\bar{m}_2}^{(l_1l_2)}(q^2,p^2;k)=\sum_{m_1}\sum_{m_2}\sum_{i} D_{\bar{m}_1m_1}^{(l_1)*}(R) \times D_{\bar{m}_2m_2}^{(l_2)*}(R)f_{m_1m_2}^{(l_1l_2)}(q^2,p^2;k), \tag{34}$$

where $D_{\bar{m}m}^{(l)}(R)$ is defined as in reference 3. For fixed k, the functions $f_{m_1m_2}^{(l_1l_2)}$ are holomorphic in

$$|\mathrm{Im}(p^2)^{\frac{1}{2}}|<\tfrac{1}{2}\alpha \quad |\mathrm{Im}(q^2)^{\frac{1}{2}}|<\tfrac{1}{2}\alpha. \tag{35}$$

6. SYMMETRIES OF THE POTENTIAL

Let G be an arbitrary subgroup of the group $SO(3)$ of all proper real rotations in three dimensions.

The potential V is said to be invariant under G if

$$V(R\mathbf{x})=V(\mathbf{x})$$

for every $R\in G$.

Denote by $T^{(l_1l_2)}(\mathbf{q},\mathbf{p})$ any function which has the following properties:

(a) It is a homogeneous polynomial of order l_1 (resp. l_2) in the components of the complex vector $\mathbf{q}$ (resp. $\mathbf{p}$).

(b) It satisfies the three-dimensional Laplace equation in $\mathbf{q}$ and in $\mathbf{p}$.

(c) It satisfies

$$T^{(l_1l_2)}(R\mathbf{q},R\mathbf{p})=T^{(l_1l_2)}(\mathbf{q},\mathbf{p})$$

for every $R\in G$.

The functions $T^{(l_1l_2)}(\mathbf{q},\mathbf{p})$ are linear combinations of products $y_{l_1m_1}(q)y_{l_2m_2}(p)$ of solid harmonics. Chose $S\geqslant 0$ functions $T_s^{(l_1l_2)}(\mathbf{q},\mathbf{p})$ which span the vector space

of functions that satisfy (a), (b), and (c).[4] For example, if G is the whole rotation group, then

$$T^{(l_1l_2)}(\mathbf{q},\mathbf{p})=\sum_{m_1}\sum_{m_2}(l_1m_1l_2m_2|l_1l_200)y_{l_1m_1}(\mathbf{q})y_{l_2m_2}(\mathbf{p})$$
$$=\delta_{l_1l_2}\delta_{m_1-m_2}(-1)^{l_1-m_1}(2l_1+1)^{-\frac{1}{2}}.$$

The index s is here superfluous, since $S=\delta_{l_1l_2}$.

If G is the group of all rotations about the z axis, then

$$T_s^{(l_1l_2)}(\mathbf{q},\mathbf{p})=\sum_{m_1}\sum_{m_2}(l_1m_1l_2m_2|l_1l_2s0)y_{l_1m_1}(\mathbf{q})y_{l_2m_2}(\mathbf{p})$$

and $S=2l+1$, where l is the smaller of numbers l_1, l_2.

The functions $T_s^{(l_1l_2)}(\mathbf{q},\mathbf{p})$ can also be calculated for the case that G is a finite group of rotations.

Theorem 5: If the potential V is invariant under G, if the conditions (25) *and* (30) *are satisfied, and if k is such that $N(k)$ exists, then the amplitude f can be expanded into the absolutely convergent series*

$$f(\mathbf{q},\mathbf{p};k)=\sum_{l_1=0}^{\infty}\sum_{l_2=0}^{\infty}\sum_{s=1}^{S}F_s^{(l_1l_2)}(q^2,p^2;k)$$
$$\times T_s^{(l_1l_2)*}(-\mathbf{q},\mathbf{p}),\quad(36)$$

where the functions $F_s^{(l_1l_2)}$ are holomorphic in the domain (35).

The assertion follows from (32) and (34) by an application of Schur's lemma.

If V is spherically symmetric and if $\mathbf{p}$ and $\mathbf{q}$ are real, then the expansion (36) reduces to the form

$$f(\mathbf{q},\mathbf{p};k)=\sum_{l=0}^{\infty}F^{(l)}(q^2,p^2;k)P_l(\xi),$$

where P_l is the Legendre polynomial and ξ is the cosine of the angle between $\mathbf{p}$ and $\mathbf{q}$.

7. MULTIPLE SCATTERING

Consider now again the general case (no symmetries assumed). Let the potential V be a sum

$$V(\mathbf{x})=\sum_{i=1}^{n}V_i(\mathbf{x}).$$

Assume that the supports of the functions V_i do not overlap. That is: If $\mathbf{x}$ is such that $V_i(\mathbf{x})\neq0$ and if $j\neq i$, then $V_j(\mathbf{x})=0$.

Assume also that, for $i=1,\cdots,n$, the potentials V_i satisfy the conditions (I.1) and (I.2). Define:

$$s_i(\mathbf{x})=\operatorname{sgn}V_i(\mathbf{x}),$$
$$a_p^{(i)}(\mathbf{x})=(2\pi)^{-\frac{3}{2}}s_i(\mathbf{x})|V_i(\mathbf{x})|^{\frac{1}{2}}e^{i\mathbf{p}\cdot\mathbf{x}},$$
$$b_q^{(i)}(\mathbf{x})=(2\pi)^{-\frac{3}{2}}|V_i(\mathbf{x})|^{\frac{1}{2}}\exp(-i\mathbf{q}^*\cdot\mathbf{x}),$$
$$A_{ij}(\mathbf{x},\mathbf{y};k)=(4\pi)^{-1}s_i(\mathbf{x})|V_i(\mathbf{x})V_j(\mathbf{y})|^{\frac{1}{2}}e^{ik|\mathbf{x}-\mathbf{y}|}|\mathbf{x}-\mathbf{y}|^{-1}.$$

[4] Compare D. L. Falkoff and G. E. Uhlenbeck, Phys. Rev. 79, 323 (1950), Appendix I.

It is easy to show that

$$a_p=\sum_{i=1}^{n}a_p^{(i)}\quad b_q=\sum_{i=1}^{n}b_q^{(i)}$$
$$A(k)=\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}(k).$$

If $j\neq s$, then

$$(a_p^{(j)},a_q^{(s)})=(b_q^{(j)},a_p^{(s)})=(b_p^{(j)},b_q^{(s)})=0,$$
$$A_{ij}(k)A_{st}(k')=0,$$
$$A_{ij}(k)a_p^{(s)}=A_{ij}(k)b_q^{(s)}=0$$

for all k, $k'\mathbf{p}$ and $\mathbf{q}$.

Define $N_j(k)$ by

$$N_j(k)=-1+[1-A_{jj}(k)]^{-1}$$

and $N(k)$ by

$$N(k)=-1+[1-A(k)]^{-1}.$$

Then, for $j\neq s$

$$N_j(k)a_p^{(s)}=N_j(k)A_{st}(k)=N_j(k)N_s(k)$$
$$=N_j^+(k)b_q^{(s)}=0.$$

The operators $N_j(k)$ describe "scattering by the individual potentials $V_j(x)$." In order to describe the "interference effects," introduce the operators

$$B_s(k)=\sum_{i}^{(i\neq s)}A_{si}(k),$$

$$H(k)=\sum_{s=1}^{n}[1+N_s(k)]B_s(k),$$

and

$$E(k)=-1+[1-H(k)]^{-1}.$$

Theorem 6: If k is such that, for $i=1,\cdots n$, the operators $N_i(k)$ exist and that the operator $E(k)$ exists, then the amplitude f corresponding to the potential (37) *is*

$$f(\mathbf{q},\mathbf{p};k)=\sum_{i=1}^{n}f_i(\mathbf{q},\mathbf{p};k)+\sum_{i=1}^{n}\sum_{j=1}^{n}f_{ij}(\mathbf{q},\mathbf{p};k),\quad(38)$$

where

$$f_i(\mathbf{q},\mathbf{p};k)=(b_q^{(i)},[1+N_i(k)]a_p^{(i)})\quad(39)$$

and

$$f_{ij}(\mathbf{q},\mathbf{p};k)=(b_{-q}^{(i)},[1+N_i(k)]E(k)a_p^{(j)}),$$
$$(i,j=1,2,\cdots n).\quad(40)$$

Proof: A straightforward computation shows that

$$(1-A)=(1-A_{11})(1-A_{22})\cdots(1-A_{nn})(1-H),$$

so that

$$(1-A)^{-1}=1+N=(1+N_n)(1+N_{n-1})\cdots(1+N_1)$$
$$+E(1+N_n)(1+N_{n-1})\cdots(1+N_1).$$

Notice that

$$[1+N_n(k)]\cdots[1+N_1(k)]a_p^{(j)}=[1+N_j(k)]a_p^{(j)}.$$

The assertion then follows from (I.20).

The total amplitude f can thus be computed by solving the scattering problem for the individual potential V_i and for the "interference operator" $H(k)$. Both calculations can be made with the techniques of Sec. 2.

A bound on the magnitude of the "interference terms" is given next as an illustration.

Assume that the distance between the supports of V_j and $V_s(j\neq s)$ is $r_{js}>0$. Introduce the functions

$$I_i(\gamma)=\int |V_i(\mathbf{x})|e^{\gamma|\mathbf{x}|}d\mathbf{x}, \quad (\gamma<\alpha), \quad (i=1,\cdots n),$$

and the functions

$$J_{ij}''(\gamma)=\int\int |V_i(\mathbf{x})V_j(\mathbf{y})| \times\exp\{\gamma|\mathbf{x}-\mathbf{y}|\}d\mathbf{x}d\mathbf{y} \quad (\gamma<\alpha).$$

Notice that, for $j\neq s$,

$$\|A_{js}(k)\|^2\leqslant(4\pi)^{-2}r_{js}^{-2}J_{js}''(-2\ \mathrm{Im}k)$$

and

$$\|B_s(k)\|\leqslant(4\pi)^{-1}\sum_j^{(j\neq s)} r_{sj}^{-1}[J_{sj}''(-2\ \mathrm{Im}\ k)]^{\frac{1}{2}}.$$

Consequently,

$$\|H(k)\|\leqslant(4\pi)^{-1}\sum_s\sum_j^{(j\neq s)}\|[1+N_s(k)]\|r_{sj}^{-1} \times[J_{sj}''(-2\ \mathrm{Im}\ k)]^{\frac{1}{2}}.$$

So:

Theorem 7: *If k is such that*

$$(4\pi)^{-1}\sum_s\sum_j^{(j\neq s)}\|[1+N_s(k)]\|r_{sj}^{-1} \times[J_{sj}''(-2\ \mathrm{Im}\ k)]^{\frac{1}{2}}\leqslant\delta<1,$$

then

$$|f_{ij}(\mathbf{q},\mathbf{p};k)|\leqslant(2\pi)^{-3}[I_i(2|\mathrm{Im}\mathbf{q}|)I_j(2|\mathrm{Im}\mathbf{p}|)]^{\frac{1}{2}} \times\|[1+N_j(\mathrm{k})]\|\delta/(1-\delta).$$

A combination of the methods of this section and of Sec. 3 is applicable to the case when there is overlap between the supports of the potentials V_i (e.g., scattering by several Yukawa potentials).

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor V. Bargmann, Professor S. S. Schweber, and A. C. Manoharan for many valuable discussions.

Theory of Multipole Radiation

S. C. SNOWDON
Midwestern Universities Research Association, Madison, Wisconsin*
(Received March 16, 1961)

The decomposition of the electromagnetic field into longitudinal, transverse electric, and transverse magnetic field types is examined in relation to a similar decomposition of the sources. In general, three independent scalar aspects of the current and charge densities must be specified to provide longitudinal and transverse electromagnetic fields. If, in addition, it is required that the sources be continuous and spatially localized functions, only two independent scalar aspects of the sources are needed to provide the electromagnetic fields outside the sources.

I. INTRODUCTION

In recent years, several authors have given rigorous treatments of multipole radiation from a classical electromagnetic system. Blatt and Weisskopf,[1] using an eigenpole expansion of the electromagnetic field introduced by Franz,[2] have expanded the current density and the curl of the current density to yield the electric and magnetic multipole moments. Wallace,[3] using a slightly different method, has also given expressions for the multipole moments. Rose[4] has given a detailed presentation of the derivation of these same multipole moments as part of a summary of electromagnetic interactions. French and Shimamoto[5] have demonstrated that the curl and divergence of the current density are not independent quantities if the source is localized in a finite region, thereby giving an explanation of the fact that the general electric multipole moment expression is given in terms of the $\nabla\times\mathbf{J}$; whereas the same expression in the long-wave limit involves the $\nabla\cdot\mathbf{J}$.

The purpose of this note is to use a method of vector decomposition employing scalar functions for examining further this boundedness restriction in association with the equation of continuity. The equation of continuity

$$\nabla\cdot\mathbf{J}-ik\rho=0, \tag{1}$$

where $k=\omega/c$, may be looked upon as a scalar restriction on the otherwise arbitrary choices of charge density and the three components of the current density. This restriction together with the boundedness restriction are to be studied below.

II. SOURCES OF ELECTROMAGNETIC FIELD

It is well known that a vector may be decomposed into longitudinal and transverse parts.[6] For the purpose at hand, it is expedient to choose this decomposition in the form

$$\mathbf{J}=-ik\nabla W+\mathfrak{L}\sigma-ik\nabla\times\mathfrak{L}T, \tag{2}$$

where the factor ik is introduced for later convenience. The operator $\mathfrak{L}$ is chosen as[7]

$$\mathfrak{L}=\mathbf{r}\times\nabla \tag{3}$$

with $\mathbf{r}$ as the radius vector since, ultimately, spherical coordinates will be employed. This form puts in evidence the fact that the transverse part of $\mathbf{J}$ may be derived from only two scalar functions,[8] σ and T. Substituting Eq. (2) into Eq. (1) gives

$$\nabla^2W=-\rho. \tag{4}$$

In general, the solution of Eq. (4) for a continuous and localized source function ρ gives W as a field quantity existing outside the source region. Equation (2) then predicts a longitudinal current flow outside the region of charge density. Whereas this may be an acceptable condition if the current carriers are of equal and opposite sign, giving zero charge density outside the region, it is not an acceptable condition in the absence of charge carriers in the space outside the specified source region. Therefore, if the current density is to be localized, the transverse current must cancel the longitudinal current outside of the source region.[5]

In order to ascertain which of the quantities in Eq. (2) are to have a field-like nature and which a source-like or bounded behavior, evaluate $\nabla\cdot\mathbf{J}$, $\mathfrak{L}\cdot\mathbf{J}$, and $(\mathfrak{L}\times\nabla)\cdot\mathbf{J}$:

$$\nabla\cdot\mathbf{J}=-ik\nabla^2W, \tag{5}$$

$$\mathfrak{L}\cdot\mathbf{J}=\mathfrak{L}^2\sigma, \tag{6}$$

and

$$(\mathfrak{L}\times\nabla)\cdot\mathbf{J}=-ik\mathfrak{L}^2\nabla^2T. \tag{7}$$

The condition that $\mathbf{J}$ be bounded requires, for Eq. (5), that ∇^2W or ρ be a source-like quantity as already anticipated. For Eq. (6), this condition requires that σ be a source-like quantity since the operator $\mathfrak{L}^2$ acts only on the transverse coordinates. In a similar manner, Eq. (7) requires that ∇^2T be a source-like quantity. In

* Supported by the U. S. Atomic Energy Commission.

[1] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), p. 796.

[2] W. Franz, Z. Physik **127**, 363 (1950).

[3] P. R. Wallace, Can. J. Phys. **29**, 393 (1951).

[4] M. E. Rose, *Multipole Fields* (John Wiley & Sons, Inc., New York, 1955).

[5] J. G. French and Y. Shimamoto, Phys. Rev. **91**, 898 (1953).

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1763.

[7] W. Franz, P. R. Wallace, and J. M. Blatt and V. F. Weisskopf, (references 1–3), each use the operator $\mathfrak{L}$ although only in connection with the decomposition of the fields and not the sources.

[8] W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1939), p. 261.

analogy with Eq. (4), let

$$\nabla^2 T = -\tau, \tag{8}$$

where τ is a source-like quantity. Of the three scalars in Eq. (2), W and T are field-like quantities; whereas σ has a source-like behavior.

