JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 2, NUMBER 2 MARCH–APRIL, 1961

Quantization of Fields with Infinite-Dimensional Invariance Groups*

BRYCE S. DEWITT
Institute of Field Physics, Department of Physics, University of North Carolina, Chapel Hill, North Carolina

A general approach to the problems of quantizing fields which have infinite-dimensional invariance groups is given. Space and time are treated on a completely equal footing. A Poisson bracket is defined by means of Green's functions, independently of the discovery or recognition of canonical variables, and is shown to satisfy all the usual identities. In accordance with the measurement theoretical foundations of the quantum theory, the Poisson bracket (i.e., commutator) is defined only for physically measurable group invariants. The Green's functions give a direct description of the propagation of small disturbances arising from a pair of mutually interfering measurements.

In order to establish a correspondence between this approach and conventional canonical theory, a motivation for the adopted definition of the Poisson bracket is outlined with the aid of the fundamental theorem of canonical transformation theory. The rest of the discussion is logically independent of this, however. The general theory of "wave operators" and their associated Green's functions is briefly reviewed. Specific details connected with the group theoretical side of the theory are handled in such a way that problems of constraints are completely avoided. In the last section the general method is applied to the Yang-Mills field, as a nontrivial example. The problem of factor ordering is not studied.

INTRODUCTION

THE development of the quantum theory of any field, or set of interacting fields, whose dynamical equations remain invariant under an infinite-dimensional group of transformations is always beset with special problems having no counterparts in simpler field theories. Historically, the first example of this situation was provided by the electromagnetic field and its associated group of gauge transformations. The problems of gauge invariance have by now been studied exhaustively, and workable techniques have been developed for dealing with them in a variety of quantum contexts. Moreover, the electromagnetic field itself has provided a prototype to which all generalizations of these techniques to other fields have been applied as a check.

Experience has shown, however, that the electromagnetic field is probably not a very good prototype. Its invariance group, being Abelian, is too simple to indicate those generalizations which are likely to reveal the group theoretical structure of more complicated theories. Because of the consequent absence of a clear route to follow, effort has tended to spread out in various directions. One line of investigation has been to focus on the problem of constraints, to which the existence of the invariance group in question gives rise but in the study of which the group itself plays a minimal role. The trouble with the problem of constraints is that its formalization has always necessitated a falling back on the canonical fundamentals of a Hamiltonian or quasi-Hamiltonian theory.[1] The resultant asymmetry in treatment of space and time coordinates does not generally fit comfortably with the invariance group, the parameters of which are space-time functions having no temporal prejudices. This is especially true in a generally covariant theory in which the gravitational field is involved,[1] but it is also true in other cases. It is even true for electromagnetism, as the history of the subject shows.

It is the purpose of this paper to present the outlines of a general approach to these problems which dispenses entirely with Hamiltonian ideas and treats space-time in a completely homogeneous fashion. The basis for

* This research was supported in whole under contract by the U. S. Air Force monitored by the AF Office of Scientific Research of the Air Research and Development Command.

[1] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); Proc. Roy. Soc. (London) **A246**, 333 (1958); S. Deser and R. Arnowitt, Phys. Rev. **113**, 745 (1959).

this approach is a definition of the classical Poisson bracket by means of Green's functions, which is independent of any definitions of pairs of conjugate variables and which is, in effect, a straightforward extension of a definition originally proposed by Peierls.[2] The point of view is here adopted that Poisson brackets (i.e., commutators) should be defined only between group invariants. This automatically eliminates the need for subsidiary conditions, which have always to be specially tailored to each individual theory and which have proved so often bothersome in the past. Furthermore, this approach is in accord with the foundations of the quantum theory as expressed in the general theory of measurement. Real physical measurements can be performed only on group invariant quantities, and the interference between two measurements which, via the uncertainty principle, in effect *defines* the commutator, is most immediately described not in terms of canonically conjugated variables at a given instant, but in terms of the Green's functions which express the laws of propagation of small disturbances and which satisfy certain fundamental reciprocal relations. In quantum electrodynamics this role of the Green's functions was demonstrated at a very early date in the classic paper of Bohr and Rosenfeld,[3] which made no use of the Lorentz or any other gauge condition. The present paper may be viewed as the first step in an analogous demonstration for the general field theoretic case.

Section 1 consists of an introductory discussion, in general terms, of differential "wave operators" and their associated Green's functions. In Sec. 2 canonical transformation theory is introduced in order to arrive at an appropriate definition for the Poisson bracket in an arbitrary field theory. Group theoretical details are examined in Sec. 3, and, with the proof that all the usual properties, including the Poisson-Jacobi identity, are satisfied by the suggested Poisson bracket, it is shown that the definition does not really depend for its consistency on the canonical arguments of the preceding section. In the form of the commutator, in fact, the definition can be justified by appealing solely to the uncertainty principle and the theory of measurement. Finally, in Sec. 4, the methods of the earlier sections are applied to a specific example. Since the author has already given elsewhere[4] a preliminary account of the application of these methods to the quantization of the gravitational field, a different example, the Yang-Mills field,[5] is chosen here.