The restriction of the charge and current density to a region of finite extent clearly requires that the sources ρ and τ be related in such a manner that the combined field-like behavior of W and T vanish outside the source region. To effect this cancellation, observe that Eq. (2) may also be written as

$$\mathbf{J} = -ik\nabla[W-(1+\mathbf{r}\cdot\nabla)T]+ik\tau\mathbf{r}+\mathfrak{L}\sigma. \tag{9}$$

The source τ must be adjusted so that its associated field T or rather $(1+\mathbf{r}\cdot\nabla)T$ cancels the field W outside of the source. In terms of the static Green's function, $G_0 = |\mathbf{r}-\mathbf{r}'|^{-1}$, it is possible to write

$$(1+\mathbf{r}\cdot\nabla)T = -(4\pi)^{-1}\int \tau(\mathbf{r}')(\mathbf{r}'\cdot\nabla')G_0(\mathbf{r},\mathbf{r}')d\mathbf{r}'. \tag{10}$$

Similarly, Eq. (4) may be integrated to give

$$W = (4\pi)^{-1}\int \rho(\mathbf{r}')G_0(\mathbf{r},\mathbf{r}')d\mathbf{r}'. \tag{11}$$

The condition that the current density be localized may be expressed as a relation between the static multipole moments of τ and ρ. In this context, a static multipole moment is understood to refer to any of the moments obtained from Eq. (11) and from the corresponding integral equivalent of Eq. (8) in which the static Green's function is employed to connect the source and field quantities. The static multipole moments of the charge density are defined as

$$Q_{LM}^{(0)} = \int r^L \rho(\mathbf{r}) Y_{LM}^*(\theta,\varphi)d\mathbf{r}, \tag{12}$$

where Y_{LM} is the normalized spherical harmonic.[1] If the condition $W-(1+\mathbf{r}\cdot\nabla)T=0$ obtains outside the source region, Eqs. (10) and (11) require that the static multipole moment of τ be

$$\int r^L \tau(\mathbf{r}) Y_{LM}^*(\theta,\varphi)d\mathbf{r} = -L^{-1}Q_{LM}^{(0)}. \tag{13}$$

III. ELECTROMAGNETIC FIELDS

In the absence of material media, Maxwell's equations in Gaussian units are

$$\nabla\cdot\mathbf{B}=0, \tag{14}$$

$$\nabla\times\mathbf{E}-ik\mathbf{B}=0, \tag{15}$$

$$\nabla\cdot\mathbf{E}=4\pi\rho, \tag{16}$$

and

$$\nabla\times\mathbf{B}+ik\mathbf{E}=4\pi\mathbf{J}, \tag{17}$$

where the time variation of the fields and sources has been taken as $e^{-i\omega t}$. The vector fields may be decomposed in a manner similar to that employed for the current density in Eq. (2). Thus,

$$\mathbf{B} = -ik\mathfrak{L}U_E+\nabla\times\mathfrak{L}U_M, \tag{18}$$

and

$$\mathbf{E} = -4\pi\nabla W+ik\mathfrak{L}R+\nabla\times\mathfrak{L}S. \tag{19}$$

If Eqs. (18) and (19) are substituted into Eqs. (15)–(17) and Eq. (2) is used for the current density, the following relations can be chosen:

$$R-U_M=0, \tag{20}$$

$$\nabla^2 S+k^2U_E=0, \tag{21}$$

$$\nabla^2 U_M+k^2R+4\pi\sigma=0, \tag{22}$$

and

$$S+4\pi T-U_E=0. \tag{23}$$

Actually, the vector independence of the terms in the decomposition of $\mathbf{B}$, $\mathbf{E}$, and $\mathbf{J}$ gives, for each of the right-hand members in Eqs. (20)–(23), a function of the longitudinal coordinate. On the other hand, $\mathfrak{L}$ (function of longitudinal coordinate)$=0$, thereby reducing to zero the contribution of this function to the fields. By eliminating R and S in Eqs. (20)–(23) and using Eq. (8), the following expressions may be found for the fields.

$$\mathbf{B} = -ik\mathfrak{L}U_E+\nabla\times\mathfrak{L}U_M \tag{24}$$

$$\mathbf{E} = -4\pi\nabla[W-(1+\mathbf{r}\cdot\nabla)T] +4\pi\tau\mathbf{r}+ik\mathfrak{L}U_M+\nabla\times\mathfrak{L}U_E, \tag{25}$$

where

$$\nabla^2 U_M+k^2U_M=-4\pi\sigma, \tag{26}$$

and

$$\nabla^2 U_E+k^2U_E=-4\pi\tau. \tag{27}$$

Equations (4) and (8) thus determine the static fields W and T from the scalar sources ρ and τ; whereas Eqs. (26) and (27) determine the dynamic fields U_M and U_E from the sources σ and τ. The independent specification of ρ and τ is not permitted, if the sources are to be bounded; their static moments are related as in Eq. (13).

Outside the source distributions, the electromagnetic fields of Eqs. (24) and (25) are determined by the two scalar fields U_M and U_E:

$$\mathbf{B} = -ik\mathfrak{L}U_E+\nabla\times\mathfrak{L}U_M, \tag{28}$$

and

$$\mathbf{E} = ik\mathfrak{L}U_M+\nabla\times\mathfrak{L}U_E. \tag{29}$$

By inference, the dynamic multipole moments of σ and τ completely determine the external electromagnetic field.

IV. MULTIPOLE EXPANSIONS

The solutions of Eqs. (26) and (27) in the region outside the source may be expanded in spherical polar coordinates as[5,9]

$$U_M=4\pi ik\sum_{L=0}^{\infty}\sum_{M=-L}^{L}G_L{}^M(\mathbf{r})\int\sigma(\mathbf{r}')g_L{}^M(\mathbf{r}')d\mathbf{r}', \quad (30)$$

and

$$U_E=4\pi ik\sum_{L=0}^{\infty}\sum_{M=-L}^{L}G_L{}^M(\mathbf{r})\int\tau(\mathbf{r}')g_L{}^M(\mathbf{r}')d\mathbf{r}', \quad (31)$$

where $G_L{}^M(\mathbf{r})=h_L{}^{(1)}(kr)Y_{LM}(\theta,\varphi)$ and $g_L{}^M(\mathbf{r}')=j_L(kr')\times Y_{LM}{}^*(\theta',\varphi')$. The integrals in Eqs. (30) and (31) are proportional to the multipole moments of the source distributions σ and τ, a proportionality factor being chosen so as to simplify the moments in the long wavelength limit ($k\rightarrow 0$). Thus, the electric multipole moments are defined as

$$Q_{LM}=-\frac{(2L+1)!}{2^L(L-1)!k^L}\int\tau(\mathbf{r})g_L{}^M(\mathbf{r})d\mathbf{r}, \quad (32)$$

and the magnetic multipole moments as

$$M_{LM}=-\frac{(2L+1)!}{2^L(L-1)!k^L}\int\sigma(\mathbf{r})g_L{}^M(\mathbf{r})d\mathbf{r}. \quad (33)$$

These definitions are seen to be equivalent to the definitions

$$Q_{LM}=\frac{i(2L+1)!}{2^L(L+1)!k^{L+1}}\int[(\mathfrak{L}\times\nabla)\cdot\mathbf{J}]g_L{}^M(\mathbf{r})d\mathbf{r}, \quad (34)$$

and

$$M_{LM}=\frac{(2L+1)!}{2^L(L+1)!k^L}\int(\mathfrak{L}\cdot\mathbf{J})g_L{}^M(\mathbf{r})d\mathbf{r}, \quad (35)$$

if for $\mathbf{J}$ one makes use of Eq. (9) and the relations: $\nabla^2g_L{}^M+k^2g_L{}^M=0$, $\mathfrak{L}^2g_L{}^M=-L(L+1)g_L{}^M$, and $\int f\mathfrak{L}^2gd\mathbf{r}=\int g\mathfrak{L}^2fd\mathbf{r}$ for bounded f. Therefore, the multipole moments of τ (electric moments) and the multipole moments of σ (magnetic moments) serve to determine completely the electromagnetic fields outside of the source region.

In the long wavelength limit ($k\rightarrow 0$), the multipole moments become

$$Q_{LM}=-L\int r^L\tau(\mathbf{r})Y_{LM}{}^*(\theta,\varphi)d\mathbf{r}=Q_{LM}{}^{(0)}, \quad (36)$$

and

$$M_{LM}=-L\int r^L\sigma(\mathbf{r})Y_{LM}{}^*(\theta,\varphi)d\mathbf{r}, \quad (37)$$

where it will be noted that the moments of the τ distribution are related to the moments of the charge distribution ρ in the manner exhibited by Eq. (13).

[9] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 399.

TABLE I. Relation of scalar fields to scalar sources.

Fields	Source density	Dynamic field	Static field	Dynamic moment	Static moment
Londitudinal	ρ		W		$Q_{LM}{}^{(0)}$
Transverse electric (magnetic type)	σ	U_M		M_{LM}	
Transverse magnetic (electric type)	τ	U_E	T	Q_{LM}	$-L^{-1}Q_{LM}{}^{(0)}$

Table I presents a summary of the relations among the various scalar quantities in the preceding development. The source densities (ρ,σ,τ) are of central interest leading to both static and dynamic field quantities. The net effect of the static fields is felt only within the source region; whereas, the dynamic fields give rise to radiation from the source. By adjusting the static fields so that their effects cancel outside the source region, a connection is obtained between the charge density ρ and the density τ.

IV. MAGNETIZATION—POLARIZATION SOURCES

Electromagnetic phenomena in the presence of material media are accounted for by dividing the current and charge densities into two types: the free or accessible sources and the bound or inaccessible sources.[10] The inaccessible sources are written in terms of source quantities $\mathbf{M}$ and $\mathbf{P}$ according to

$$\mathbf{J}_{\text{bound}}=-ik\mathbf{P}+\nabla\times\mathbf{M}, \quad (38)$$

and

$$\rho_{\text{bound}}=-\nabla\cdot\mathbf{P}, \quad (39)$$

where $\mathbf{P}$ and $\mathbf{M}$ are interpreted in terms of polarization and magnetization per unit volume. Additional contributions to the multipole moments are then given by substituting Eq. (38) into Eqs. (34) and (35) for the current density. The contribution of $\mathbf{P}$ to the electric moments is

$$Q_{LM}{}'=\frac{(2L+1)!}{2^L(L+1)!k^L}\int[(\mathfrak{L}\times\nabla)\cdot\mathbf{P}]g_L{}^M(\mathbf{r})d\mathbf{r}, \quad (40)$$

and the contribution of $\mathbf{M}$ to the electric moment is

$$Q_{LM}{}''=\frac{i(2L+1)!}{2^L(L+1)!k^{L-1}}\int(\mathfrak{L}\cdot\mathbf{M})g_L{}^M(\mathbf{r})d\mathbf{r}. \quad (41)$$

In the case of the magnetic moments, $\mathbf{M}$ contributes

$$M_{LM}{}'=\frac{(2L+1)!}{2^L(L+1)!k^L}\int[(\mathfrak{L}\times\nabla)\cdot\mathbf{M}]g_L{}^M(\mathbf{r})d\mathbf{r}, \quad (42)$$

and $\mathbf{P}$ contributes

$$M_{LM}{}''=-\frac{i(2L+1)!}{2^L(L+1)!k^{L-1}}\int(\mathfrak{L}\cdot\mathbf{P})g_L{}^M(\mathbf{r})d\mathbf{r}. \quad (43)$$

[10] W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955), Chap. 22.

In the long-wave limit $k \rightarrow 0$, expansion of $g_L{}^M$ in powers of kr yields

$$Q_{LM}' = \frac{1}{L+1} \int r^L Y_{LM}^*(\theta,\varphi)[(\mathfrak{L}\times\nabla)\cdot\mathbf{P}]d\mathbf{r}, \quad (44)$$

$$Q_{LM}'' = \frac{ik}{L+1} \int r^L Y_{LM}^*(\theta,\varphi)(\mathfrak{L}\cdot\mathbf{M})d\mathbf{r}, \quad (45)$$

$$M_{LM}' = \frac{1}{L+1} \int r^L Y_{LM}^*(\theta,\varphi)[(\mathfrak{L}\times\nabla)\cdot\mathbf{M}]d\mathbf{r}, \quad (46)$$

and

$$M_{LM}'' = -\frac{ik}{L+1} \int r^L Y_{LM}(\theta,\varphi)(\mathfrak{L}\cdot\mathbf{P})d\mathbf{r}. \quad (47)$$

If the polarization is decomposed into independent vector contributions in a manner similar to that used for the current density, the requirement that $\mathbf{P}$ have a source-like behavior connects the static multipole moments of $(\mathfrak{L}\times\nabla)\cdot\mathbf{P}$ with the static moments of $\nabla\cdot\mathbf{P}$. The arguments parallel those presented in Sec. II for the free current density. Since the Q_{LM}' in Eq. (44) are static moments, the boundedness argument yields

$$Q_{LM}' = -\int r^L Y_{LM}^*(\theta,\varphi)\nabla\cdot\mathbf{P}d\mathbf{r}. \quad (48)$$

Following the same argument, the static multipole moments of $(\mathfrak{L}\times\nabla)\cdot\mathbf{M}$ are related to the static moments of $\nabla\cdot\mathbf{M}$. Since the M_{LM}' in Eq. (46) are static moments, the boundedness argument gives

$$M_{LM}' = -\int r^L Y_{LM}^*(\theta,\varphi)\nabla\cdot\mathbf{M}d\mathbf{r}. \quad (49)$$

Thus, the expressions for the multipole moments are seen to be identical with those given by Blatt and Weisskopf[1] if one observes that their use of current density is to be interpreted as the partial current density $-ik\mathbf{P}$, the effects of $\nabla\times\mathbf{M}$ being introduced through the use of the field $\mathbf{H}$ instead of $\mathbf{B}$.

Molecular Collisions. III. Symmetric Top Molecules

GEORGE GIOUMOUSIS
Shell Development Company, Emeryville, California

(Received April 11, 1961)

The question of rotational energy transfer in molecular collisions is treated for the case of symmetric top molecules. The Schrödinger equation for this problem, which can be taken as an integral equation over nine variables, is reduced to a set of coupled one dimensional integral equations by means of expansions over the irreducible representations of the three-dimensional rotation group. Various selection rules are found, of which two, in indices (in the matrix elements of the potential) related to the quantum number k, have no counterpart in the previously developed theory for diatomic molecules.

The results are exhibited as cross sections for the excitation of rotational states, given the states before the collision. The rotational states for the symmetric top rotors are specified by the usual quantum numbers k, l, m, and the translational states by the velocity and direction of motion. The angular dependence of the cross section is treated by means of an expansion in spherical harmonics which is new in this context and which promises to be extremely useful in the application to transport processes.

A FORMAL theory of inelastic molecular collisions has previously been developed,[1] and, at first glance, one would assume that it had been developed to its logical limit as a perfectly general theory. However, for rotating molecules, the dynamics can be developed in a form which exhibits a coupling between the rotation of the molecule and the rotation of one molecule relative to the other. This coupling is conveniently expressed in terms of the matrix representations of the three dimensional rotation group, as a result of which the subsequent analysis is greatly simplified and various selection rules appear.

A previous paper[2] applied the group theoretical formalism to the collisions of rotating and vibrating linear molecules. The part concerning the vibrational motion was carried along only for completeness and, in fact, was only an amplified restatement of the relevant parts of the original work. However, the rotational motion of linear molecules is a degenerate case of that of nonlinear molecules, so that it cannot serve as even a qualitative model for the latter. The case of symmetric top molecules was chosen for this work since, on the one hand, it is essentially of the order of complexity of the case of general nonlinear molecules; while on the other hand, the rotational wave functions are still expressible in terms of the group representations. The treatment of the vibrational part is to be understood as being exactly analogous to that in the corresponding parts of the previous work.[1,2]

The treatment here parallels that of the previous work very closely. For the sake of brevity, these two previous papers will be denoted by I and II, respectively. Equations will be similarly denoted, so that, for example, the first equation of the third section in the reference will be Eq. (II 3.1).

1. COORDINATES

As in II, the coordinates used here are based on the three-dimensional rotation group. It must be stressed that the group as used here is a point, rather than a coordinate, transformation group, so that an element of group R rotates the whole space, carrying material bodies along, and the x, y, and z axes remain fixed. A basic reference is Wigner.[3] For the sake of the printer, the group representations are denoted by

$$D(jmn|R) \text{ rather than } D^j(R)_{mn}$$

and the Clebsch-Gordon coefficients by

$$C(jj'Jnn') \text{ rather than } S_{Jnn'}{}^{jj'}.$$

The orientation of a molecule is properly measured from the configuration with the vibrational coordinates at their equilibrium values.[4] Assuming this, a symmetric top molecule is one in which two of the principal moments of inertia (I_1 and I_2, say) are equal. The orientation of the molecule will be specified by the rotation which places the unequal principal moment parallel to the z axis and the other two parallel to the x and y axes, respectively. A certain arbitrariness is needed with respect to the two equal moments, which, however, can be deferred until a special case is under consideration, since all choices are equally valid. If this rotation which places the molecule "a" in a standard configuration be denoted by R^a, then the wave function of the symmetric top is[3]

$$(8\pi^2)^{-\frac{1}{2}}(2l^a+1)^{\frac{1}{2}}D(l^ak^am^a|R^a), \tag{1.1}$$

with energy

$$\frac{1}{2I_1}l^a(l^a+1)\hbar^2+\tfrac{1}{2}\left(\frac{1}{I_3}-\frac{1}{I_1}\right)(k^a)^2\hbar^2. \tag{1.2}$$

The wave function for the molecule "b" is similarly written.

As in II, let R be the rotation which places the vector from the center of mass of "a" to the center of mass of "b" parallel to the z axis. Since the first Eulerian angle is undefined by the above, it is convenient to define it by some condition on molecule "a." The exact condition is of little importance, but for definiteness

[1] G. Gioumousis and C. F. Curtiss, J. Chem. Phys. 29, 996 (1958).
[2] G. Gioumousis and C. F. Curtiss, J. Math. Phys. 2, 96 (1961).
[3] E. Wigner, *Group Theory* (Academic Press, Inc., New York, 1959).
[4] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill Book Company, Inc., 1955), p. 11.

it will be taken that the third principal moment of "a" be parallel to the first-third quadrant half of the xz plane.

Again, as in II, the vectors S^a and S^b are defined by first performing the rotation R on the whole system, then S^a and S^b in such a way as to put "a" and "b," respectively, in their standard configurations. As before, this definition is equivalent to

$$R^a=S^aR \quad \text{and} \quad R^b=S^bR.$$

Similarly, rotations with first Eulerian angle zero may be used to describe directions, specifically the direction of the velocity of "b" relative to that of "a." If such rotations be denoted by T, then the desired cross sections will have the functional form

$$\sigma(l^ak^am^a; l^bk^bm^b; T|E|l'^ak'^am'^a; l'^bk'^bm'^b; T'),$$

where the primed variables refer to the values after the collision, and E is the total energy in relative coordinates.

2. EXPANSION IN THE REPRESENTATION COEFFICIENTS

It is evident that the intermolecular potential function V is a function only of the intermolecular distance r, the normal coordinates of "a" and "b" (which, however, are to be neglected in this treatment), and the two rotations S^a and S^b. The dependence on the rotations can be expressed by a series expansion

$$V(rS^aS^b)=\sum_{\lambda_1\lambda_2\nu_1\nu_2\mu_2} v(\lambda_1\lambda_2\nu_1\nu_2\mu_2|r) \times D(\lambda_1\nu_1-\mu_2|S^a)D(\lambda_2\nu_2\mu_2|S^b). \quad (2.1)$$

Only five indices appear in the series because the definition of R fixes the third Eulerian angle of S^a.

It is convenient to couple the three rotating systems which must be considered, the two molecules and the rotation of one molecule relative to the other, into a state of sharp total angular momentum. The rule for composition of angular momenta yields the function

$$I(l^ak^al^bk^bl\lambda LM|RR^aR^b) = \sum_{m^am} C(l\lambda LM, M-m)C(l^al^blm^a, m-m^a) \times D(l^ak^am^a|R^a)D(l^bk^bm-m^a|R^b) \times D(\lambda OM-m|R) \quad (2.2)$$

as a result of coupling "a" and "b" to give a system with total angular momentum l and z component m, and this, in turn, is coupled with the relative rotation to give L and M. Equation (2.2) may be inverted, thus

$$D(l^ak^am^a|R^a)D(l^bk^bm^b|R^b)D(\lambda 0\mu|R) =\sum_{lL} C(l\lambda L, m^a+m^b, \mu)C(l^al^bm^am^b) \times I(l^ak^al^bk^bl\lambda L, m^a+m^b+\mu|RR^aR^b). \quad (2.3)$$

A new function J can be defined which is I expressed in terms of RS^aS^b rather than RR^aR^b, that is

$$J(\cdots|RS^aS^b)=I(\cdots|R, S^aR, S^bR). \quad (2.4)$$

The same method used to derive Eq. (II 2.8) from (II 2.7) is valid here, with result

$$J(l^ak^al^bk^bl\lambda LM|RS^aS^b) =\sum_{st} C(l^al^bst)C(l\lambda L, s+t, 0)D(l^ak^bs|S^a) \times D(l^b0t|S^b)D(L, s+t, M|R). \quad (2.5)$$

3. INTEGRAL EQUATIONS

The integral equation to be solved is of the form[1]

$$\psi=\varphi+G^+V\varphi$$

where ψ is the wave function, φ the initial-state wave function, G^+ the Green's function operator, and V the potential. The initial state is to be one of sharp angular momentum

$$\varphi(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M}) =j(\bar{L}|\bar{k}r)I(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M}|RR^aR^b), \quad (3.1)$$

and only later is a wave function with plane-wave initial state to be built up by linearity.

The operator $G^+=G^+(E)$ may be written[2] as a bilinear series in the internal variables,

$$G^+(E)f(\mathbf{r}R^aR^b) =\int\int\int \sum_{\substack{l^ak^am^a\\ l^bk^bm^b}} \left(\frac{2l^a+1}{8\pi^2}\right) D(l^ak^am^a|R^a) \times D(l^ak^am^a|R'^a)^*\left(\frac{2l^b+1}{8\pi^2}\right) D(l^bk^bm^b|R^b) \times D(l^bk^bm^b|R'^b)^*\left[-\frac{\mu}{2\pi\hbar^2} \times\frac{\exp[ik(l^ak^al^bk^b)|\mathbf{r}-\mathbf{r}'|]}{|\mathbf{r}-\mathbf{r}'|}\right] f(\mathbf{r}'R'^aR'^b) \times d\mathbf{r}'dR'^adR'^b. \quad (3.2)$$

Here the energy dependence of the operator is evidenced through k by the relation

$$\tfrac{1}{2}\mu\hbar^2k(l^ak^al^bk^b)^2=E-E^a(l^ak^a)-E^b(l^bk^b)$$

and in no other way.

Substitution of the series for the exponential factor into Eq. (3.2), and changing the dummy indices $l^ak^am^a \times l^bk^bm^b$ to $l_1k_1m_1l_2k_2m_2$ gives

$$G^+(E)f=\frac{1}{2\pi}\int\cdots\int \sum_{\substack{l_1k_1m_1\\ l_2k_2m_2}} \left(\frac{2l_1+1}{8\pi^2}\right) D(l_1k_1m_1|R^a) \times D(l_1k_1m_1|R'^a)^*\left(\frac{2l_2+1}{8\pi^2}\right) D(l_2k_2m_2|R^b) \times D(l_2k_2m_2|R'^b)^*\left[\left(-\frac{\mu ki}{2\pi\hbar^2}\right)(2l+1) \times j(l|kr_<)h(l|kr_>)(-1)^sD(lOs|R) \times D(lO-s|R')\right] f(r'R'^aR'^b)\mathbf{r}'^2dr'dR'^adR'^b.$$

Substitution of Eq. (2.3) twice into the above then gives

$$G^+(E)f=\frac{1}{2\pi}\int\cdots\int\sum_{\substack{l_1l_2l_3l_4l_5\\Mk_1k_2}}\left(-\frac{\mu ki}{2\pi\hbar^2}\right)\times\frac{(2l_1+1)(2l_2+1)(2l_5+1)}{(8\pi^2)^2}\times I(l_1k_1l_2k_2l_3l_5l_4M\,|\,RR^aR^b)\times I(l_1k_1l_2k_2l_3l_5l_4M\,|\,R'R^aR'^b)^*\times j(l_5|kr_<)h(l_5|kr_>)f(r'R'R'^aR'^b)\times r'^2dr'dR'dR'^adR'^b \quad (3.3)$$

where k denotes $k(l_1k_1l_2k_2)$.

Let the wave function be expanded in terms of the I's as follows

$$\psi=\sum_{\substack{l''^ak''^al''^bk''^b\\l''\lambda''L''M''}}\psi(l''^ak''^al''^bk''^bl''\lambda''L''M''\,|\,r)\times I(l''^ak''^al''^bk''^bl''\lambda''L''M''\,|\,RR^aR^b). \quad (3.4)$$

For brevity, let the set of indicial arguments of I be denoted by Λ. Then the series form of the integral equation becomes

$$\sum_{\Lambda''}\psi(\Lambda''\,|\,r)I(\Lambda''\,|\,RR^aR^b)=j(\bar{L}\,|\,kr)I(\bar{\Lambda}\,|\,RR^aR^b)+\frac{1}{2\pi}\int\cdots\int\sum_{\substack{\Lambda_1\Lambda'\lambda_1\lambda_2\\\nu_1\nu_2\mu_2}}\times\left(-\frac{\mu ki}{2\pi\hbar^2}\right)\frac{(2l_1+1)(2l_2+1)(2l_5+1)}{8\pi^2}\times I(\Lambda_1|RR^aR^b)J(\Lambda_1|R'S'^aS'^b)j(l_5|kr_<)h(l_5|kr_>)\times v(\lambda_1\lambda_2\nu_1\nu_2\mu_2|r')D(\lambda_1\nu_1-\mu_2|S'^a)D(\lambda_2\nu_2\mu_2|S'^b)\times\psi(\Lambda'|r')J(\Lambda'|R'S'^aS'^b)r'^2dr'dR'dS'^adS'^b. \quad (3.5)$$

The justification of the change of variable from $R'^aR'^b$ to $S'^aS'^b$ follows from the basic property of group integration.[2,3]

The functions I can be shown to satisfy the orthogonality relation[3]

$$\int\int\int I(\Lambda|RR^aR^b)I(\Lambda'|RR^aR^b)dRdR^adR^b=\delta(\Lambda,\Lambda')\frac{(8\pi^2)^3}{(2l^a+1)(2l^b+1)(2\lambda+1)}. \quad (3.6)$$

Then, if an integral operator be defined by

$$G(\lambda l^ak^al^bk^b)f(r)=\int j(\lambda\,|\,k(l^ak^al^bk^b)r_<)\times h(\lambda\,|\,k(l^ak^al^bk^b)r_>)r'^2f(r')dr', \quad (3.7)$$

the integral equation becomes

$$\psi(\Lambda|r)=\delta(\Lambda,\bar{\Lambda})j(\bar{\lambda}|\bar{k}r)+\sum_{\Lambda'\lambda_1\lambda_2\nu_1\nu_2\mu_2}\left(-\frac{\mu ki}{4\pi^2\hbar^2}\right)\times\frac{(2l^a+1)(2l^b+1)(2\lambda+1)}{(8\pi^2)^2}\times\left[\int\int\int J(\Lambda\,|\,R'S'^aS'^b)^*D(\lambda_1\nu_1-\mu_2\,|\,S'^a)\times D(\lambda_2\nu_2\mu_2\,|\,S'^b)J(\Lambda'\,|\,R'S'^aS'^b)dR'dS'^adS'^b\right]\times G(\lambda l^ak^al^b)\{v(\lambda_1\lambda_2\nu_1\nu_2\mu_2|r)\psi(\Lambda'\,|\,r)\}. \quad (3.8)$$

It should be noted that since Λ is shorthand for eight indices, the sum above is really over thirteen indices.

The factor within brackets in Eq. (3.7) above may be evaluated by a trivial variation of that following[2] Eq. (II 3.9), with result here

$$[\cdots]=\frac{(8\pi^2)^3}{(2l^a+1)(2l^b+1)(2L+1)}\times\delta(LMk^ak^b;L'M',k'^a+\nu_1,k'^b+\nu_2)\sum_{st}C(l^al^blst)\times C(l\lambda L,s+t,0)C(l'^al'^bl',s+\mu_2,t-\mu_2)\times C(l'\lambda'L,s+t,0)C(l'^a\lambda_1l^a,s+\mu_2,-\mu_2)\times C(l'^b\lambda_2l^b,t-\mu_2,\mu_2)C(l'^a\lambda_1l^ak'^a,k^a-k'^a)\times C(l'^b\lambda_2l^bk'^b,k^b-k'^b). \quad (3.9)$$

Now, if the dependence of the wave function on the initial state be indicated explicitly, and the Λ shorthand dropped, the integral equation [Eq. (3.8)] becomes

$$\psi(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M};l^ak^al^bk^bl\lambda LM\,|\,r)=\delta(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M};l^ak^al^bk^bl\lambda LM)j(\bar{\lambda}\,|\,\bar{k}r)-\frac{2\mu ki}{\hbar^2}\sum_{\substack{l'^ak'^al'^bk'^b\\l'\lambda'\lambda_1\lambda_2\mu_2}}\frac{2\lambda+1}{2L+1}C(l^al^blst)C(l'^al'^bl',s+\mu_2,t-\mu_2)\times C(l\lambda L,s+t,0)C(l'\lambda'L,s+t,0)\times C(l'^a\lambda_1l^a,s+\mu_2,-\mu_2)C(l'^b\lambda_2l^b,t-\mu_2,\mu_2)\times C(l'^a\lambda_1l^ak'^a,k^a-k'^a)C(l'^b\lambda_2l^bk'^b,k^b-k'^b)\times G(\lambda l^ak^al^bk^b)\cdot\{v(\lambda_1\lambda_2,k^a-k'^a,k^b-k'^b,\mu_2|r)\times\psi(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M};l'^ak'^al'^bk'^bl'\lambda'LM\,|\,r|\}, \quad (3.10)$$

where $k=k(l^ak^al^bk^b)$ and $\bar{k}=k(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b)$. The number of indices in the sum is diminished by the delta functions involving the k's, the ν's, and L and M in Eq. (3.9). The former lead to a selection rule in the values of ν_1, ν_2 in $v(\lambda_1\lambda_2\nu_1\nu_2\mu_2)$. The latter lead to a lack of mixing

among the values of L and M in the homogeneous term of the integral equation. Since $\delta_{L\bar{L}}\delta_{M\bar{M}}$ occurs in the inhomogeneous term, unless $M=\bar{M}$ and $L=\bar{L}$, the equation is homogeneous and can have only zero as a solution. Further, since M and $\bar{M}$ do not appear in Eq. (3.10) other than in the wave function ψ, the wave function is not a function of $\bar{M}$ at all. These statements can be expressed thus

$$\psi(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M};\, l^ak^al^bk^bl\lambda LM\,|\,r)$$
$$=\delta_{L\bar{L}}\delta_{M\bar{M}}\psi(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}0;\, l^ak^al^bk^bl\lambda\bar{L}0\,|\,r). \quad (3.11)$$

As a result, the expansion of Eq. (3.4) is over six indices instead of eight, which is the reason for the introduction of initial states with sharp angular rather than linear momentum.

4. CROSS SECTIONS: POLARIZED AND UNPOLARIZED BEAMS

The solution to the integral equation [Eq. (3.10)] can be seen to have an asymptotic form

$$\psi(\bar{l}^a\cdots\bar{L}0;\, l^a\cdots\bar{L}0\,|\,r)-\delta(\bar{l}^a\cdots;\, l^a\cdots)j(\bar{\lambda}\,|\,\bar{k}r)$$
$$\sim f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda};\, l^ak^al^bk^bl\lambda)r^{-1}e^{ikr}, \quad (4.1)$$

where f is a constant with respect to r. This is a solution with initial state as given in Eq. (3.1). What is needed for the cross section is a solution with initial state a plane wave, that is

$$\varphi(\bar{l}^a\bar{k}^a\bar{m}^a\bar{l}^b\bar{k}^b\bar{m}^b;\,\mathbf{k})=e^{i\mathbf{k}\cdot\mathbf{r}}(8\pi^2)^{-1}(2\bar{l}^a+1)^{\frac{1}{2}}(2\bar{l}^b+1)^{\frac{1}{2}}$$
$$\times D(\bar{l}^a\bar{k}^a\bar{m}^a\,|\,R^a)D(\bar{l}^b\bar{k}^b\bar{m}^b\,|\,R^b). \quad (4.2)$$

Now it is evident that the φ of Eq. (4.2) can be expressed as a linear combination of the φ's of Eq. (3.1), and that so can the corresponding solutions ψ of the integral equation. The coefficients of the linear combination are identical to those of Eq. (II 4.3), except for a factor of 2π to adjust the normalization of the rotational functions,

$$d(\bar{l}^a\bar{k}^a\bar{m}^a\bar{l}^b\bar{k}^b\bar{m}^bT;\, \bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M})$$
$$=\frac{[(2\bar{l}^a+1)(2\bar{l}^b+1)]^{\frac{1}{2}}}{8\pi^2}i^{\lambda^*}(2\bar{\lambda}+1)$$
$$\times D(\bar{\lambda}0,\bar{m}^a+\bar{m}^b-M\,|\,T)C(\bar{l}\bar{\lambda}\bar{L},\bar{m}^a+\bar{m}^b,\bar{M}-\bar{m}^a-\bar{m}^b)$$
$$\times C(\bar{l}^a\bar{l}^b\bar{l}\bar{m}^a\bar{m}^b), \quad (4.3)$$

where $\lambda^*=\bar{\lambda}+2(\bar{M}-\bar{m}^a-\bar{m}^b)$ and T is the direction of $\mathbf{k}$.

Equation (4.3) may be combined with Eqs. (4.1), (3.4), and (1.1) to give the following asymptotic form for the wave function with plane-wave initial state

$$f(T\,|\,\bar{l}^a\bar{k}^a\bar{m}^a\bar{l}^b\bar{k}^b\bar{m}^b;\, l^ak^am^al^bk^bm^b\,|\,R)$$
$$\sum_{\bar{l}\bar{\lambda}\bar{L}\bar{M}l\lambda}(8\pi^2)[(2l^a+1)(2l^b+1)]^{-\frac{1}{2}}d(\bar{l}^a\bar{k}^a\bar{m}^a\bar{l}^b\bar{k}^b\bar{m}^bT;\,\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}\bar{L}\bar{M})f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda};\,\bar{L};\,l^ak^al^bk^bl\lambda)$$
$$\times C(l\lambda L,\,m^a+m^b,\,\bar{M}-m^a-m^b)C(l^al^blm^am^b)D(\lambda0,\,\bar{M}-m^a-m^b\,|\,R)$$
$$=\sum_{\bar{l}\bar{\lambda}\bar{L}\bar{M}l\lambda}[(2\bar{l}^a+1)(2\bar{l}^b+1)]^{\frac{1}{2}}[(2l^a+1)(2l^b+1)]^{-\frac{1}{2}}i^{\bar{\lambda}}(-1)^{\bar{M}-\bar{m}^a-\bar{m}^b}(2\bar{\lambda}+1)C(\bar{l}\bar{\lambda}\bar{L},\,\bar{m}^a+\bar{m}^b,\,\bar{M}-\bar{m}^a-\bar{m}^b)$$
$$\times C(l\lambda L,\,m^a+m^b,\,\bar{M}-m^a-m^b)C(\bar{l}^a\bar{l}^b\bar{l}\bar{m}^a\bar{m}^b)C(l^al^blm^am^b)f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda};\,\bar{L};\,l^ak^al^bk^bl\lambda)$$
$$\times D(\bar{\lambda}0,\,\bar{m}^a+\bar{m}^b-\bar{M}\,|\,T)D(\lambda0,\,\bar{M}-m^a-m^b\,|\,R), \quad (4.4)$$

where the sum is over $l\lambda\bar{l}\bar{\lambda}\bar{L}\bar{M}$ and R is the direction of $\mathbf{r}$.