It will be seen in the course of these derivations that the group theoretical side of the theory still plays a relatively small role in spite of the fact that the derivations are now fully "covariant" in the sense that space and time are placed on a completely equal footing. In the author's opinion, however, the role of the invariance group can be expected to increase when a study is undertaken of purely quantum problems, such as the factor ordering ambiguity which is not considered here. For, as has been emphasized by Klein,[6] the quantum theory can be regarded in many ways as a theory of the infinite dimensional unitary representations of the invariance groups which characterize the physical system under consideration.

1. WAVE OPERATORS AND GREEN'S FUNCTIONS

The propagation of a small disturbance, whether in a set of interacting fields, or in an elastic physical medium, or in a collection of bodies interacting via the laws of celestial mechanics, is described by a linear differential equation of finite order, usually not higher then the second. Let us denote the dynamical variables appearing in such an equation by symbols such as ψ^i, ϕ^a. These variables will be functions of one or more continuous parameters, or "coordinates." For definiteness we shall regard them as functions of four space-time coordinates x^μ. It will be obvious, however, that everything we say will be equally applicable to theories with either more or fewer parameters, in particular to systems having only a finite number of degrees of freedom, with "time" as the single parameter.

Different points of space-time will be distinguished by means of primes: x, x', x'', etc. For compactness the point at which a given variable, such as ψ^i, is evaluated will be indicated by affixing primes to the index appearing on the variable, e.g., $\psi^{i''}$. For economy in the use of primes, the symbol z will also sometimes be used in place of x to designate a point in space time. Lower case Latin letters from the beginning of the alphabet $(a,b,c,\cdots)$ will always be associated with the symbol z, while those from the middle of the alphabet $(i,j,k,\cdots)$ will be associated with the symbol x.

It will be convenient to express the differential operator appearing in the propagation equation formally as a continuous matrix $F_{ij'}$, the equation itself taking the homogeneous form

$$\int F_{ij'}\psi^{j'}d^4x'=0. \tag{1.1}$$

Typically $F_{ij'}$ will be a linear combination of first or second derivatives of the delta function of $x-x'$, with coefficients which may be functionals of some or all of the field variables appearing in the theory, for example involving these variables together with their derivatives up to some small (first or second) finite order.

The $F_{ij'}$ will generally have three characteristic

[2] R. E. Peierls, Proc. Roy. Soc. (London) **A214**, 143 (1952).
[3] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **12**, 8 (1933).
[4] B. S. DeWitt, Phys. Rev. Letters **4**, 317 (1960).
[5] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954). See also R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[6] O. Klein, "Quantum theory and relativity," essay in *Niels Bohr and the Development of Physics*, edited by W. Pauli (McGraw-Hill Book Company, Inc., New York, 1955).

properties: (1) the property of being self-adjoint; (2) the property that well-behaved nonvanishing solutions of Eq. (1.1) really exist; and (3) the possession of unique "retarded" and "advanced" Green's functions. The property of being self-adjoint is characterized by the existence of a matrix operator $f^{\mu}{}_{ij'}$ such that

$$\int (\phi^i F_{ij'}\psi^{j'}-\psi^i F_{ij'}\phi^{j'})d^4x' = \frac{\partial}{\partial x^\mu}\int (\phi^i f^{\mu}{}_{ij'}\psi^{j'}-\psi^i f^{\mu}{}_{ij'}\phi^{j'})d^4x' \quad (1.2)$$

for arbitrary ϕ^i and ψ^i. An immediate corollary of Eq. (1.2) is

$$\int d^4x \int d^4x'(\phi^i F_{ij'}\psi^{j'}-\psi^i F_{ij'}\phi^{j'})d^4x'=0 \quad (1.3)$$

for fields ϕ^i, ψ^i which vanish sufficiently rapidly in remote regions of space time, which, in view of the arbitrariness of ϕ^i and ψ^i, may be expressed simply as

$$F_{ij'}=F_{j'i}. \quad (1.4)$$

The Green's functions associated with $F_{ij'}$ are characterized by the equation

$$\int F_{ik''}G^{\pm k''j'}d^4x''=-\delta_i{}^{j'}, \quad (1.5)$$

with the conditions

$$\begin{aligned} G^{-ij'}&=0 \quad \text{for} \quad x<x', \\ G^{+ij'}&=0 \quad \text{for} \quad x>x'. \end{aligned} \quad (1.6)$$

Here the symbol $\delta_i{}^{j'}$ denotes in obvious fashion a product of a Kronecker delta with a delta function, while "$<$" is an abbreviation for "lies to the past of" and "$>$" is an abbreviation for "lies to the future of," well-defined motions of "past" and "future" being assumed to exist in the space time of parameters x^μ. Typically these notions will be based on a *metric* of signature $-+++$ which may or may not, itself, belong to the category of "field variables." In this case there will generally be a region of overlap of the domains past and future which may be removed by relabeling it the "present." Both $G^{-ij'}$ and $G^{+ij'}$ vanish simultaneously in this region. In a nonrelativistic theory the "present" may collapse to a hypersurface.

It is to be noted that $F_{ij'}$, unlike an ordinary finite matrix, does not possess a unique inverse; both $G^{-ij'}$ and $G^{+ij'}$, as well as linear combinations of the two, are its "inverses." This fact is a direct consequence of the existence of well behaved solutions to Eq. (1.1), which may always be added to any "inverse."[7] On the other hand, it is to be recalled that although solutions to "wave equations" [here collectively described by Eq. (1.1)] are everywhere bounded (which is what is meant by "well-behaved") they do not vanish in remote regions of space-time sufficiently rapidly to be normalizable. For if they did, then $F_{ij'}$ would possess true null eigenvectors and have no inverses at all. Operators having the three properties enumerated above will be called *wave operators*. A continuous matrix having a single unique inverse, on the other hand, will be called *regular*. The undifferentiated delta function is the simplest example of a regular matrix.

A very important property of the Green's functions $G^{\pm ij'}$ is their ability, in the combination

$$G^{ij'}\equiv G^{+ij'}-G^{-ij'}, \quad (1.7)$$

to express *Huygens principle* for a disturbance ψ^i:

$$\psi^i=-\int d\Sigma_{\mu'}\int d^4x'' \times (G^{j'i}f^{\mu'}{}_{j'k''}\psi^{k''}-\psi^{j'}f^{\mu'}{}_{j'k''}G^{k''i}). \quad (1.8)$$

Here the value at an arbitrary point x of a function ψ^i satisfying Eq. (1.1) is expressed in terms of *Cauchy data*,[8]

$$\psi^i \quad \text{and} \quad \int f^{\mu}{}_{ij'}\psi^{j'}d^4x',$$

on a hypersurface Σ (surface element $d\Sigma_{\mu'}$) all points of which lie in the "present" with respect to one another, and which is customarily referred to as "spacelike." The proof of Eq. (1.8) is carried out by changing the surface integral into a volume integral with the aid of Gauss' theorem, and then using Eq. (1.2). For $x>\Sigma$ Eq. (1.8) becomes

$$\psi^i=\int_{\Sigma}^{\text{future}} d^4x' \int d^4x'' \times (G^{+j'i}F_{j'k''}\psi^{k''}-\psi^{j'}F_{j'k''}G^{+k''i}), \quad (1.9)$$

while for $x<\Sigma$ it becomes

$$\psi^i=\int^{\Sigma} d^4x' \int d^4x'' \times (G^{-j'i}F_{j'k''}\psi^{k''}-\psi^{j'}F_{j'k''}G^{-k''i}), \quad (1.10)$$

the validity of both forms following immediately from Eqs. (1.1) and (1.5). The extension of the domains of integration arbitrarily far into the future and past respectively is permitted since the Green's functions in each case "cutoff" sharply beyond the point x. In the case of x lying *on* Σ, the singularities of the Green's functions are to be interpreted in such a way that

[7] In theories for which Fourier transforms may be introduced (e.g., differential equation with constant coefficients), the existence of well-behaved solutions is revealed in the presence of poles on the real axis in the "energy plane."

[8] The operator $f^{\mu}{}_{ij'}$ is of one differential order lower than $F_{ij'}$. If $F_{ij'}$ is already of the first order, then there is no distinction between the Cauchy data ψ^i and $\int f^{\mu}{}_{ij'}\psi^{j'}d^4x'$.

$G^{j'i}$ vanishes while $\int f^{u'}{}_{j'k''}G^{k''i}d^4x''$ has the form of a three-dimensional delta function.[9]

The function $G^{ij'}$ is known as the *propagation function* for the disturbance ψ^i. It satisfies the homogeneous equation

$$\int F_{ik''}G^{k''j'}d^4x''=0. \tag{1.11}$$

Because ψ^i satisfies the homogeneous Eq. (1.1) and because the Cauchy data on Σ may be chosen completely arbitrarily, we may infer from Eq. (1.8) that $G^{ij'}$ satisfies not only Eq. (1.11) but also the equation

$$\int F_{ik''}G^{j'k''}d^4x''=0. \tag{1.12}$$

However, since there is only one unique function with the kinematical properties of $G^{ij'}$ which satisfies this equation, namely, the negative transpose of $G^{ij'}$, we infer from this the laws of reciprocity

$$G^{ij'}=-G^{j'i}, \tag{1.13}$$

$$G^{\pm ij'}=G^{\mp j'i}, \tag{1.14}$$

which in turn permit us to rewrite Huygens principle in the form

$$\psi^i=\int_\Sigma d\Sigma_{\mu'}\int d^4x''$$
$$\times(G^{ij'}f^{\mu'}{}_{j'k''}\psi^{k''}-\psi^{j'}f^{\mu'}{}_{j'k''}G^{ik''}). \tag{1.15}$$

We note incidentally, from Eq. (1.14), that $F_{ij'}$ possesses among its various inverses a symmetric inverse

$$\bar{G}^{ij'}=\tfrac{1}{2}(G^{+ij'}+G^{-ij'})=\bar{G}^{j'i}, \tag{1.16}$$

as befits a symmetric (i.e., self-adjoint) operator.