By Eq. (I 4.7) the cross section for scattering from barred states $\bar{l}^a\cdots$, and direction T to unbarred states $l^a,\ \cdots$, and direction R is

$$\sigma(\bar{l}^a\bar{k}^a\bar{m}^a\bar{l}^b\bar{k}^b\bar{m}^b;\,T\,|\,l^ak^am^al^bk^bm^b;\,R)$$
$$=\frac{k}{\bar{k}}|f(T\,|\,\bar{l}^a\cdots;\,l^a\cdots\,|\,R)|^2. \quad (4.5)$$

Let the absolute square be written as the product of f with its complex conjugate, whence (if primed dummy indices are used for the latter) a series over twelve indices results. Precisely the arguments used to derive Eq. (II 4.6) are valid here, with result that the cross section may be expanded in the form

$$\sigma(\bar{l}^a\cdots;\,T\,|\,l^a\cdots;\,R)=\sum_{\lambda_3\lambda_4\mu_4}\sigma(\bar{l}^a\cdots\,|\,\lambda_3\lambda_4\mu_4\,|\,l^a\cdots)$$
$$\times D(\lambda_30-\mu_4\,|\,T)D(\lambda_40\mu_4\,|\,R), \quad (4.6)$$

where the cross-section coefficient has the form

$$\sigma(\bar{l}^a\cdots\,|\,\lambda_3\lambda_4\mu_4\,|\,l^a\cdots)$$
$$=k/\bar{k}\frac{(2\bar{l}^a+1)(2\bar{l}^b+1)}{(2l^a+1)(2l^b+1)}\sum i^{\mu^*}(2\bar{\lambda}'+1)(2\lambda+1)$$
$$\times C(\bar{l}\bar{\lambda}\bar{L},\,\bar{m},\,\mu_3-\bar{m})C(\bar{l}'\bar{\lambda}'L'\bar{m},\,\mu_3-\mu_4-\bar{m})$$
$$\times C(\bar{l}^a\bar{l}^b\bar{l}\bar{m}^a\bar{m}^b)C(\bar{l}^a\bar{l}^b\bar{l}'\bar{m}^a\bar{m}^b)C(l^al^blm^am^b)$$
$$\times C(l^al^bl'm^am^b)C(\lambda\lambda'\lambda_300)C(\lambda\lambda'\lambda_400)$$
$$\times C(\bar{\lambda}\bar{\lambda}'\lambda_3\bar{m}-\mu_3,\,\mu_3-\mu_4-\bar{m})$$
$$\times C(\lambda\lambda'\lambda_4,\,\mu_3-m,\,m+\mu_4-\mu_3)$$
$$\times f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda};\,L;\,l^ak^al^bk^bl\lambda)$$
$$\times f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}'\bar{\lambda}';\,L';\,l^ak^al^bk^bl'\lambda')^*, \quad (4.7)$$

where $\mu^*=\bar{\lambda}-\bar{\lambda}'+2(\mu_4+m+\bar{m})$ and the sum is over $\bar{l}\bar{l}'l'\bar{\lambda}\bar{\lambda}'\lambda'LL'\mu_3$.

The corresponding result for unpolarized beams is very similar to that for diatomic molecules. Comparison of Eq. (4.7) and (II 4.7) would show that the mathematics is identical in the two cases, so that only the result will be given here. The cross section for unpolarized beams is that of Eqs. (4.6) and (4.7) averaged over $\bar{m}^a$ and $\bar{m}^b$ and summed over m^a and m^b. It can be expanded in the form

$$\sigma(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b; T|l^ak^al^bk^b; R) = \sum_{\lambda^3}\sigma(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b; l^ak^al^bk^b|\lambda^3)D(\lambda^300|RT^{-1}), \quad (4.8)$$

where the expansion coefficient has the form

$$\begin{aligned}\sigma(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b; l^ak^al^bk^b|\lambda^3) &= (k/\bar{k})[(2l^a+1)(2l^b+1)]^{-1}\sum i^{\lambda^*}(2L+1)(2L'+1)\\ &\times(2\bar{\lambda}+1)(2\bar{\lambda}'+1)C(\bar{\lambda}\bar{\lambda}'\lambda_300)C(\lambda\lambda'\lambda_300)\\ &\times W(L\lambda L'\lambda'; l\lambda_3)W(L\bar{\lambda}L'\bar{\lambda}'; \bar{l}\lambda_3)\\ &\times f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}\bar{\lambda}; L; l^ak^al^bk^bl\lambda)\\ &\times f(\bar{l}^a\bar{k}^a\bar{l}^b\bar{k}^b\bar{l}'\bar{\lambda}'; L'; l^ak^al^bk^bl'\lambda')^*,\end{aligned}$$

where

$$\lambda^* = \bar{\lambda} - \bar{\lambda}' + 2(l+\bar{l})$$

and the sum is over $\bar{l}l\bar{\lambda}\lambda\bar{\lambda}'\lambda'LL'$.

5. DISCUSSION

The relationship of the two sorts of cross sections, for polarized and unpolarized beams, deserves some further comment. Ordinarily, in scattering theory, one chooses the coordinate system so that the relative velocity is along the z axis, so that the variable denoted by T in the foregoing is not needed. However, when particles with internal structure are scattered, a coordinate system is required to which to refer the internal wave function. One has the choice of fixing this coordinate system once and for all, and accepting the extra variable T, or else changing it with each change in the relative velocity and thereby needing a variable $\bar{T}$ to give the orientation of each coordinate system relative to a standard. However, in the case of unpolarized beams the distinction loses its meaning, since the cross section is a function only of the angle between the incoming and outgoing beams, rather than each direction separately.

The choice made in this work, to leave the coordinates fixed, was dictated by the hoped-for application of this theory to the kinetic theory of gases. One typically describes the state of a gas by a distribution in velocity and internal quantum numbers, and the collisions are the mechanism by which this distribution changes. The distribution function must refer to the same coordinate system for all states, and it is clearly simpler to require the same of the collision cross section.

There is a further problem in the kinetic theory, that of detailed balance, for which the present conventions seem better. The classical statement of the principle of detailed balance is that for every collision there is also one with incoming and outgoing velocities interchanged. The statement of it for symmetric tops with the present convention on coordinate systems is that

$$\begin{aligned}k(l^ak^al^bk^b)^2\sigma(l^ak^am^a; l^bk^bm^b; T|E|l'^ak'^am'^a; l'^bk'^bm'^b; T') &= k(l'^ak'^al'^bk'^b)^2\\ &\times\sigma(l'^ak'^am'^a; l'^bk'^bm'^b; T'|E|l^ak^am^a, l^bk^bm^b; T),\end{aligned} \quad (5.1)$$

whereas it would be much more complicated with the other convention, since each side of the equation would have its own coordinate system. It may be stated, parenthetically, that Eq. (5.1) is false, but that it becomes true if averaged over the m's, but that is another matter.

Finally, there is the problem of the numerical calculation of these cross sections. Direct numerical solution of the integral equations seems out of the question. There is always the Born approximation, but I gather that it is unsuitable for such low-energy collisions. The method of distorted waves has been used for H_2-H_2 collisions,[5] and, in view of the selection rule on ν_1 and ν_2 [in Eq. (3.10)], should not be any more difficult to apply to symmetric tops. A method involving less calculation is the rigid near-sphere approximation, which also has been previously applied[6] only to hydrogen.

It is felt that worse, rather than better, approximations are needed; if this implies an easing of the calculation problem. Possibilities include rigid potentials, as well as delta function potentials with angle-dependent factors. The virtue of the delta function potential is that the matrix element is simply the product of the values of the two wave functions at a particular point. At the present state of the theory, the choice of which molecules to consider for further work depends on the availability of experimental data.

[5] K. Takayanagi, Progr. Theoret. Phys. (Kyoto) 8, 497 (1952).
[6] G. Gioumousis and C. F. Curtiss (unpublished).

Comparison of Atom and Bond Percolation Processes

J. M. HAMMERSLEY
Institute of Statistics, University of Oxford, England
(Received March 16, 1961)

Various inequalities, some of them strict, are proved concerning probabilities associated with percolation processes. In particular, it is shown that the critical probability of an atom percolation process is not less than the critical probability of the corresponding bond percolation process.

STATEMENT OF RESULTS

PERCOLATION processes deal with the flow of a deterministic fluid through a random medium, whereas diffusion processes deal with a random fluid in a deterministic medium. The terms *fluid* and *medium* are abstract and interpretable according to context. Both types of process have many and various applications.[1,2] We begin with a definition of our terms.

A *medium* is defined to be an infinite set of abstract objects, called *atoms* and *bonds*. A bond may be a one-way oriented path from one atom to another, or it may be a two-way path between a pair of atoms. The medium may contain bonds of either or of both types. In alternative mathematical language, the atoms are the nodes and the bonds are the loops of an (unoriented, oriented, or partially oriented) infinite linear graph. A *walk* is defined to be an ordered sequence of bonds, each *step* of the walk being along (and, whenever relevant, in the direction of orientation of) a bond leading from an atom reached by the previous step. A walk is *self-avoiding* if it visits no atom more than once. Let A be a given atom of the medium, fixed hereafter and called the *source atom*. We shall always suppose that (a) the number of bonds leading from (though not necessarily to) any atom is finite (though perhaps unbounded); and (b) there is at least one infinitely long self-avoiding walk leading from the source atom. Conditions (a) and (b) are not at all restrictive, and are introduced merely to simplify the statement and proof of results. Indeed, I cannot think of a situation where neither the results are trivial nor can the situation be easily reduced to one for which (a) and (b) apply.

In an *atom percolation process*, each atom of the medium has, independently of all other atoms, a fixed probability p of being *undammed* and a probability $q=1-p$ of being *dammed*. A walk, starting at the source atom, is said to be undammed if all the atoms which it visits (except perhaps the source atom itself) are undammed. (The point of excluding the source atom in the previous sentence is that, if the source atom were included, all our results would hold *a fortiori*.) Let S be an arbitrary set of self-avoiding walks leading from the source atom; and write $\alpha(p,S)$ for the probability that at least one of these walks is undammed. Thus $\alpha(p,S)$ is the probability that fluid supplied to the source atom and travelling along paths which avoid dammed atoms, percolates along and as far as the end of some walk of S, or (when S is the set of all infinite self-avoiding walks) percolates to infinity. When S consists of all n-stepped self-avoiding walks from the source atom, we shall write $\alpha_n(p)$ for $\alpha(p,S)$. Clearly $\alpha_n(p)$ is a nonincreasing function of n for fixed p, and thus

$$\alpha(p)=\lim_{n\to\infty}\alpha_n(p) \tag{1}$$

exists and represents the probability of percolation to infinity. The supremum of all values of p such that $\alpha(p)=0$ is called the critical probability and is denoted by π_α. Since $\alpha(p)$ is a nondecreasing function of p, infinite percolation is inhibited with probability 1 whenever $p<\pi_\alpha$.

In a *bond percolation process*, each bond of the medium has, independently of all other bonds, a fixed probability p of being *undammed* and a fixed probability $q=1-p$ of being *dammed*. A walk, starting at the source atom, is said to be undammed if all its bonds are undammed. This time we write $\beta(p,S)$ for the probability that at least one of the walks of S is undammed, and $\beta_n(p)$ for $\beta(p,S)$ when S is the set of all n-stepped self-avoiding walks. As before,

$$\beta(p)=\lim_{n\to\infty}\beta_n(p) \tag{2}$$

exists and represents the probability of percolation to infinity. The critical probability π_β is the supremum of all p such that $\beta(p)=0$.

M. E. Fisher[2] proved that, if the medium is either the unoriented square lattice in the plane or the unoriented triangular lattice in the plane, then

$$\alpha(p)\leqslant\beta(p), \tag{3}$$

from which it follows immediately that (for these two particular media)

$$\pi_\alpha\geqslant\pi_\beta. \tag{4}$$

The main purpose of this paper is to show that (3) and (4) hold for any medium possessing properties (a) and (b) above. In fact, for any such arbitrary medium,

$$\alpha(p,S)\leqslant\beta(p,S) \tag{5}$$

for arbitrary S. The problem of determining when strict inequality holds in (5) is an interesting one. We shall give a complete answer to it in the case when all

[1] S. R. Broadbent and J. M. Hammersley, Proc. Cambridge Phil. Soc. **53**, 629–641 (1957).
[2] M. E. Fisher, J. Math. Phys. **2**, 620 (1961).

the walks of S are finite, and a partial answer when S contains infinite walks. As yet, however, I cannot invent a technique for determining strict inequality in (4). The technique used for proving (5) is an elaboration of that used in[3] for finding an upper bound to $\beta_n(p)$. Clearly, such an upper bound is also an upper bound to $\alpha_n(p)$, $\alpha(p)$, and $\beta(p)$.

Consider first the case when all walks of S have n or fewer steps. Let r, a function of S, be the greatest number of bonds which leave any single atom on a walk of S, including the source atom but excluding the final atoms of each walk of S. Let ν denote the excess of the number of distinct bonds, occurring on all walks of S taken together, over the number of distinct atoms, occurring on all walks of S excluding the source atom but including the final atoms of each walk. We shall show that

$$\alpha(p,S)+\nu(pq^{r-1})^{2n-1}\leqslant\beta(p,S)\leqslant\alpha(p,S)+\nu pq. \quad (6)$$

From (6) it follows that equality holds in (5) if and only if at least one of the quantities ν, p, q is zero. It is easy to see that $\nu=0$ if and only if the linear graph, formed by the atoms and bonds belonging to S, is a tree. When S does not form a tree, (6) gives information about the magnitude of the strict inequality in (5).

Next, consider the cases when S contains an infinite walk. It still remains true that equality holds in (5) when S forms a tree. But, when S does not form a tree, the cases of inequality are more difficult and require a further assumption which is sufficient though unnecessarily strong. We shall say that an atom is *n accessible* if there is a walk of n or fewer steps (which without loss of generality may be supposed self-avoiding), from the source atom to the atom in question. We also say that an atom is doubly n accessible if there exist two distinct walks each of n or fewer steps from the source atom to the atom in question such that both walks are (this time necessarily and *not* without loss of generality) self-avoiding. The additional assumption is that every infinitely long self-avoiding walk from the source atom shall contain at least one doubly n-accessible atom for some fixed value of n independent of the walk considered. Then, under this additional assumption,

$$\alpha(p)\leqslant\{1-(pq^{r-1})^{2n-1}/\mu\}\beta(p), \quad (7)$$

where r is now the greatest number of bonds leaving any single $(n-1)$-accessible atom (including the source atom itself), and where μ is the total number of distinct n-accessible atoms (excluding the source atom itself). Thus, strict inequality holds in (3) under the above additional assumption provided that $\pi_\beta<p<1$. Of course, when $0\leqslant p<\pi_\beta$, we have $\alpha(p)=\beta(p)=0$, which explains why we can only hope a relationship of the type of (7) rather than (6) valid for all p.

The inequality (7) is by no means the best possible, and it can be improved in a number of special cases by relatively trivial modifications in the proof. For example, if the medium is unoriented we may sharpen both (6) and (7) by replacing $r-1$ by $r-2+(2n-1)^{-1}$. For the unoriented square lattice in the plane (for which we may take $n=3$, $r=4$, and $\mu=24$), this sharpening gives

$$\alpha(p)\leqslant(1-p^5q^{11}/24)\beta(p); \quad (8)$$

and by using the particular geometry of the square lattice, we can further sharpen (8) by replacing $p^5q^{11}/24$ by $p^5q^9/12$. Even this last result is doubtless not as sharp as possible.

Finally, we exhibit two simple examples to show that π_α may equal π_β even though the medium is not a tree. Let T denote an infinite tree of one-way bonds, such that T has a single root and each atom has three bonds leading from it. Let each atom of the unoriented square lattice be the root of an individual tree T. The resulting medium is not a tree (and incidentally it does not fulfil the foregoing additional assumption); but, for a source atom on the square lattice (and at the root of some T), $\pi_\alpha=\pi_\beta=\frac{1}{3}$. Moreover, for $\frac{1}{3}<p<1$, we have $\alpha(p)<\beta(p)$; thus exhibiting the unnessarily strong nature of the additional assumption. As a second example, suppose that two bonds lead from the source atom to the root of T and that these two bonds together with T constitute the complete medium. Then the additional assumption is satisfied, with $n=\mu=1$ and $r=2$; and (7) holds although $\pi_\alpha=\pi_\beta=\frac{1}{3}$. For this second example, $\alpha(p)\times\{1+q\}=\beta(p)$; so (7) is again not the best possible.

PROOF OF RESULTS

We begin by proving a lemma on independent equidistributed random variables which can only assume the values 0 and 1. Let X be a finite set of distinct elements $X_1, X_2, \cdots, X_n$. To each element X_r let there correspond a distinct random variable $x_r=0$ or 1, such that its expectation $Ex_r=p$ and such that $x_1, x_2, \cdots, x_n$ are mutually independent. For each $i=1, 2, \cdots, m$, let Y_i denote a subset of the elements of X, say

$$Y_i=\{Y_{i1}, Y_{i2}, \cdots, Y_{iI(i)}\}, \quad (9)$$

where each $Y_{ij}=X_r$ for some r. When $Y_{ij}=X_r$, we define the random variable y_{ij} by $y_{ij}=x_r$. We shall say that a given subset Y_i is *admissible* if $Y_{i1}, Y_{i2}, \cdots, Y_{iI(i)}$ are all distinct elements of X. Suppose that, for each fixed value of i, Y_i is an admissible subset. Thus, $I(i)\leqslant n$ in an admissible subset. Notice that the complete set of Y_{ij} for all i, j need not consist of distinct elements, since we could have $Y_{11}=Y_{21}$ for example.

Let Y and Y' be two particular distinct Y_{ij} with the property that, if Y_i^* is the subset obtained from Y_i by replacing Y' (wherever it occurs) by Y, then Y_i^* is an admissible subset for each fixed $i=1, 2, \cdots, m$. If y and y' denote the distinct random variables y_{ij} associated with Y and Y', we replace y' by y whenever

[3] J. M. Hammersley and R. S. Walters, J. Soc. Indust. Appl. Math. (to be published).

Y' is replaced by Y. Define for the subsets Y_i

$$Z_i=1-y_{i1}y_{i2}\cdots y_{iI(i)} \tag{10}$$

and

$$Z_0=Z_1Z_2\cdots Z_m. \tag{11}$$

Corresponding to the subsets Y_i^*, we may define similar quantities Z_i^* and Z_0^* by replacing y' by y in (10), and Z_i by Z_i^* in (10) and (11).