We end this section by recording for later use the following identities:

$$G^{+ij'}=\theta(x',x)G^{ij'}, \tag{1.17}$$

$$G^{-ij'}=-\theta(x,x')G^{ij'}, \tag{1.18}$$

where

$$\theta(x,x')=\begin{cases}1 & \text{when} \quad x>x'\\ 0 & \text{when} \quad x<x'.\end{cases} \tag{1.19}$$

Also,

$$\begin{aligned}&G^{+ia}G^{+b'j'}-G^{-ia}G^{-b'j'}\\ &\quad=(G^{+ia}-G^{-ia})G^{+b'j'}+G^{-ia}(G^{+b'j'}-G^{-b'j'})\\ &\quad=[\theta(x',z')-\theta(x,z)]G^{ia}G^{b'j'}.\end{aligned} \tag{1.20}$$

[9] This rule requires modification when $F_{ij'}$ is of the first differential order. The two terms in the integrand of Eq. (1.8) then make identical contributions, and both share the delta-function property when x lies on Σ.

2. DEFINITION OF THE POISSON BRACKET

Consider now a field ψ^i whose dynamical properties are specified by an action functional S. The field equations which it satisfies may be written in the form

$$S_{,i}=0, \tag{2.1}$$

where the comma followed by an index is used to denote the *variational derivative* with respect to ψ^i at a point. Our approach to the Poisson bracket for this field will be based on canonical transformation theory and on the recognition of the action as the generator of a finite canonical transformation. Here it is necessary to keep in mind that the functional S appearing in Eq. (2.1) is the over-all space-time action which connects dynamical variables in the remote past with those in the remote future. We may make this explicit by writing

$$S=S(\psi_\infty|\psi_{-\infty}), \tag{2.2}$$

where the symbols ψ_∞ and $\psi_{-\infty}$ designate any set of field quantities associated with the remote future and past respectively which suffice, without redundancy, to determine the "history" of the field. In this section we *assume* that such quantities exist without, however, at any time having need either to find them explicitly or to make other than symbolic use of them.

In order to describe the dynamics of the field ψ^i in canonical terms it is necessary to "break into" the action at an arbitrary spacelike hypersurface Σ, and to express it as the sum of two parts:

$$S=S(\psi_\infty|\psi_\Sigma)+S(\psi_\Sigma|\psi_{-\infty}). \tag{2.3}$$

Here the symbol ψ_Σ designates the same quantities as ψ_∞ and $\psi_{-\infty}$, taken, however, on the hypersurface Σ and determined by the stationary action principle (2.1) from the boundary conditions expressed by ψ_∞ and $\psi_{-\infty}$. It is to be noted that in a generally covariant theory involving the gravitational field, the hypersurface Σ must itself be specified in terms of field variables, since the definition of spacelike then depends on the metric.

We now invoke the fundamental theorem of canonical transformation theory:

> *The variation in the functional form of the generator of a finite canonical transformation, due to independent infinitesimal canonical transformations of its arguments, is equal to the difference between the corresponding independent infinitesimal generators.*[10]

Nearly every important theorem in classical mechanics is a corollary of this one. In order to apply it to the characterization of the Poisson bracket we consider the following change in the action:

$$S\rightarrow S+\epsilon B, \tag{2.4}$$

[10] A general proof of this theoren will be found in a previous paper by the author: Bryce S. DeWitt, Revs. Modern Phys. **29**, 377 (1957).

where ϵ is an infinitesimal constant and B is an arbitrary functional of the field variables which is invariant under all infinite dimensional transformation groups possessed by the theory. This change, which may be regarded as a comparison between two slightly different physical systems, will induce a change in the dynamical variables, the precise nature of which depends upon the boundary conditions selected. For example, we may adopt advanced boundary conditions in which the dynamical states of the two systems are taken to coincide in the remote future. Since both the original and modified actions, when broken up as in Eq. (2.3), generate canonical transformations describing the unfolding-in-time of their respective "histories," it is evident that the dynamical variables of the two systems on any space-like hypersurface Σ are themselves connected by an infinitesimal canonical transformation. Denoting the generator of this transformation by $s^+(\Sigma)$ and the corresponding variation in any Σ-associated field quantity f_Σ by $\delta^+ f_\Sigma$, we have, by definition,

$$\delta^+ f_\Sigma = -(f_\Sigma, s^+(\Sigma)), \tag{2.5}$$

the parentheses denoting the Poisson bracket. From the fundamental theorem, on the other hand, we have

$$\bar{\delta} S(\psi_\infty | \psi_\Sigma) = s^+(\infty) - s^+(\Sigma), \tag{2.6}$$

where the symbol $\bar{\delta}$ is used to denote the change in the functional form. Since the advanced boundary conditions assure us that $s^+(\infty)=0$, it therefore follows that

$$\delta^+ f_\Sigma = (f_\Sigma, \delta S(\psi_\infty | \psi_\Sigma)). \tag{2.7}$$