No subset Y_i can contain both Y and Y', for otherwise Y_i^* would not be admissible. Hence, we can classify the subsets Y_i into three classes: a class C such that Y but not Y' belongs to each Y_i of C, a class C' such that Y' but not Y belongs to each Y_i of C', and a class C'' such that neither Y nor Y' belongs to any Y_i of C''. Here C'' may be empty, but C and C' cannot be empty. We define Z to be the product of those Z_i for which Y_i belongs to C, and similarly Z' to be the product of those Z_i for which Y_i belongs to C'. Finally, Z'' is defined to be 1 if C'' is empty, and otherwise to be the product of those Z_i for which Y_i belongs to C''. Thus

$$Z_0=ZZ'Z''. \tag{12}$$

If Y_i belongs to C, we can write

$$Z_i=1-yz_i, \tag{13}$$

where z_i is independent of y and y'. Since y equals any positive power of itself, being 0 or 1, we get from (13)

$$Z=1-yz, \tag{14}$$

where z is independent of y and y'. This independence shows that z can take any of its possible values when $y=1$; and therefore all possible values of $1-z$ are included amongst the possible values of Z. However, by (10), each Z_i and therefore Z also can only take the values 0 or 1. Hence $1-z$ can only take the values 0 or 1. Thus,

$$z=0 \text{ or } 1. \tag{15}$$

By a similar argument we can write

$$Z'=1-y'z', \tag{16}$$

where z' is independent of y and y', and

$$z'=0 \text{ or } 1. \tag{17}$$

Finally,

$$Z''=z''=0 \text{ or } 1, \tag{18}$$

where z'' is independent of y and y'. From (12), (14), (16), and (18),

$$Z_0=z''(1-yz)(1-y'z'). \tag{19}$$

We obtain Z_0^* by replacing y' by y in (19). Thus

$$Z_0^*-Z_0=z'z''(1-yz)(y'-y)=z'z''(y'-y)+zz'z''(y-yy'), \tag{20}$$

because $y^2=y$. Since $Ey=Ey'=p$, we get

$$EZ_0^*=EZ_0+pqE(zz'z''). \tag{21}$$

By (15), (17), and (18),

$$zz'z''=0 \text{ or } 1; \tag{22}$$

and hence,

$$E(zz'z'')=\text{Prob}(zz'z''=1)=\text{Prob}(z=z'=z''=1). \tag{23}$$

Similarly, since Z_0 can only take the values 0 or 1, we have

$$EZ_0=\text{Prob}(Z_0=1)=1-\text{Prob}(Z_0=0) \tag{24}$$

together with a similar result for Z_0^* in place of Z_0. From (21), (23), and (24) we obtain the required lemma:

$$\text{Prob}(Z_0=0)=\text{Prob}(Z_0^*=0)+pq\ \text{Prob}(z=z'=z''=1). \tag{25}$$

We now apply this lemma to the proof of (6). Let S denote an arbitrary set of self-avoiding walks, each starting from the source atom and having n or fewer steps. We take X to be the set of all distinct bonds belonging to the walks of S. Let the walks of S be $S_1, S_2, \cdots, S_m$; and let $Y_{i1}, Y_{i2}, \cdots, Y_{iI(i)}$ be the successive bonds on S_i $(i=1, 2\cdots, m)$. Since S_i is self-avoiding, Y_i is an admissible set. We take $y_{ij}=0$ or 1 according as Y_{ij} is dammed or undammed in the bond percolation process. From (10) we see that $Z_i=0$ or 1 according as S_i is undammed or dammed; and then, from (11), $Z_0=0$ if and only if at least one walk of S is undammed. Thus

$$\beta(p,S)=\text{Prob}(Z_0=0). \tag{26}$$

Similarly, let W be the complete set of atoms on walks of S, excluding the source atom; and write $W_{i1}, W_{i2}, \cdots, W_{iI(i)}$ for the successive atoms on S_i, excluding the source atom. Let $w_{ij}=0$ or 1 according as w_{ij} is dammed or undammed in the atom percolation process. If W_0 is defined in terms of the w_{ij} in the same way as Z_0 is defined in terms of the y_{ij}, we have in conformity with (26)

$$\alpha(p,S)=\text{Prob}(W_0=0). \tag{27}$$

To each atom of W we now select a bond of X by the following association procedure, which ensures that distinct atoms of W are associated with distinct bonds of X. With each atom on S_1 associate the bond of S_1 leading to that atom. With an atom, which lies on S_i but does not lie on $S_1, S_2, \cdots, S_{i-1}$ (where $1<i\leq m$), associate the bond of S_i leading to that atom. This may, and in general will, leave some unselected bonds of X. The number ν, occurring in (6), is the excess of the number of members of X over the number of members of W. If $\nu=0$, there are no unselected members of X; and the foregoing mapping provides a (1,1) correspondence between the Y_{ij} and the W_{ij}. In this case, the distribution of the w_{ij} will be the same as the distribution of the y_{ij} and we shall have

$$\alpha(p,S)=\beta(p,S). \tag{28}$$

If, however, $\nu \neq 0$, there will be some unselected bonds. Let Y' denote any given unselected bond, and let Y denote the selected bond leading to the same atom as Y' does; and let y and y' be the random y_{ij} belonging to Y and Y', respectively. We now replace y' by y. This will change $\text{Prob}(Z_0=0)$ by the quantity

$$pq\ \text{Prob}(z=z'=z''=1) \tag{29}$$

occurring on the right of (25). We can, in fact, repeat this process of replacement ν times, until there are no unselected bonds left. The eventually residual set of y_{ij} will have the same distribution as the w_{ij}. Hence, the complete set of replacements reduces (26) to (27), the total reduction being the sum of terms (29), one term for each reduction. Thus, (6) will be established as soon as we have proved that

$$(pq^{r-1})^{2n-1} \leqslant pq\ \text{Prob}(z=z'=z''=1) \leqslant pq \tag{30}$$

holds at each stage of the reduction process. The right-hand inequality of (30) is trivial. To see the validity of the left-hand inequality in (30), consider two particular walks S_0 and S_0' of S such that S_0 contains Y and S_0' contains Y'. Suppose that all bonds of S_0 are undammed. Then,

$$0=Z=1-yz. \tag{31}$$

Hence, $z=1$ when $y=1$. But z is independent of y. Therefore, $z=1$ if all bonds of S_0 except perhaps Y are undammed. Similarly $z'=1$ if all bonds of S_0' except perhaps Y' are undammed. The total number of bonds in S_0 and S_0' apart from Y and Y' is at most $2(n-1)$. If, further, $A_1, A_2, \cdots, A_k$ are the distinct atoms on S_0 and S_0', including the source atom but excluding the two final atoms, then $k \leqslant 2n-1$ since the source atom is common to both S_0 and S_0'. If every (existing) bond from each A_j $(j=1, 2, \cdots, k)$ is dammed except when it is a bond of S_0 or S_0', there can be no other undammed walk of S apart from S_0 and S_0', and in this case $Z''=z''=1$. The total number of distinct bonds, which have to be dammed in the previous sentence, is at most $(r-1)k-1 \leqslant (r-1)(2n-1)-1$ by virtue of the definition of r as the maximum number of bonds from an atom. The final -1 in the previous sentence arises from the consideration that, since S_0 and S_0' are not identical, at least one of the $A_1, A_2, \cdots, A_k$, has two distinct bonds leading from it which both belong to either S_0 or S_0'. Hence, by damming a given set of $s \leqslant (r-1)(2n-1)-1$ distinct bonds and by undamming a given set of $t \leqslant 2(n-1)$ distinct bonds (where the two sets are mutually disjoint), we can ensure that $z=z'=z''=1$. Thus,

$$\text{Prob}(z=z'=z''=1) \geqslant p^t q^s \geqslant p^{2(n-1)} q^{(r-1)(2n-1)-1}; \tag{32}$$

and the left-hand side of (30) follows at once. It is easy to see that, in an unoriented medium, we can replace $(r-1)(2n-1)-1$ by $(r-2)(2n-1)$.

To complete the proof we have to verify that the subsets Y_i, obtained at each stage of replacement, are admissible. They will all certainly be admissible if the particular subsets, obtained after the final replacement, are admissible; and this admissibility is satisfied because the finally obtained subsets are in (1,1)-correspondence with the subsets W_i of W, which are admissible because a self-avoiding walk does not visit any *atom* more than once.

Next we deduce (5) from (6). When S is an arbitrary system of self-avoiding walks starting from the source atom, let S^N denote the set of all walks of S, each truncated after its Nth step whenever it has such a step. Then, by (6),

$$\alpha(p,S^N) \leqslant \beta(p,S^N); \tag{33}$$

and on letting $N \to \infty$, we obtain (5).

Finally, and hereafter, we consider the case when S is the set of all infinitely long self-avoiding walks starting at the source atom. S may, and in general will, contain an uncountable infinity of walks. But, because only a finite number of bonds lead from any given atom, the number of members of S cannot exceed the power of the continuum. The walks of S may therefore be indexed as S_θ where θ is a real number. Let S^N denote the set of walks obtained by truncating each walk of S after its Nth step. Since $\alpha(p,S^N)$ and $\beta(p,S^N)$ tend to $\alpha(p,S)$ and $\beta(p,S)$, respectively, as $N \to \infty$; there exists a number $M=M(\epsilon)$ corresponding to any prescribed $\epsilon>0$ such that

$$\begin{aligned} \alpha(p,S) &\leqslant \alpha(p,S^M) \leqslant \alpha(p,S)+\epsilon, \\ \beta(p,S) &\leqslant \beta(p,S^M) \leqslant \beta(p,S)+\epsilon. \end{aligned} \tag{34}$$

Let n be fixed positive integer. Since there are only finitely many bonds from any given atom, there are only finitely many n-accessible atoms, an n-accessible atom being one which can be reached from the source atom by a self-avoiding walk of n or fewer steps. Hence, each walk S_θ of S passes through only finitely many n-accessible atoms and therefore possesses an atom A_θ, which is the last n-accessible atom that it visits. Let S_θ' be the walk obtained from S_θ by truncating S_θ precisely M steps after it has passed through A_θ. Then S_θ' is a walk of finite (though possibly unbounded) length whose last M atoms are not n accessible. Let S' denote the set of all S_θ'. There are only a finite number of self-avoiding walks starting at the source atom and having a fixed number of steps. Hence, S' contains either a finite or a countably infinite number of walks, because each walk of S' is of finite length. Hereafter, therefore, we can write S_i' for a member of S' where i is a positive integer. Since S' is either countable or finite, we may select a finite subset S'' of S' such that

$$\begin{aligned} \alpha(p,S')-\epsilon &\leqslant \alpha(p,S'') \leqslant \alpha(p,S'), \\ \beta(p,S')-\epsilon &\leqslant \beta(p,S'') \leqslant \beta(p,S'). \end{aligned} \tag{35}$$

Here S'' depends, of course, upon ϵ. If (in either atom or bond process respectively) there exists at least one

undammed walk of S, there exists at least one undammed walk of S'; and if there exists at least one undammed walk of S', there exists at least one undammed walk of S^M, since every walk of S' contains at least M steps. Thus,

$$\alpha(p,S)\leqslant\alpha(p,S')\leqslant\alpha(p,S^M),$$
$$\beta(p,S)\leqslant\beta(p,S')\leqslant\beta(p,S^M). \tag{36}$$

We say that an atom is doubly n accessible if it can be reached from the source atom by either of two distinct self-avoiding walks both of n or fewer steps. Suppose that every walk of S contains at least one doubly n-accessible atom. Since any S_i' of S' is obtained from an S_i of S by omitting certain atoms, all of which are not n accessible, S_i' must contain at least one doubly n-accessible atom. Let A_i denote the least n-accessible atom visited on a given arbitrary S_i' of S'; and let T_i denote the part of S_i' lying beyond A_i. Let U_i denote a self-avoiding walk of n or fewer steps starting at the source atom and ending at A_i; such a walk exists because A_i is n accessible. T_i is not empty because it contains M steps, by virtue of the definition of S'. Let V_i denote the walk obtained by starting at the source atom, following U_i until A_i is reached, and then following T_i. Since each part, U_i and T_i, of V_i is self-avoiding, and since no atom on T_i is n accessible whereas every atom on U_i is n accessible, V_i is a self-avoiding walk; and hence V_i belongs to S', because it is terminated M steps after its last n accessible atom. Therefore, V_i contains a doubly n-accessible atom, say A_i', which necessarily lies on U_i. If there is more than one doubly n-accessible atom on U_i, we take A_i' to be the last one of these. Let U_i' be the part of U_i from the source atom to A_i'. Since A_i' is doubly n accessible, there exists a self-avoiding walk U_i'' of n or fewer steps from the source atom to A_i' such that U_i'' is not identical with U_i'. Moreover, U_i'' cannot intersect any atom (if such exists) lying on U_i' and occurring after A_i', for otherwise *either* U_i'' would not be self-avoiding *or* A_i' would not be the last doubly n-accessible atom on U_i. Hence, V_i', defined as the walk starting at the source atom, following U_i'' as far as A_i', then following U_i from A_i' to A_i, and finally following S_i' from A_i onwards, is a self-avoiding walk; and V_i' belongs to S' and is not identical with V_i. It may, or may not, be the case that S_i' is identical with either V_i or V_i'. We shall say that V_i and V_i' are a pair of associates of S_i'. The essential properties of associates, obtained by the above construction, are that the two associates of S_i' (i) both coincide with S_i' beyond the last n-accessible atom, (ii) both belong to S', and (iii) are not identical (the one with the other) although the lack of identity is confined to their first n steps.

Next, let R denote the set consisting of all walks of the form S_i', V_i, V_i', where S_i' runs over all walks of S'', and V_i and V_i' are a pair of associates of S_i'. (If S_i' possesses more than one pair of associates, arbitrarily choose just one such pair for inclusion in R.) The construction of the previous paragraph is such that V_i and V_i' are a pair of associates of V_i itself, and are a pair of associates of V_i' itself. Hence, R is closed under association: that is to say, every walk of R possesses a pair of associates in R. The number of walks in R is finite, being at most three times the number of walks in S''. Since S'' is a subset of R, and R is a subset of S', we have

$$\alpha(p,S'')\leqslant\alpha(p,R)\leqslant\alpha(p,S'),$$
$$\beta(p,S'')\leqslant\beta(p,R)\leqslant\beta(p,S'). \tag{37}$$

Now S consists of all infinitely long self-avoiding walks starting from the source atom; and accordingly,

$$\alpha(p,S)=\alpha(p),\quad \beta(p,S)=\beta(p), \tag{38}$$

by the definition of $\alpha(p)$ and $\beta(p)$. From (34)–(38) we conclude that

$$|\alpha(p,R)-\alpha(p)|\leqslant\epsilon,\quad |\beta(p,R)-\beta(p)|\leqslant\epsilon, \tag{39}$$

where R is a finite class of finite self-avoiding walks and where R is closed under association.

Let μ denote the total number of n-accessible atoms. Then the set of atoms, each of which is a last n-accessible atom on some walk of R, has at most μ members. Consequently, there exists an atom A_0, which is the last n-accessible atom on some walk of R, such that there is a probability not less than $\beta(p,R)/\mu$ that at least one walk of R passing through A_0 and having no subsequent n-accessible atoms is undammed in the bond percolation process for R. *A fortiori*,

$$\beta(p,R)/\mu\leqslant\mathrm{Prob}(F), \tag{40}$$

where F is the event that at least one walk of R passing through A_0 and having no subsequent n-accessible atoms is undammed from A_0 onwards. Clearly, the event F is independent of the state of any bond which is traversed during or before an nth step of any walk starting at the source atom. Let R_0 be any walk of R passing through A_0 and having no subsequent n-accessible atoms; and let V_0 and V_0' be the associates of R_0. Let U_0 and U_0' be the parts of V_0 and V_0' respectively, between the source atom and A_0. Since U_0 and U_0' are distinct walks terminating at a common atom, they each contain bonds B_0 and B_0', respectively, such that B_0 and B_0' both lead to the same atom. Let G be the event that (a) every bond of U_0 and U_0' except perhaps B_0 and B_0' is undammed, and (b) every bond (other than a bond of U_0 or U_0') which leads from an atom on either U_0 or U_0' (including the source atom but excluding A_0) is dammed. As in the proof of (32), we have

$$\mathrm{Prob}(G)\geqslant p^{2(n-1)}q^{(r-1)(2n-1)-1}. \tag{41}$$

Since G only relates to the states of bonds which can be traversed during or before the nth step of some walk starting from the source atom, G is independent of F.

Thus, from (40) and (41),

$$\text{Prob}(FG)=\text{Prob}(F)\text{Prob}(G) \geqslant p^{2(n-1)}q^{(r-1)(2n-1)-1}\beta(p,R)/\mu. \quad (42)$$

Now apply Lemma (25) with X being the set of all distinct bonds of the finite set R of finite walks, and where Z_0 and $Z_0{}^*$ differ only in that the latter is obtained after identification of the random variables associated with B_0 and B_0'. If both F and G occur, then $z=z'=z''=1$. Hence,

$$\beta(p,R)=\text{Prob}(Z_0=0)\geqslant\text{Prob}(Z_0{}^*=0) +pq\,\text{Prob}(z=z'=z''=1), \quad (43)$$

$$\text{Prob}(z=z'=z''=1)\geqslant\text{Prob}(FG), \quad (44)$$

and

$$\text{Prob}(Z_0{}^*=0)\geqslant\alpha(p,R), \quad (45)$$

since any further identification of random variables cannot increase the expressions corresponding to $\text{Prob}(Z_0{}^*=0)$ and leading to $\alpha(p,R)$. From (42)–(45) we obtain

$$\beta(p,R)\geqslant\alpha(p,r)+(pq^{r-1})^{2n-1}\beta(p,R)/\mu. \quad (46)$$

In (46), n, and hence r and μ, are fixed independently of ϵ. On letting $\epsilon\to 0$ and using (39), we deduce (7) from (46).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

Eigenvalue Problems in Matrix Mechanics*

RICHARD R. CHASMAN
Argonne National Laboratory, Argonne, Illinois

(Received March 3, 1961)

The techniques of the Heisenberg matrix mechanics are extended to treat all potentials of the form Q^{2n}, for n a positive integer. The square well (Q^{∞}) and the potential Q^4 are treated in detail.