For retarded boundary conditions, described by variations δ^- and generators $s^-(\Sigma)$, with $s^-(-\infty)=0$, the corresponding equations are

$$\bar{\delta} S(\psi_\Sigma | \psi_{-\infty}) = s^-(\Sigma) - s^-(-\infty), \tag{2.8}$$

$$\delta^- f_\Sigma = -(f_\Sigma, s^-(\Sigma)) = -(f_\Sigma, \bar{\delta} S(\psi_\Sigma | \psi_{-\infty})). \tag{2.9}$$

Under the variations $\delta^\pm$ associated with advanced or retarded boundary conditions, the segments $S(\psi_\infty|\psi_\Sigma)$ and $S(\psi_\Sigma|\psi_{-\infty})$ of the over-all action suffer two independent changes: (1) a change δ in *value* due to the changing values of their arguments, and (2) the change $\bar{\delta}$ in functional form. The change δ is determined simply by inserting the new field $\psi^i + \delta^\pm \psi^i$ into the old action. The fact that the new field does not satisfy the field equations of the original system is unimportant. Because of the stationary action principle only the variations in the arguments at the endpoints contribute. The change $\bar{\delta}$, on the other hand, is determined by the replacement (2.4):

$$\bar{\delta} S = \epsilon B = \bar{\delta} S(\psi_\infty | \psi_\Sigma) + \bar{\delta} S(\psi_\Sigma | \psi_{-\infty}). \tag{2.10}$$

In evaluating this infinitesimal change it suffices to use the old field ψ^i in B.

Equations (2.5), (2.9), and (2.10) together yield

$$\delta^+ f_\Sigma - \delta^- f_\Sigma = \epsilon(f_\Sigma, B). \tag{2.11}$$

Since any dynamical variable A may be constructed out of hypersurface-associated variables f_Σ, Eq. (2.11) may immediately be generalized to

$$\delta^+ A - \delta^- A = \epsilon(A, B), \tag{2.12}$$

in which reference to spacelike hypersurfaces no longer appears. Equation (2.12) expresses essentially the content of Peierls' definition of the Poisson bracket.[2] Its utility rests on the fact that an independent evaluation of the variations $\delta^\pm A$ is possible, as will be seen in the next section. It is important, however, to make two remarks about this definition. Firstly, our derivation of Eq. (2.12) is not an explicit one as was Peierls' in which the propagation of small disturbances in a system having a constraintless Lagrangian was studied directly. It proceeds instead from the fundamental theorem of canonical transformation theory, and although the existence of canonically conjugated variables which can be used to give the ordinary definition of the Poisson bracket [e.g., when it first appears in Eq. (2.5)] is, of course, assumed, it never makes explicit use of them. In a theory possessing an infinite dimensional invariance group, in fact, proper canonical variables are usually extremely difficult to find. Nevertheless, in attempts to quantize such theories a great deal of effort has been devoted to the search for precisely these variables. On the other hand, these variables are rarely, if ever, of immediate physical interest, and therefore it is desirable to have a theory which is more accessible to direct physical intuition.

The second remark is this: By focusing our attention on variables of immediate physical interest we restrict ourselves, in the definition of Poisson brackets, to physically measurable quantities, which are of necessity group invariants. But this is quite satisfactory from the point of view of Eq. (2.12) since the variations $\delta^\pm A$ in any group invariant A will be well defined even though the variations $\delta^\pm \psi^i$ in the field variables themselves are not because of the possibility of performing infinitesimal group transformations. We see that with this restriction Eq. (2.12) in effect amounts to a generalization of Peierls' definition. By working directly with Green's functions, in fact, we shall demonstrate in the next section that the Poisson bracket (2.12) satisfies all the usual indentities and that it can therefore be disconnected completely from its canonical origins.

3. THE INVARIANCE GROUP AND ITS EFFECT ON THE DETAILS OF THE THEORY

The representation of the infinite-dimensional invariance group of the theory, which is provided by the field variables ψ^i, may be expressed in the infinitesimal form

$$\delta\psi^i = \int R^i{}_{L'} \delta\xi^{L'} d^4x', \tag{3.1}$$

where the functions $\delta\xi^L$ are infinitesimal group param-

eters. Here, capital Latin indices from the middle of the alphabet $(L,M,N,\cdots)$ will be associated with the symbol x, while those from the beginning of the alphabet $(A,B,C,\cdots)$ will be associated with the symbol z.

The representation (3.1) need not be linear in ψ^i but may be quite general. The only restriction on it is the identity

$$\int (R^i{}_{A,j'}R^{j'}{}_{B'}-R^i{}_{B',j'}R^{j'}{}_A)d^4x' = \int R^i{}_{L'}c^{L'}{}_{AB'}d^4x', \quad (3.2)$$

where the $c^{L'}{}_{AB'}$ are the structure constants of the group, which in turn satisfy the identity

$$\int (c^L{}_{AM'}c^{M'}{}_{B'C''}+c^L{}_{B'M'}c^{M'}{}_{C''A} +c^L{}_{C''M'}c^{M'}{}_{AB'})d^4x'=0. \quad (3.3)$$

Typically $R^i{}_{L'}$ will be a differential operator; that is, a linear combination of the delta function and its derivatives, with coefficients involving the field variables and their derivatives. It is characteristic of the field theories which are of interest in physics that a homogeneous quadratic function of this operator, of the form

$$F_{AB'}\equiv \int d^4x \int d^4x' g_{ij'}R^i{}_A R^{j'}{}_{B'}, \quad (3.4)$$

where $g_{ij'}$ is a symmetric regular continuous matrix having a unique inverse $g^{ij'}$, can always be found which is a wave operator possessing unique Green's functions $G^{\pm AB'}$. The choice of $g_{ij'}$ is not necessarily unique, but having made it we stick to it, and use it together with its inverse to raise and lower the field indices i, j', $\cdots$, etc.

It is evident from Eq. (3.1) that a group invariant A is characterized by the condition

$$\int A_{,i}R^i{}_A d^4x=0. \quad (3.5)$$

The action S, in particular, will satisfy this condition independently of the field equations. This means that the field equations themselves are not all independent of one another but are subject to constraints. It is assumed, of course, that the invariance group alone gives rise to the totality of conditions (3.5). That no further conditions can be obtained by taking variational derivatives is assured by the identity (3.2).

By taking the variational derivative of Eq. (3.5), with A replaced by S, it is easy to show that under the group transformation (3.1) the field equations (2.1) are replaced by linear combination of themselves. We have

$$\delta S_{,i}=\int S_{,ij'}\delta\psi^{j'}d^4x' = \int d^4x' \int d^4z S_{,ij'}R^{j'}{}_A\delta\xi^A = -\int d^4x' \int d^4z S_{,j'}R^{j'}{}_{A,i}\delta\xi^A. \quad (3.6)$$

In this way we see how the invariance of the action insures the invariance of the field equations. It will be noted that when the field equations are satisfied the second variational derivative of the action satisfies

$$\int S_{,ij'}R^{j'}{}_A d^4x'=0. \quad (3.7)$$

The continuous matrix $S_{,ij'}$ therefore possesses true null eigenvectors and can have no inverses. For the field theories which are of interest in physics, however, it is always possible to find a symmetric regular continuous matrix $g^{AB'}$ such that the combination

$$F_{ij'}\equiv S_{,ij'}+\int d^4z \int d^4z' g^{AB'}R_{iA}R_{j'B'} \quad (3.8)$$

is a proper wave operator with unique Green's functions $G^{\pm ij'}$. The matrix $g^{AB'}$ together with its inverse $g_{AB'}$ will be used to raise and lower the group indices A, B', $\cdots$, etc.

Let us now consider the change (2.4) in the action. Under this change the field variables will suffer a corresponding modification $\delta^{\pm}\psi^i$ which satisfies the equation

$$\delta^{\pm}S_{,i}\equiv \int S_{,ij'}\delta^{\pm}\psi^{j'}d^4x'=-\epsilon B_{,i}. \quad (3.9)$$

The $\delta^{\pm}\psi^i$ are, of course, not well defined, being determined only up to a transformation of the form (3.1). In virtue of Eq. (3.5), however, the corresponding change in any invariant A *is* well defined. It is evident that the general solution of Eq. (3.9) is obtained by adding (3.1) to an arbitrary linear combination of particular solutions determined by appropriate boundary and supplementary conditions. The boundary conditions to be adopted are already implied by the $\pm$ signs. As the supplementary condition it is convenient to choose

$$\int R_{iA}\delta^{\pm}\psi^i d^4x=0. \quad (3.10)$$

If this condition is not already satisfied then it is easy to see that it may nevertheless always be imposed by

carrying out an infinitesimal group transformation (3.1) with the group parameters choosen according to

$$\delta\xi^A=\int d^4x\int d^4z' G^{\pm AB'}R_{iB'}\delta^{\pm}\psi^i. \tag{3.11}$$

When the supplementary condition (3.10) is satisfied, it follows at once from (3.8) that

$$\int F_{ij'}\delta^{\pm}\psi^{j'}d^4x'=\int S_{,ij'}\delta^{\pm}\psi^{j'}d^4x'. \tag{3.12}$$

The solution of Eq. (3.9) may, therefore, be written immediately in the form

$$\delta^{\pm}\psi^i=\epsilon\int G^{\pm ij'}B_{,j'}d^4x'. \tag{3.13}$$

It is important to check, however, that this solution in fact satisfies the supplementary condition (3.10) which was used to get it in the first place. This can be done by first deriving an important relation between the Green's functions $G^{\pm AB'}$ and $G^{\pm ij'}$. We note, using Eqs. (3.4), (3.7), and (3.8), that

$$\int R^i{}_A F_{ij'}d^4x=\int F_{AB'}R_{j'}{}^{B'}d^4z'. \tag{3.14}$$

Therefore,

$$\int d^4x\int d^4z' F_{AB'}R_i{}^{B'}G^{\pm ij'}$$
$$=\int d^4x\int d^4x'' R^{k''}{}_A F_{k''i}G^{\pm ij'}=-R^{j'}{}_A. \tag{3.15}$$