1. INTRODUCTION

ALTHOUGH the Heisenberg matrix mechanics gives us a very convenient way to handle the harmonic oscillator, little else has been done with it. The purpose of this paper is to apply the matrix techniques to obtain approximate bound-state eigenvalues for all potentials of the form Q^{2n}, for n a positive integer. The technique used here will be based on the form of the Q matrix of the harmonic oscillator. As Q^{∞} (square well) differs most from Q^2 of all potentials Q^{2n}, it is most instructive to treat the square well in detail. Because the square well can be solved exactly by other methods, a detailed calculation using the techniques to be developed here will give us a good idea of the accuracy of these techniques to various orders of approximation. We also consider the potential Q^4 in detail to see how well the techniques work as we approach the harmonic oscillator potential.

2. DEVELOPMENT OF EQUATIONS

We start with the postulate

$$PQ-QP=i\hbar, \quad (1)$$

where P is the momentum matrix and Q is the position matrix, and the further postulate that we may choose a representation in which the Hamiltonian H is diagonal.

For the potentials under consideration here, we have the general equation

$$H=(P^2/2m)+(\alpha/n)Q^n. \quad (2)$$

The commutators of P and Q with the Hamiltonian can be calculated using Eq. (1), and we obtain

$$[Q,H]\equiv(QH-HQ)=-(i\hbar P/m), \quad (3)$$

$$[P,H]\equiv(PH-HP)=i\hbar\alpha Q^{n-1}. \quad (4)$$

The harmonic oscillator ($n=2$) can be solved quite readily by taking the i, j matrix elements in Eqs. (3) and (4) and combining the two equations to get

$$Q_{ij}(H_{jj}-H_{ii})^2=(\alpha/m)\hbar^2Q_{ij}. \quad (5)$$

The solutions to Eq. (5) are

$$\left.\begin{aligned} &Q_{ij}\neq 0\\ &(H_{jj}-H_{ii})^2=(\alpha/m)\hbar^2\end{aligned}\right\}, \quad (6)$$

or

$$Q_{ij}=0. \quad (7)$$

First, we convince ourselves that there can be no eigenvalue H_{i_0,i_0} for which $Q_{i_0,j}$ vanishes for all j. We do this by taking the i_0,j matrix element of Eq. (3) and the i_0,i_0 matrix element of Eq. (1) and we are led to the contradiction, for vanishing Q_{i_0j},

$$\textstyle\sum_j P_{i_0,j}Q_{j,i_0}-Q_{i_0,j}P_{j,i_0}=i\hbar=0. \quad (8)$$

Because of this result, we can be certain that all levels have been accounted for in Eq. (6), and if we now arrange the energy levels in order of increasing energy, we obtain

$$Q_{ij}=0 \quad \text{when } j\neq i\pm 1. \quad (9)$$

Solving the harmonic oscillator in this way gives no clue as to how one treats other potentials of the form

* Based on work performed under the auspices of the U. S. Atomic Energy Commission.

Thus, from (40) and (41),

$$\text{Prob}(FG)=\text{Prob}(F)\text{Prob}(G) \geqslant p^{2(n-1)}q^{(r-1)(2n-1)-1}\beta(p,R)/\mu. \quad (42)$$

Now apply Lemma (25) with X being the set of all distinct bonds of the finite set R of finite walks, and where Z_0 and $Z_0{}^*$ differ only in that the latter is obtained after identification of the random variables associated with B_0 and B_0'. If both F and G occur, then $z=z'=z''=1$. Hence,

$$\beta(p,R)=\text{Prob}(Z_0=0)\geqslant\text{Prob}(Z_0{}^*=0) +pq\,\text{Prob}(z=z'=z''=1), \quad (43)$$

$$\text{Prob}(z=z'=z''=1)\geqslant\text{Prob}(FG), \quad (44)$$

and

$$\text{Prob}(Z_0{}^*=0)\geqslant\alpha(p,R), \quad (45)$$

since any further identification of random variables cannot increase the expressions corresponding to $\text{Prob}(Z_0{}^*=0)$ and leading to $\alpha(p,R)$. From (42)–(45) we obtain

$$\beta(p,R)\geqslant\alpha(p,r)+(pq^{r-1})^{2n-1}\beta(p,R)/\mu. \quad (46)$$

In (46), n, and hence r and μ, are fixed independently of ϵ. On letting $\epsilon\to 0$ and using (39), we deduce (7) from (46).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 5 SEPTEMBER–OCTOBER, 1961

Eigenvalue Problems in Matrix Mechanics*

RICHARD R. CHASMAN
Argonne National Laboratory, Argonne, Illinois

(Received March 3, 1961)

The techniques of the Heisenberg matrix mechanics are extended to treat all potentials of the form Q^{2n}, for n a positive integer. The square well (Q^{∞}) and the potential Q^4 are treated in detail.

1. INTRODUCTION

ALTHOUGH the Heisenberg matrix mechanics gives us a very convenient way to handle the harmonic oscillator, little else has been done with it. The purpose of this paper is to apply the matrix techniques to obtain approximate bound-state eigenvalues for all potentials of the form Q^{2n}, for n a positive integer. The technique used here will be based on the form of the Q matrix of the harmonic oscillator. As Q^{∞} (square well) differs most from Q^2 of all potentials Q^{2n}, it is most instructive to treat the square well in detail. Because the square well can be solved exactly by other methods, a detailed calculation using the techniques to be developed here will give us a good idea of the accuracy of these techniques to various orders of approximation. We also consider the potential Q^4 in detail to see how well the techniques work as we approach the harmonic oscillator potential.

2. DEVELOPMENT OF EQUATIONS

We start with the postulate

$$PQ-QP=i\hbar, \quad (1)$$

where P is the momentum matrix and Q is the position matrix, and the further postulate that we may choose a representation in which the Hamiltonian H is diagonal.

For the potentials under consideration here, we have the general equation

$$H=(P^2/2m)+(\alpha/n)Q^n. \quad (2)$$

The commutators of P and Q with the Hamiltonian can be calculated using Eq. (1), and we obtain

$$[Q,H]\equiv(QH-HQ)=-(i\hbar P/m), \quad (3)$$

$$[P,H]\equiv(PH-HP)=i\hbar\alpha Q^{n-1}. \quad (4)$$

The harmonic oscillator ($n=2$) can be solved quite readily by taking the i, j matrix elements in Eqs. (3) and (4) and combining the two equations to get

$$Q_{ij}(H_{jj}-H_{ii})^2=(\alpha/m)\hbar^2Q_{ij}. \quad (5)$$

The solutions to Eq. (5) are

$$\left.\begin{aligned} &Q_{ij}\neq 0\\ &(H_{jj}-H_{ii})^2=(\alpha/m)\hbar^2\end{aligned}\right\}, \quad (6)$$

or

$$Q_{ij}=0. \quad (7)$$

First, we convince ourselves that there can be no eigenvalue H_{i_0,i_0} for which $Q_{i_0,j}$ vanishes for all j. We do this by taking the i_0,j matrix element of Eq. (3) and the i_0,i_0 matrix element of Eq. (1) and we are led to the contradiction, for vanishing Q_{i_0j},

$$\textstyle\sum_j P_{i_0,j}Q_{j,i_0}-Q_{i_0,j}P_{j,i_0}=i\hbar=0. \quad (8)$$

Because of this result, we can be certain that all levels have been accounted for in Eq. (6), and if we now arrange the energy levels in order of increasing energy, we obtain

$$Q_{ij}=0 \quad \text{when } j\neq i\pm 1. \quad (9)$$

Solving the harmonic oscillator in this way gives no clue as to how one treats other potentials of the form

* Based on work performed under the auspices of the U. S. Atomic Energy Commission.

Q^n, as only for the harmonic oscillator is there a cancellation of Q_{ij} in Eq. (5) or its equivalent for other potentials.

Next, we develop a technique for solving the harmonic oscillator which can be extended to other potentials. Equation (3) is multiplied on the right by Q, and Eq. (4) is multiplied on the left by P. We take the diagonal matrix elements of both of the new equations and obtain

$$\alpha(Q^n)_{ii}=(P^2/m)_{ii}, \tag{10}$$

which is just the virial theorem. We next take diagonal matrix elements in Eq. (2), combine with Eq. (10) to get

$$H_{ii}=\frac{2+n}{2n}\left(\frac{P^2}{m}\right)_{ii}. \tag{11}$$

We calculate the commutator

$$[Q,[Q,H]]=-\hbar^2/m \tag{12}$$

and take the diagonal matrix element which gives

$$\textstyle\sum_j Q_{ij}(H_{jj}-H_{ii})Q_{ji}=\hbar^2/(2m). \tag{13}$$

We then square Eq. (3) and again take a diagonal matrix element to obtain

$$\sum_j Q_{ij}(H_{jj}-H_{ii})^2Q_{ji}=\frac{\hbar^2}{m}\left(\frac{P^2}{m}\right)_{ii}=\frac{\hbar^2}{m}\left(\frac{2n}{2+n}\right)H_{ii}. \tag{14}$$

Equations (13) and (14) can be solved for the harmonit oscillator by starting from the ground state, i.e., sec $i=0$ in the equations, and because j can then be only 1, we calculate

$$Q_{01}(H_{11}-H_{00})Q_{10}=\hbar^2/(2m), \tag{13}'$$

$$Q_{01}(H_{11}-H_{00})^2Q_{10}=(\hbar^2/m)H_{00}, \tag{14}'$$

$$\left.\begin{aligned}\tfrac{1}{2}(H_{11}-H_{00})&=H_{00}\\ H_{11}&=3H_{00}\end{aligned}\right\}. \tag{15}$$

We can then work up to higher energies, getting each energy in terms of H_{00}.

To get a first approximation for cases in which $n\neq 2$, we shall again assume a Q matrix of the same form as that of the oscillator, i.e.,

$$Q_{ij}=0 \quad \text{for } j\neq i\pm 1. \tag{16}$$

We may then use the energies obtained from this approximation to calculate values of the matrix elements for $j=i\pm 3$. (Q_{ij} must vanish for $j=i\pm k$, for k even.) We do this by substituting from Eq. (3) into Eq. (1) and taking the $i,i+2$ matrix element of Eq. (1)

$$\sum_{j=i\pm 1,\,j=i\pm 3} Q_{ij}Q_{j,i+2}[(H_{i+2,i+2}-H_{j,j})-(H_{jj}-H_{ii})]=0. \tag{17}$$

If we start from $i=0$, and use the matrix elements $Q_{i,i\pm 1}$ and the eigenvalues H_{ii} obtained from Eqs. (13) and (14), we can get ratios

$$\frac{Q_{i,i\pm 3}(H_{i\pm 3,i\pm 3}-H_{ii})Q_{i\pm 3,i}}{Q_{i,i+1}(H_{i+1,i+1}-H_{ii})Q_{i+1,i}},$$

and substitute the ratios so obtained into Eqs. (13) and (14), and again solve for $Q_{i,i\pm 1}$ and $H_{i,i}$, this time summing over $j=i\pm 1$, $j=i\pm 3$. To obtain the third-order approximation, we use the equations

$$\sum_{j=i\pm 1,\,i\pm 3,\,i\pm 5} Q_{ij}Q_{j,i\pm 2}[(H_{i+2,i+2}-H_{jj})-(H_{jj}-H_{ii})]=0, \tag{18}$$

$$\sum_{j=i\pm 1,\,i\pm 3,\,i\pm 5} Q_{ij}Q_{j,i\pm 4}[(H_{i+4,i+4}-H_{jj})-(H_{jj}-H_{ii})]=0. \tag{19}$$

Equations (18) and (19) must be iterated until the $Q_{i,i\pm 3}$ and $Q_{i,i\pm 5}$ matrix elements do not change.

If we are interested in obtaining the ground state energy in terms of the parameters α, m, $\hbar$; it can be done by combining Eq. (3) and Eq. (4) to get

$$Q_{i,j}(H_{jj}-H_{ii})^2=(\alpha\hbar^2/m)(Q^{n-1})_{i,j}, \tag{20}$$

and put all quantities in terms of H_{00}. This is not very convenient when n is large. The ratios of energy levels will be independent of the parameters α, $\hbar$, and m.

3. RESULTS OF CALCULATIONS

The equations developed in Sec. II were applied to the square well, which we define as

$$\lim_{n\to\infty}\alpha Q^n/n.$$

The eigenvalues of the Hamiltonian were calculated to three orders of approximation and these numbers, together with the exact values of the energy, are given in Table I. It is worthwhile to get some approximate notions of the rate of convergence from Table I. Except for H_{11}, which is accurate to 1% in the second approximation, it seems to be true that the difference

TABLE I. Energy levels of the square well (in units of H_{00}).

Energy	1st approx	2nd approx	3rd approx	Exact value
H_{00}	1	1	1	1
H_{11}	5	3.98	3.98	4
H_{22}	13	9.27	8.86	9
H_{33}	25	17.20	15.93	16
H_{44}	41	27.84	25.28	25
H_{55}	61	41.08	36.98	36
H_{66}	85	57.06	51.30	49
H_{77}	113	75.76	67.10	64
H_{88}	145	97.20	85.96	81
H_{99}	181	121.30	106.70	100

TABLE II. $Q_{i,i+3}/Q_{i,i+1}$ ratios for square well.

$Q_{i,i+3}/Q_{i,i+1}$	1st approx	2nd approx	Exact value
Q_{03}/Q_{01}	0.096	0.085	0.080
Q_{14}/Q_{12}	0.131	0.106	0.094
Q_{25}/Q_{23}	0.150	0.116	0.101
$Q_{3,6}/Q_{34}$	0.160	0.117	0.104
$Q_{4,7}/Q_{4,5}$	0.161	0.118	0.106
$Q_{5,8}/Q_{5,6}$	0.162	0.118	0.107
$Q_{6,9}/Q_{6,7}$	0.163	0.119	0.108
$Q_{7,10}/Q_{7,8}$	0.163	0.121	0.109
$Q_{8,11}/Q_{8,9}$	0.164	0.122	0.109

between successive approximations, n and $n-1$, is at least twice as large as the difference between the nth approximation and the true value. This semi-quantitative notion will be useful in studying other potentials, when we do not know the exact eigenvalues. We also note that the agreement becomes worse as we go to higher energy levels, but not very rapidly looking at the percent deviation. In Table II, we give ratios of matrix elements $Q_{i,i+3}/Q_{i,i+1}$ obtained for the second and third approximations, together with the exact values. The first time that we calculate the $i,i+3$ matrix elements (for the second approximation), we ignore all higher-order matrix elements which are especially important in Eq. (17) and Eq. (14), so the $i,i+3$ matrix element must take the place of all the higher-order matrix elements. In the third approximation, the $i,i+5$ matrix elements play this role [i.e., we overestimate them by neglecting $i,i\pm7$ matrix elements in Eq. (19)], and in the third approximation, the $Q_{i,i\pm3}/Q_{i,i\pm1}$ matrix element ratios are given fairly accurately ($\sim10\%$).

TABLE III. Quartic oscillator energy eigenvalues (in units of H_{00}).

Energy	1st approx	2nd approx	Numerical calculation[a]
H_{00}	1	1	1
H_{11}	3.66	3.58	3.58
H_{22}	7.21	7.02	7.04
H_{33}	11.25	10.97	10.98
H_{44}	15.71	15.31	
H_{55}	20.51	20.01	
H_{66}	25.62	24.99	
H_{77}	30.99	30.24	
H_{88}	36.61	35.72	
H_{99}	42.45	41.43	

[a] L. I. Shiff, Phys. Rev. 92, 766 (1953).

We have also calculated the energy eigenvalues of the potential Q^4 to two orders of approximation. Here we would expect the first-order approximation to be much more exact than in the case of the square well. The small difference in eigenvalues from the first- and second-order approximations indicates that this is the case, and from our examination of the rate of convergence of the square-well eigenvalues, we would estimate that the second-order calculation is good to $\sim1\%$ for the quartic oscillator. The energy eigenvalues of the quartic oscillator are given in Table III. In Table IV, we include the ratios of matrix elements for the quartic oscillator.

TABLE IV. Ratio of matrix elements for quartic oscillator.

$Q_{i,i+3}/Q_{i,i+1}$	1st approx
Q_{03}/Q_{01}	0.054
Q_{14}/Q_{12}	0.049
Q_{25}/Q_{23}	0.049
Q_{36}/Q_{34}	0.048
Q_{47}/Q_{45}	0.048
Q_{58}/Q_{56}	0.047
Q_{69}/Q_{67}	0.047
$Q_{7,10}/Q_{7,8}$	0.047
$Q_{8,11}/Q_{8,9}$	0.047

4. SUMMARY

In this paper we have developed a unified treatment for all potentials of the form Q^{2n} from $n=1$ to $n=\infty$. By using the square well as a test case, we obtain a feel for the accuracy of the treatment to various orders of approximation. The calculation of the eigenvalues of the potential Q^4 shows how rapidly the calculation converges as we approach $n=1$, which is the expected result as the treatment is based on the harmonic-oscillator Q matrix. It is also worth pointing out that the same techniques can be used to solve problems of the form

$$H=\tfrac{1}{2}\alpha Q^2+m^{-1}\beta P^m. \tag{21}$$

ACKNOWLEDGMENT

I would like to thank Dr. Gordon Goodman for several illuminating comments made in the course of this work.

Remarks on the Continued Fraction Calculation of Eigenvalues and Eigenvectors*

JEROME D. SWALEN†‡
Department of Chemistry, Harvard University, Cambridge, Massachusetts and Shell Development Company, Emeryville, California

AND

LOUIS PIERCE
Department of Chemistry, University of Notre Dame, Notre Dame, Indiana

(Received September 19, 1960)

For eigenvalue problems in which the secular determinant has tridiagonal form, e.g., the rigid asymmetric rotor; the secular equation may be written in the form $f(\lambda')=0$, where $f(\lambda')$ is a continued fraction and λ' an eigenvalue. Furthermore, if the secular problem is of nth order, then the continued fraction $f(\lambda')$ may be developed in n different ways. Since the eigenvalues are roots of a function $f(\lambda)$, it is convenient to find the eigenvalues by means of the Newton-Raphson iterative procedure. This requires that the derivative of $f(\lambda)$ with respect to $\lambda(f'(\lambda))$ be determined. An exact expression for $f'(\lambda)$ is derived and it is shown that $f'(\lambda')$ is in fact the norm of the eigenvector belonging to the eigenvalue λ'. A simple recursion formula, in continued fraction form, for the eigenvector elements is also derived. The Newton-Raphson procedure is further shown to be equivalent to the variational method for iterative calculation of eigenvalues. The former procedure has, however, the advantage of bypassing the necessity of solving a set of simultaneous equations. Advantage is taken of the relation between $f'(\lambda')$ and the eigenvector of λ' to formulate a reasonable criterion for choosing the best possible development of $f(\lambda)$ is order to avoid convergence to an undesired root of $f(\lambda)$.