But also,

$$\int d^4z'\int d^4z'' F_{AB'}G^{\pm B'C''}R^{j'}{}_{C''}=-R^{j'}{}_A. \tag{3.16}$$

Equations (3.15) and (3.16) are both wave equations in the operator $F_{AB'}$, having the same inhomogeneous term, $-R^{j'}{}_A$. The functions satisfying these equations have the same kinematical properties and must, therefore, be identical; that is,

$$\int R_i{}^A G^{\pm ij'}d^4x=\int G^{\pm AB'}R^{j'}{}_{B'}d^4z'. \tag{3.17}$$

On using this equation with the index A in the lower position, we then have, from (3.13),

$$\int R_{iA}\delta^{\pm}\psi^i d^4x=\epsilon\int d^4x'\int d^4z' G^{\pm}{}_A{}^{B'}R^{j'}{}_{B'}B_{,j'}, \tag{3.18}$$

which vanishes in virtue of Eq. (3.5) (with A replaced by B), thus showing the complete self-consistency of the condition (3.10).

The explicit form for the Poisson bracket (2.12) now follows from Eqs. (1.7) and (3.13). On writing

$$\delta^{\pm}A=\int A_{,i}\delta^{\pm}\psi^i d^4x, \tag{3.19}$$

we have

$$(A,B)=\int d^4x\int d^4x' A_{,i}G^{ij'}B_{,j'}, \tag{3.20}$$

a result which is immediately interpretable in terms of the mutual interference of measurements performed on A and B.[11] That the Poisson bracket (3.20) is unique and independent of possible freedom of choice of the regular matrices $g_{ij'}$ and $g^{AB'}$ follows from the uniqueness of this mutual interference.

The infinitesimal canonical transformation $\delta^+A-\delta^-A$ generated by the group invariant B has a special characteristic worth noting. Since $\delta^+S_{,i}-\delta^-S_{,i}=0$ it follows that the field $\psi^i+\delta^+\psi^i-\delta^-\psi^i$ satisfies the field equations if ψ^i does. Hence we see that group invariants transform solutions of the field equations into other solutions. The consequent role of group invariants as infinitesimal generators for the *group of mappings* of the set of all physically distinct solutions of the field equations into itself guarantees that the Poisson bracket (3.20) satisfies all of the identities usually associated with Poisson brackets; for the Poisson brackets may be mapped into the commutators of the Lie ring associated with the mapping group.

These identities may also be verified directly. The antisymmetry of the Poisson bracket follows from the reciprocity law (1.13). The identity

$$(A,BC)\equiv(A,B)C+B(A,C) \tag{3.21}$$

is obvious. The verification of the Poisson-Jacobi identity, on the other hand, requires a little computa-

---

[11] The mutual interference is more naturally expressed in terms of retarded Green's functions. Denoting by $\delta_B A$ the retarded change in A produced by the change (2.4) in the action, and by $\delta_A B$ the corresponding change in B with B replaced by A in (2.4), we have

$$\delta_B A=\epsilon\int d^4x\int d^4x' A_{,i}G^{-ij'}B_{,j'},$$
$$\delta_A B=\epsilon\int d^4x\int d^4x' B_{,i}G^{-ij'}A_{,j'}.$$

In virtue of the reciprocity law (1.14), therefore, Eq. (2.12) may be reexpressed in the form

$$\delta_A B-\delta_B A=\epsilon(A,B),$$

which is the original form of Peierls' definition. We note also that the retarded change in A produced by B is equal to the advanced change in B produced by A, and vice versa.

tion. By straightforward application of (3.20), we find

$$(A,(B,C))+(B,(C,A))+(C,(A,B))$$
$$=\int d^4x\int d^4x'\int d^4x''\int d^4z$$
$$\times[A_{,ia}B_{,j'}C_{,k''}(G^{k''a}G^{ij'}+G^{j'a}G^{k''i})$$
$$+A_{,i}B_{,j'a}C_{,k''}(G^{ia}G^{j'k''}+G^{k''a}G^{ij'})$$
$$+A_{,i}B_{,j'}C_{,k''a}(G^{j'a}G^{k''i}+G^{ia}G^{j'k''})$$
$$+A_{,i}B_{,j'}C_{,k''}(G^{ia}G^{j'k''}{}_{,a}+G^{j'a}G^{k''i}{}_{,a}$$
$$+G^{k''a}G^{ij'}{}_{,a})]. \quad (3.22)$$

The first three terms in the square brackets vanish because of the reciprocity law (1.13) and the interchangeability of the order of variational differentiation. In order to evaluate the fourth term it is necessary to investigate the variational derivative of the propagation function itself. By taking the variational derivative of Eq. (1.5), we find

$$\int F_{ik''}G^{\pm k''j'}{}_{,a}d^4x''=-\int F_{ik''}{}_{,a}G^{\pm k'j'}d^4x'', \quad (3.23)$$

of which the solution, on taking into account the kinematics of the Green's functions and using Eqs. (3.8) and (3.17), is

$$G^{\pm ij'}{}_{;a}=\int d^4z'\int d^4z''G^{\pm ib'}F_{b'c''}{}_{,a}G^{\pm c''j'}$$
$$=\int d^4z'\int d^4z''G^{\pm ib'}S_{,ab'c''}G^{\pm c''j'}+\int d^4z'$$
$$\times\int d^4z''\int d^4x''(G^{\pm ib'}R_{b'L''}{}_{,a}G^{\pm L''A''}R^{j'}{}_{A''}$$
$$+R^{iA''}G^{\pm}{}_{A''L''}R_{b'}{}^{L''}{}_{,a}G^{\pm b'j'}). \quad (3.24)$$