I. INTRODUCTION

In certain quantum-mechanical problems, the natural basis for a matrix representation of the Hamiltonian operator leads to a matrix with tridiagonal form, i.e., a matrix with its only nonzero elements on and immediately above and below the principal diagonal. Two examples of such problems are the rigid asymmetric rotor[1] and Mathieu's[2] equation. When this special form of the Hamiltonian matrix occurs, a particularly efficient method for calculating eigenvalues, involving a continued fraction form of the secular equation, may be employed. While the continued fraction procedure has been used and described by several authors,[3] little consideration has been given to the vitally important problem of convergence. The present work is concerned principally with this convergence problem.

Except for small matrices, the continued fraction procedure does not allow explicit formulas to be given for the eigenvalues. In general, iterative procedures must be used, and most published methods make use of a first-order process. An exception is the work of Posener[4] on the rigid rotor in which an approximate Newton-Raphson procedure is used. Our objective here is to derive explicit formulas for *exact* application of the Newton-Raphson procedure[5] to Hermitian matrices of tridiagonal form. We also consider here the problems of calculating the eigenvectors and of choosing the best development of the continued fraction form of the secular equation (there are n ways of writing the secular equation in continued fraction form for an nth-order secular problem). The latter problem is of special importance because a poor choice of development may lead to slow convergence and often convergence to an undesired root of the secular equation.

The methods given here are closely related to Löwdin's interation-variation procedure.[6] Löwdin's method is essentially an application of the Newton-Raphson procedure to determination of eigenvalues of a general (not tridiagonal) Hermitian matrix. He further is able to show that his method is equivalent to the variational method (as is ours also). The advantage of our method is that the problem of solving simultaneous equations is bypassed. The limitation, of course, is that it is applicable only to tridiagonal matrices. However, Given's[7] has recently described a method for transforming a general Hermitian matrix into tridiagonal form, thus greatly increasing the domain of applicability of the continued fraction method.

II. THEORY OF THE CONTINUED FRACTION METHOD FOR EIGENVALUE PROBLEMS

A. The Continued Fraction Form of the Secular Equation

It is assumed, in what follows, that an orthonormal basis for the matrix representation of the Hamiltonian

* This work was made possible by support extended to Harvard University by the Office of Naval Research and by support extended to the University of Notre Dame by the National Science Foundation.

† Present address: Shell Development Company, Emeryville, California.

‡ National Science Foundation Fellow, 1955–56.

[1] G. W. King, R. M. Hainer, and P. C. Cross, J. Chem. Phys. 11, 27 (1943).

[2] R. W. Kilb, *Tables of Mathieu Eigenvalues and Eigenfunctions for Special Boundary Conditions*, copies available from Professor E. B. Wilson, Jr., Department of Chemistry, Harvard University; C. C. Lin and J. D. Swalen, Revs. Modern Phys. 31, 841 (1959).

[3] See, for example, M. W. P. Strandberg, *Microwave Spectroscopy* (John Wiley & Sons, Inc., New York, 1954).

[4] D. W. Posener, J. Chem. Phys. 24, 546 (1956).

[5] J. D. Swalen, thesis, Harvard University (1956).

[6] P. O. Löwdin, *Advances in Chemical Physics* (Interscience Publishers, Inc., New York, 1959), Vol. 2, pp. 270–273; and "An elementary iteration-variation method for solving the Schrödinger equation," Technical Note No. 11, Quantum Chemistry Group, Uppsala, Sweden (1958).

[7] W. Givens, J. Assoc. Comp. Mach. 4, 298 (1957), and Oak Ridge National Laboratory Report ORNL-1574, March 3, 1954.

operator has been so chosen so that the matrix $\mathbf{H}$ has tridiagonal form. Denoting the unitary transformation which diagonalizes $\mathbf{H}$ by $\mathbf{S}$, we have then the equation

$$\mathbf{S}^{+}\mathbf{H}\mathbf{S}=\mathbf{\Lambda}, \tag{1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix. For a particular eigenstate,

$$(\mathbf{H}-\lambda\mathbf{I})\mathbf{s}=0, \tag{2}$$

where λ is an element of $\mathbf{\Lambda}$, and $\mathbf{s}$ a column vector of $\mathbf{S}$. The determinental form of the secular equation is then explicitly

$$\begin{vmatrix} H_{11}-\lambda & H_{12} & & & \\ H_{21} & H_{22}-\lambda & H_{23} & & \\ & & & 0 & \\ & H_{k,k-1} & H_{k,k}-\lambda & H_{k,k+1} & \\ 0 & & & & \\ & & H_{n,n-1} & H_{nn}-\lambda & \end{vmatrix} =0. \tag{3}$$

The continued fraction form of the secular equation may be obtained by at least two distinct procedures. One of the methods (see Sec. IIC) involves the elimination of all but one of the components of $\mathbf{s}$ from the homogeneous linear system of equations (2). A single equation of compatibility having the desired continued fraction form is then obtained.

The second method consists of reduction of Eq. (3) to a determinant of order one by means of successive outer pivotal condensations. The condensation procedure has been described in detail by Strandberg[3] in connection with the asymmetric rotor problem. Briefly, if the rows and columns of Eq. (3) having $k'>k$ and $k'<k$ are removed by successive outer pivotal condensations, the following equation results

$$H_{kk}-\lambda+h_{k+1}^{2}/R_{k+1}(\lambda)+h_{k-1}^{2}/R_{k-1}(\lambda)=0. \tag{4}$$

The quantities $h_{k\pm1}$ and $R_{k\pm1}$ introduced in (4) simplify the notation and are defined as

$$h_{k\pm\alpha}^{2}=|H_{k\pm\alpha,k\pm\alpha\mp1}|^{2}, \tag{5a}$$

$$R_{k\pm\alpha}(\lambda)=\lambda-H_{k\pm\alpha,k\pm\alpha}-h_{k\pm\alpha\pm1}^{2}/R_{k\pm\alpha\pm1}(\lambda), \tag{5b}$$

where $\alpha=1, 2, 3, \cdots$.[8] Note that for an nth order secular determinant, there are n different forms of Eq. (4), corresponding to the n possible values of k.

[3] Throughout this paper the index k will be used to denote the particular development which is being employed, while α will be used as a running index. Note that α takes on only positive values. Wherever the subscript $k+\alpha$ occurs, the maximum value of α is $\alpha=n-k$; while where $k-\alpha$ occurs, the maximum value of α is $\alpha=k-1$. The peculiar limits for α are required because the quantities such as R_m or h_m are zero when $m<1$ or $m>n$, for a secular equation of order n. To avoid cumbersome expressions, the limits on α are not given in (5) or in any of the succeeding equations.

We shall refer to equations, such as Eq. (4), which involve the quantities $R_{k\pm\alpha}(\lambda)$ defined by the recursion relation (5b), as arising from the kth development of the continued fraction form of the secular equation or, more simply, as arising from the kth development. It will later be made evident that choice of development is a very important practical consideration.

B. Calculation of Eigenvalues

When the off-diagonal elements of H are small in comparison with the difference between diagonal elements, H_{kk} itself will be a good approximation to the kth eigenvalue of H. If Eq. (4) is rewritten as

$$\lambda=H_{kk}+h_{k+1}^{2}/R_{k+1}+h_{k-1}^{2}/R_{k-1}, \tag{6}$$

then the substitution of H_{kk} for λ on the right-hand side of Eq. (6) generates an improved approximation to the kth eigenvalue. This procedure may then be iterated until the desired accuracy is attained. However, when off-diagonal elements are not small, this simple first-order iterative procedure will in general be divergent or, at best, slowly convergent.

A particularly convenient method for dealing with this convergence problem is to employ the Newton-Raphson technique, which corresponds to an iterative procedure of the second order. We define a function $f_k(\lambda)$ as

$$f_k(\lambda)=\lambda-H_{kk}-h_{k+1}^{2}/R_{k+1}-h_{k-1}^{2}/R_{k-1}. \tag{7}$$

For λ' such that $f_k(\lambda')=0$, Eq. (7) reduces to Eq. (4), and thus the eigenvalues λ' of $\mathbf{H}$ are the roots of $f_k(\lambda)$. Note again that n different functions $f(\lambda)$ may be defined. Equation (7) gives explicitly the kth development of $f(\lambda)$. Suppose that $\lambda^{(m)}$ is the mth approximation to an eigenvalue of H; then the expansion of $f_k(\lambda)$ about $\lambda^{(m)}$ gives

$$f_k(\lambda)=f_k(\lambda^{(m)})+f_k'(\lambda^{(m)})[\lambda-\lambda^{(m)}]+\cdots. \tag{8}$$

Neglecting derivatives of $f_k(\lambda)$ of higher order than the first, we find that $f_k(\lambda)$ will be zero if the next approximation $\lambda^{(m+1)}$ is taken as

$$\lambda^{(m+1)}=\lambda^{(m)}-f_k(\lambda^{(m)})/f_k'(\lambda^{(m)}), \tag{9}$$

where

$$\begin{aligned} f_k'(\lambda)&=df_k(\lambda)/d\lambda \\ &=1+(h_{k+1}/R_{k+1})^{2}[1+(h_{k+2}/R_{k+2})^{2}(1+\cdots)] \\ &\quad+(h_{k-1}/R_{k-1})^{2}[1+(h_{k-2}/R_{k-2})^{2}(1+\cdots)]. \end{aligned} \tag{10}$$

In terms of $f_k(\lambda)$, the simple first-order procedure described above consists of repeated application of the formula

$$\lambda^{(m+1)}=\lambda^{(m)}-f_k(\lambda^{(m)}). \tag{11}$$

Since from (10) it is evident that $f_k'(\lambda)\geq1$, it follows from (9) that Eq. (11) is convergent only if $1\leq f_k'(\lambda)<2$.

Note that even when (11) is applicable, it is still necessary to calculate the R's. Thus, relatively little extra effort is required to calculate $f_k'(\lambda)$, and this effort is more than compensated for in speed of convergence afforded by use of Eq. (9).

When degeneracies occur, the continued fraction must be modified in order to prevent one of the R's from approaching zero and the continued fraction from violently diverging. If various eigenvalues are degenerate in the zero order, then a proper linear combination can lead to factoring of the secular equation. Usually the symmetry of the problem determines this factoring. For doubly degenerate eigenvalues in zero order, a symmetric and antisymmetric combination are used. In the asymmetric rotor problem, the Wang transformation is used.[1]

For cases of near degeneracies, which cannot conveniently be removed by a transformation, a very good first approximation to a root is needed in order to avoid a convergence to a wrong root. The use of the derivative is mandatory to avoid divergence.

C. Calculation of Eigenvectors

The use of the Newton-Raphson procedure is especially efficient when the eigenvectors as well as the eigenvalues are desired. The reason being that $f'(\lambda')$ is directly related to the eigenvector belonging to the eigenvalue λ'. To demonstrate this, we re-express the eigenvalue problem (2) in terms of a vector

$$\sigma_k = \begin{pmatrix} \vdots \\ \sigma_{k-1} \\ 1 \\ \sigma_{k+1} \\ \vdots \end{pmatrix}. \tag{12}$$

If σ_k is an unnormalized eigenvector of $\mathbf{H}$, with the kth element equal to unity, then the relation of σ_k to $\mathbf{s}$ is

$$\mathbf{s} = N_k \sigma_k, \tag{13a}$$

$$|N_k|^{-2} = \sigma_k^+ \sigma_k = 1 + \sum_\alpha [\sigma_{k+\alpha}^* \sigma_{k+\alpha} + \sigma_{k-\alpha}^* \sigma_{k-\alpha}]. \tag{13b}$$

In terms of σ_k the kth equation of relation (2) is

$$H_{k,k-1}\sigma_{k-1} + (H_{kk} - \lambda) + H_{k,k+1}\sigma_{k+1} = 0. \tag{14}$$

From the $(k+\alpha)$th and $(k-\alpha)$th equations of (2), we obtain the recursion formulas

$$\frac{\sigma_{k\pm\alpha}}{\sigma_{k\pm\alpha\mp1}} = \frac{-H_{k\pm\alpha,k\pm\alpha\mp1}}{H_{k\pm\alpha,k\pm\alpha} - \lambda + H_{k\pm\alpha,k\pm\alpha\pm1}(\sigma_{k\pm\alpha\pm1}/\sigma_{k\pm\alpha})},$$

which reduce in terms of the notation of IIB to

$$\sigma_{k\pm\alpha}/\sigma_{k\pm\alpha\mp1} = H_{k\pm\alpha,k\pm\alpha\mp1}/R_{k\pm\alpha}. \tag{15}$$

Note that: (1) The elements of σ_k defined by Eq. (15) are components of an eigenvector of $\mathbf{H}$ only when Eq. (14) is also satisfied. (2) The substitution of σ_{k-1} and σ_{k+1} from Eq. (15) into Eq. (14) yields the continued fraction form [Eq. (4)] of the secular equation.

Let us drop the requirement that σ_k satisfy Eq. (14) and define it through Eq. (15) alone. These equations, in general, have a solution for any value of λ. A detailed comparison of Eq. (15) and Eq. (10) shows that for any value of λ

$$f_k'(\lambda) = \sigma_k^+(\lambda)\sigma_k(\lambda), \tag{16}$$

and in particular for λ', an eigenvalue of $\mathbf{H}$, $f_k'(\lambda')$ is the norm of the eigenvector $\sigma_k(\lambda')$.

It is interesting to note that the Newton-Raphson procedure of IIB is completely equivalent to the variational method for solving secular equations.[6] To demonstrate this, suppose that $\lambda^{(m)}$ is the mth approximation to an eigenvalue of H. The approximate eigenvector $\sigma_k(\lambda^{(m)})$, corresponding to $\lambda^{(m)}$, may then be obtained from Eq. (15), setting $\sigma_k = 1$ and using $\lambda^{(m)}$ to evaluate the $R_{k\pm\alpha}$. According to the variational method, we obtain the next approximation $\lambda^{(m+1)}$

$$\lambda^{(m+1)} = \sigma_k^+(\lambda^{(m)})\mathbf{H}\sigma_k(\lambda^{(m)})/[\sigma_k^+(\lambda^{(m)})\sigma_k(\lambda^{(m)})]. \tag{17}$$

With the aid of Eqs. (14), (16), and (7), it is seen that Eq. (17) reduces to the kth development of the Newton-Raphson formula (9).

The relation between $f_k'(\lambda)$ and $\sigma_k(\lambda)$ may further be used to write a simple expression for the average value of any operator $\mathbf{\Pi}$ which is diagonal in the basis for the representation of $\mathbf{H}$ leading to Eq. (3). Thus if φ_k is the kth basis function and

$$\mathbf{\Pi}\varphi_k = \Pi_k \varphi_k, \tag{18}$$

then from Eq. (15) it can be shown that

$$\langle \Pi \rangle_{\lambda'} = [1/f_k'(\lambda')]\{\Pi_k + (h_{k+1}/R_{k+1})^2[\Pi_{k+1} + (h_{k+2}/R_{k+2})^2(\Pi_{k+2} + \cdots)] + (h_{k-1}/R_{k-1})^2 \times [\Pi_{k-1} + (h_{k-2}/R_{k-2})^2(\Pi_{k-2} + \cdots]\}, \tag{19}$$

where $\langle \Pi \rangle_{\lambda'}$ is the average value of $\mathbf{\Pi}$ for the eigenstate belonging to the eigenvalue λ'. This relation was first derived in a different manner for the special case of the rigid asymmetric rotor by Kivelson and Wilson.[9]

D. Choice of Development

We have made rather extensive use of the above described procedures in calculating the energy levels of the asymmetric rotor both with desk calculators and with electronic digital computers. In our experience we have always found the iterative procedure Eq. (9) to be convergent. However, we also found that convergence is not always to the desired root of $f(\lambda)$. This is often the case even when the starting approximation is quite good. Convergence to the wrong root is clearly a serious problem since considerable effort may be expended in obtaining an undesired eigenvalue. There seem to be only two controls over this problem: (1) The

[9] D. Kivelson and E. B. Wilson, Jr., J. Chem. Phys. **20**, 1975 (1952).

choice of initial approximation to the true eigenvalue and (2) the choice of n developments of $f(\lambda)$ for an nth-order secular problem. Controls (1) and (2) are really not independent. Simple considerations show that the quality of an approximation to an eigenvalue is strongly dependent on the particular choice of development of $f(\lambda)$.

Consider the general behavior of $f(\lambda)$ for an nth-order secular problem with n distinct eigenvalues. We have already noted that $f'(\lambda)$ is always greater than unity regardless of development. Thus, $f(\lambda)$ must cross the λ axis n times, and the n roots are separated by $n-1$ poles. Furthermore, if the k'th development has been used, the $n-1$ poles are located at the roots of $R_{k'+1}$ and $R_{k'-1}$. It is important to note that roots of $R_{k'\pm1}$ are the roots of a secular determinant of order $n-1$, and that this determinant is obtained from $|H_{k'k''}-\lambda\delta_{k'k''}|=0$ by deleting the k'th row and k'th column, or alternatively by omitting the k'th basis function in setting up the matrix of the Hamiltonian.

Suppose now that for a given eigenvalue λ' of H, that $s_{k'}\ll1$, where $s_{k'}$ is the k'th element of the eigenvector $\mathbf{s}$ belonging to λ'. From the above considerations, it is clear that if $f(\lambda)$ is developed about the k'th position, then one of the poles of $f_{k'}(\lambda)$ will very nearly coincide with the root λ' of $f_k(\lambda)$. (If $s_{k'}=0$ then the coincidence is of course exact.) Therefore, if $s_{k'}\ll1$ and the k'th development used, then the first approximation must indeed be very good, or else convergence to the wrong root will almost certainly occur.

Clearly, it is desirable to choose a development for which the poles of $f(\lambda)$ are as far removed from the root of interest as is possible. We have not been able to formulate an absolute criterion for choosing this "best development"; however, the following criterion seems to be quite reasonable:

For a given eigenvalue λ' of H, the "best development" is the kth if

$$f_k'(\lambda')\leq f_{k\pm\alpha}'(\lambda'),\quad \alpha=1,2,3\cdots. \tag{20}$$

If the condition (20) is satisfied, it follows that

$$\sigma_{k\pm\alpha}{}^*\sigma_{k\pm\alpha}\leq\sigma_k{}^*\sigma_k$$

for all values of α and this assures that

$$s_k{}^*s_k\geq1/n.$$

Furthermore, if Eq. (20) is satisfied, then the kth basis function, which is the largest contributor in the expansion of the true eigenfunction, makes no contribution to the roots of $R_{k\pm1}$.

In practice, the "best development," as defined by Eq. (20), is easily found if a reasonably good approximation λ^0 to the true eigenvalue λ' is available. One first develops $f(\lambda)$ about some arbitrary position, say k'. Then $f_{k'}'(\lambda^0)$ is evaluated in its fully expanded form

$$f_{k'}'(\lambda^0)=1+\sum_\alpha[\sigma_{k'+\alpha}{}^*(\lambda^0)\sigma_{k'+\alpha}(\lambda^0)+\sigma_{k'-\alpha}{}^*(\lambda^0)\sigma_{k'-\alpha}(\lambda^0)]. \tag{21}$$

Since λ^0 has been assumed to be a reasonably good approximation to λ', $\sigma_{k'}(\lambda^0)$ will be a reasonably good approximation to the true eigenvector. If the largest term in $f_{k'}'(\lambda^0)$ is $\sigma_{k'+j}{}^*\sigma_{k'+j}$, then it is clear that the best development is the kth where $k=k'+j$.

Choice of development is of importance in the eigenvector as well as eigenvalue calculation. King, Hainer, and Cross[1] have derived, for the case of the rigid asymmetric rotor, a two term recursion formula which is a polynomial in λ, as opposed to the continued function relation (15). It has been noted[10,11] that their method is often subject to large round-off errors. The reason for this is that the eigenvector elements are always calculated relative to the first element. In terms of the method of Sec. IIC, this corresponds to using only the first development. Should it happen that $s_1\ll1$ for the root λ', then the norm of σ will be very large due to one or more nearly vanishing denominators of $R_{1+\alpha}$. Clearly, it seems reasonable to expect that the use of the kth development, where $f_k'(\lambda')$ satisfies Eq. (20), will eliminate excessive round-off errors. In practice (the rigid asymmetric rotor problem), we have found this to be the case.

III. CONCLUSION

Most of the above described methods have been extensively applied by the authors and others[2,10,12] to the asymmetric rotor problem and the Mathieu problem. Compared to previously published methods, they have proven to be very efficient indeed. However, most of these applications have been to nondegenerate eigenvalues. While we have not investigated the case of degenerate or nearly degenerate eigenvalues in much detail, the work of Herschbach and Swalen[12] indicates a special technique which, in some cases, may be useful when the present methods fail. The reader is referred to their paper for their variation of the continued fraction method, which is designed to avoid large round-off errors in the calculations of nearly degenerate energy levels for internal and over-all molecular rotation.

ACKNOWLEDGMENTS

We wish to gratefully acknowledge discussions on this work with Professor E. Bright Wilson, Jr. We also thank Professor Harrison Shull, Dr. T. W. Evans, and Professor Karl Kronstein for discussions concerning various aspects of this problem.

[10] R. Schwendeman and V. Laurie, *Line Strengths for the Rigid Asymmetric Rotor* (Pergamon Press, New York, 1958). V. Laurie, thesis, Harvard University (1957). Schwendeman and Laurie were to our knowledge first to use recursion formulas in continued fraction form for the rigid rotor eigenvector problem; they essentially used the equations preceding Eq. (15) but apparently did not notice that they could be reduced to the simpler form (15).

[11] G. Erlandsson, Arkiv Fysik **16**, 181 (1959).

[12] D. R. Herschbach and J. D. Swalen, J. Chem. Phys. **29**, 761 (1958).

Calculation of the Eigenvalues of a Tridiagonal Hermitian Matrix*

LOUIS PIERCE
Department of Chemistry, University of Notre Dame, Notre Dame, Indiana
(Received September 19, 1960)

For real symmetric or Hermitian matrices with tridiagonal form, the secular equation may be written as a continued fraction equation $f(\lambda)=0$. $f(\lambda)$ is a member of a recursively defined sequence $\mathfrak{R}^{(n)}(\lambda)$ of n continued fractions if the secular equation is of the nth order. The basis for a new method of computing the eigenvalues of such tridiagonal matrices is given. The method requires the determination of an integer-valued function $P_n(\gamma)$ for a succession of values of γ, where $P_n(\gamma)$ is a function only of n and the signs of the n terms in $\mathfrak{R}^{(n)}(\gamma)$.

THE preceding paper[1] dealt with the application of the Newton-Raphson procedure to the problem of finding the eigenvalues and eigenvectors of a tridiagonal Hermitian matrix, when the secular equation is cast into continued fraction form. While in general the methods described in I are very efficient, they are seriously lacking in the case of nearly-degenerate eigenvalues. The difficulty stems from the fact that the continued fraction $f(\lambda)$ has nearly coinciding poles and roots in cases of near-degeneracy. Givens[2] has derived a procedure, employing the polynomial form of the secular equation, for which near-degeneracy presents no special problems. His method is based on the fact that the leading principal minors of $|\mathbf{H}-\lambda\mathbf{I}_n|$ form in general a Sturm sequence $g_0(\lambda)$, $g_1(\lambda)$, $\cdots$, $g_n(\lambda)$ of polynomials when $\mathbf{H}$ is tridiagonal and Hermitian. The sequence can be recursively defined, and knowing it to be a Sturm sequence, it is possible to determine the number of eigenvalues α_i of $\mathbf{H}$ such that $\alpha_i<\gamma$ simply by counting the number of changes in sign in the sequence $g_0(\gamma)$, $g_1(\gamma)$, $\cdots g_n(\gamma)$.

The purpose of this note is to show that the recursively defined sequence of continued fractions, associated with the continued fraction form of $|\mathbf{H}-\lambda\mathbf{I}_n|=0$, has a property whose consequences are just as powerful as the properties of a Sturm sequence of polynomials. In order to demonstrate this property, it will be convenient to use a notation slightly different from that of I. The form of the secular determinant is taken as

$$|\mathbf{H}-\lambda\mathbf{I}_n|=\begin{vmatrix} H_{11}-\lambda & H_{12} & 0 & & \\ H_{21} & H_{22}-\lambda & H_{23} & & \\ & \ddots & \ddots & \ddots & \\ & H_{k,k-1} & H_{kk}-\lambda & H_{k,k+1} & \\ & & \ddots & \ddots & \ddots \\ 0 & & & H_{n,n-1} & H_{nn}-\lambda \end{vmatrix}=0. \tag{1}$$

Th kth development of the continued fraction form of (1) is

$$f_k(\lambda)=\lambda-H_{kk}-h_{k+1}^2/R_{k,k+1}-h_{k-1}^2/R_{k,k-1}=0, \tag{2}$$

where[3]

$$h_{k\pm\alpha}=|H_{k\pm\alpha,k\pm\alpha+1}|, \tag{3}$$

$$R_{k,k\pm\alpha}=\lambda-H_{k\pm\alpha,k\pm\alpha}-h_{k\pm\alpha\pm1}^2/R_{k,k\pm\alpha\pm1}. \tag{4}$$

It was proved in I that $f_k'(\lambda)=df_k(\lambda)/d\lambda\geqq1$. Thus, if $f_k(\lambda)$ has n distinct roots (we consider here only this case), these roots are separated by $n-1$ poles of $f_k(\lambda)$, and the $n-1$ poles are furthermore the roots of $R_{k,k+1}(\lambda)\times R_{k,k-1}(\lambda)$. Comparison of (2) and (4) makes it clear that $R_{k,k\pm\alpha}'(\lambda)\geqq1$. Thus the roots of $R_{k,k\pm\alpha}(\lambda)$ are separated by the roots of $R_{k,k\pm\alpha\pm1}(\lambda)$. This important property of $f(\lambda)$ and the $R(\lambda)$'s makes it possible to prove inductively the following theorem:

Let $\mathfrak{R}_{kk}^{(n)}(\lambda)$ represent the n-term sequence of continued fractions

$$f_k(\lambda), R_{k,k+1}(\lambda), R_{k,k+2}(\lambda), \cdots, R_{k,n}(\lambda), R_{k,k-1}(\lambda), R_{k,k-2}(\lambda), \cdots, R_{k,1}(\lambda).$$

Then, given any real number γ such that no term in the sequence $\mathfrak{R}_{kk}^{(n)}(\gamma)$ is zero, the number of roots α_i of

* This research was made possible by a grant from the National Science Foundation.

[1] J. D. Swalen and L. Pierce, J. Math. Phys. 2, 736 (1961), hereafter referred to as I.

[2] W. Givens, *Numerical Computation of the Characteristic Values of a Real Symmetric Matrix*, 1953 (unpublished).

[3] Whenever $k+a$ appears, a is limited to the values 1, 2, $\cdots$, $n-k$, and whenever $k-a$ appears, a is limited to the values 1, 2, $\cdots$, $k-a$.

$f_k(\lambda)$ *such that* $\alpha_i<\gamma$ *is given by*

$$P_n(\gamma)=[n+n_+(\gamma)-n_-(\gamma)]/2,$$

where $n_+(\gamma)$ *is the number of positive terms, and* $n_-(\gamma)$ *the number of negative terms in the sequence* $\mathfrak{R}_{kk}^{(n)}(\gamma)$. The proof of the theorem is given in three steps, the cases of $k=1$, $k=n$, and $k\neq 1$, n being separately considered. For the proof, it is convenient to define an l-term sequence $\mathfrak{R}_{k,k+\alpha}^{(l)}$ and an m-term sequence $\mathfrak{R}_{k,k-\alpha}^{(m)}$ as

$$\mathfrak{R}_{k,k+\alpha}^{(l)}=R_{k,k+\alpha},\ R_{k,k+\alpha+1},\ \cdots,\ R_{k,n}\cdot l=n-k-\alpha+1;\quad \alpha=1,2,\cdots,n-k. \tag{5}$$

$$\mathfrak{R}_{k,k-\alpha}^{(m)}=R_{k,k-\alpha},\ R_{k,k-\alpha-1},\ \cdots,\ R_{k,1}\cdot m=k-\alpha;\quad \alpha=1,2,\cdots,k-1.$$

The theorem as stated applies to $\mathfrak{R}_{kk}^{(n)}(\lambda)$ as regards the roots of its leading term $f_k(\lambda)$. However, as will be made evident, it applies as well to $\mathfrak{R}_{k,k+\alpha}^{(l)}$ as regards the roots of its leading term $R_{k,k+\alpha}$ and to $\mathfrak{R}_{k,k-\alpha}^{(m)}$, as regards the roots of $R_{k,k-\alpha}$.

Case I. $k=1$: Let $\alpha_i(<\alpha_{i+1})$ denote the ith root of $f_1(\lambda)$, and $\beta_i(<\beta_{i+1})$ denote the ith root of $R_{12}(\lambda)$. Then,

$$\alpha_1<\beta_1<\alpha_2\cdots<\alpha_j<\beta_j<\alpha_{j+1}<\beta_{j+1}<\alpha_{j+2}\cdots\beta_{n-1}<\alpha_n. \tag{6}$$

$R_{12}(\lambda)$, of course, is the continued fraction form of a secular problem of order $n-1$, related to the problem of interest by striking out the first row and column of the secular determinant. Assume the theorem to be true for the sequence $\mathfrak{R}_{12}^{(l)}(\lambda)$ [as regards the roots of its leading term $R_{12}(\lambda)$], where $l=n-1$. If γ is chosen such that $\beta_j<\gamma<\beta_{j+1}$, then

$$P_l(\gamma)=[l+l_+(\gamma)-l_-(\gamma)]/2=j.$$

Since $\alpha_j<\gamma<\alpha_{j+2}$, it is clear that the number of $\alpha_i<\gamma$ is either j or $j+1$. Remembering that $f_k'(\lambda)\geqq 1$, it follows that there are j roots $\alpha_i<\gamma$ if $f_1(\lambda)<0$, and $j+1$ roots $\alpha_i<\gamma$ if $f_1(\lambda)>0$. Furthermore, as regards $\mathfrak{R}_{11}^{(n)}$,

$$P_n(\gamma)=\{(l+1)+[l_+(\gamma)+1]-l_-(\gamma)\}/2=j+1$$

if $f_1(\gamma)>0$, and

$$P_n(\gamma)=\{(l+1)+l_+(\gamma)-[l_-(\gamma)+1]\}/2=j$$

if $f_1(\gamma)<0$. The theorem is easily proved true for the sequences $\mathfrak{R}_{1,n}^{(1)}$, $\mathfrak{R}_{1,n-1}^{(2)}$, $\mathfrak{R}_{1,n-2}^{(3)}$ and is thus true for $\mathfrak{R}_{11}^{(n)}$.

Case II. $k=n$: The nth development of $f(\lambda)$ from the secular determinant (1) is clearly equivalent to the first development ($k=1$) of $f(\lambda)$ from a secular determinant related to (1) by reverse ordering of rows and columns. Thus, the proof for Case I applies equally to Case II.

Case III. $k\neq 1, n$: Clearly, since the theorem is true, for $\mathfrak{R}_{11}^{(n)}$ and $\mathfrak{R}_{1,1+\alpha}^{(l)}$, it must be true for $\mathfrak{R}_{k,k+\alpha}^{(l)}$ [as regards the roots of $R_{k,k+1}(\lambda)$]; and since it is true for $\mathfrak{R}_{nn}^{(n)}$ and $\mathfrak{R}_{n,n-\alpha}^{(m)}$, it must be true for $\mathfrak{R}_{k,k-1}^{(m)}$ [as regards the roots of $R_{k,k-1}(\lambda)$]. Note for $\alpha=1$, $l+m=n-1$. Let $\alpha_i(<\alpha_{i+1})$ denote the ith root of $f_k(\lambda)$, and $\beta_i(<\beta_{i+1})$ the ith root of $R_{k,k+1}(\lambda)\times R_{k,k-1}(\lambda)$. The inequality (6) then applies. If γ is chosen as above (Case I), then

$$P_l(\gamma)+P_m(\gamma)=\{(l+m)+[l_+(\gamma)+m_+(\gamma)]-[l_-(\gamma)+m_-(\gamma)]\}/2=j.$$

If $f_k(\gamma)>0$, it follows that there are $j+1$ roots $\alpha_i<\gamma$ and that $P_n(\gamma)=j+1$; while if $f_k(\gamma)<0$, there are j roots $\alpha_i<\gamma$ and $P_n(\gamma)=j$. Thus, the theorem is true for $\mathfrak{R}_{kk}^{(n)}$, as regards the roots of $f_k(\lambda)$, for $k=1,2,\cdots,n$.

The utility of the theorem is obvious. It can be used to "localize" or "trap" any root of the secular equation, and the size of the interval in which a root has been trapped can be narrowed, in principle, indefinitely by computing $P_n(\gamma)$ for a succession of values of γ. Furthermore, with the "trapping" procedure, unlike the Newton-Raphson procedure, there is no danger of converging to an undesired eigenvalue. Since trapping requires only that the signs of $f(\lambda)$ and $R(\lambda)$ be properly determined, this method is less sensitive to round-off errors than is the Newton-Raphson procedure. Thus, it is to be preferred over the Newton-Raphson procedure when calculating nearly-degenerate eigenvalues.

With the digital computer, it is primarily the number of multiplications which determines the time required for a given computation. The calculation of a single sequence $\mathfrak{R}_{kk}^{(n)}(\gamma)$ requires $n-1$ multiplications (squaring of the $h_{k\pm\alpha}$ is not counted since this is done once and for all at the start of the computation). The present method is, of course, very closely analogous to the Givens method. The only important difference is that Givens uses the polynomial form of the secular equation, while we use the continued fraction form. Insofar as the criterion of "number of multiplications" is concerned, a definite comparison of the two methods can be made; the present method is twice as "good" since a single computation of a Sturm sequence require $2(n-1)$ multiplications, as opposed to only $n-1$ multiplications for the corresponding continued fraction sequence.

ACKNOWLEDGMENTS

The author is grateful to Professor Karl Kronstein for several stimulating discussions of this work and related problems, and to Dr. J. D. Swalen and Bernard Kirtman for their critical comments on a preliminary draft of this paper.

Erratum: Contraction of Lie Groups

[J. Math. Phys. 2, 1 (1961)]

EUGENE J. SALETAN
Rutgers University, New Brunswick, New Jersey

IN my paper[1] it was stated incorrectly that group contraction was first introduced by Inonü and Wigner.[2] In fact, these authors, as they stated in a footnote, specialize in a concept of the limits of Lie algebras introduced first by Segal.[3] My paper deals only with this specialization. I take this opportunity to make the necessary correction.

[1] E. J. Saletan, J. Math. Phys. **2**, 1 (1961).
[2] E. Inonü and E. P. Wigner, Proc. Natl. Acad. Sci. U. S. **39**, 510 (1953).
[3] I. E. Segal, Duke Math. J. **18**, 221 (1951).