Upon insertion into (3.22) the last term of (3.24) gives no contribution, since the contraction of $R^{iA''}$ and $R^{j'}{}_{A''}$ with $A_{,i}$ and $B_{,j'}$, etc., vanishes on account of the invariance property of A, B, and C as expressed by Eq. (3.5). With use of Eqs. (1.7) and (1.20),[12] therefore, the expression (3.22) reduces to

$$\int d^4x\int d^4x'\int d^4x''\int d^4z\int d^4z'\int d^4z''A_{,i}B_{,j'}C_{,k''}S_{,ab'c''}$$
$$\times\{[\theta(x'',z'')-\theta(x',z')]G^{ia}G^{j'b'}G^{c''k''}$$
$$+[\theta(x,z'')-\theta(x'',z')]G^{j'a}G^{k''b'}G^{c''i}$$
$$+[\theta(x',z'')-\theta(x,z')]G^{k''a}G^{ib'}G^{c''j'}\}. \quad (3.25)$$

[12] The author is indebted to Dr. T. Imamura for pointing out the utility of the identity (1.20) in the present connection.

By permuting the indices a, b', c'' and, correspondingly, the points z, z', z'' (which is permitted because of the complete symmetry of $S_{,ab'c''}$ in its indices) it is easy to see that the various terms of this expression cancel one another, thus confirming the Poisson-Jacobi identity.

The foregoing theory of the Poisson bracket may be regarded (particularly in its measurement-theoretical interpretation) as a correspondence principle limit of the quantized theory. In a more accurate treatment, in which the Poisson bracket is viewed as an actual commutator—which is to say, in which one is interested in effects beyond the lowest order evaluation of the mutual interference of two measurements—the verification of the Poisson-Jacobi identity is not so simple, for the reason that the propagation function is itself a q-number in all except completely trivial linear theories, and may stand in different orders in pairs of terms which would otherwise cancel one another in expressions such as (3.25). The operator properties of the propagation function indicate that the commutator form of Eq. (3.20) should correctly be written

$$[A,B]=i\int d^4x\int d^4x'\frac{\delta A}{\delta\psi^i}\cdot G^{ij'}\cdot\frac{\delta B}{\delta\psi^{j'}}. \quad (3.26)$$

(We use units in which $\hbar=c=1$.) Here the dots signify that the propagation function is first to be inserted as a replacement for $\delta\psi^{j'}$ in all the places in which it occurs in the variation δB (assuming that A and B are expressible in terms of products of ψ's) and that the resulting "product" is then to be inserted as a replacement for $\delta\psi^i$ in the variation δA, or, alternatively, that the process of insertion is first performed in δA and then in δB.[13] That the two procedures are equivalent is evident from the familiar properties of commutator brackets when one takes note of the fact that in computing the commutator of two group invariants [of which the characterization (3.5) should now be replaced by the more rigorous form $\int(\delta A/\delta\psi^i)\cdot R^i{}_A d^4x=0$], one may work directly with the ψ's out of which they are built, proceeding as if the simple commutation rule

$$[\psi^i,\psi^{j'}]=iG^{ij'} \quad (3.27)$$

were valid.[14] These observations, of course, merely

[13] If Fermi statistics are required than the ψ's must come in pairs in any group invariant, and $\delta A/\delta\psi^i$ should be taken as a "right derivative" and $\delta B/\delta\psi^{j'}$ as a "left derivative" with respect to each pair. The process of insertion of the propagation function then involves the pairing of a ψ from A with a ψ from B in an antisymmetrized combination.

[14] The anticommutator should, of course, appear here when Fermi statistics are involved. It will be observed that this requires the propagation function to satisfy a symmetrical reciprocity law $G^{ij'}=G^{j'i}$ in place of Eq. (1.13). This does not lead to an inconsistency, however, because the wave operator $F_{ij'}$ turns out to be antisymmetric instead of symmetric in precisely those cases in which Fermi statistics are required. The alterations which this requires in the results of Sec. 1 are straightforward and will be left to the reader.

shift the question of the consistency of the rigorous quantum theory onto the propagation function itself and its correct definition as an operator. We shall not, however, pursue this problem further here.

A final remark should be made about the structure of the group invariants, A and B, appearing in a Poisson bracket. Heretofore, we have always had in mind, for each of these quantities, some explicit functional expression involving the ψ's. Actually these quantities are defined only *modulo* the field equations. It is straightforward to show, however, that this freedom leaves the value of the Poisson bracket unaffected. Let us, for example, replace B by

$$B'=B+\int f^i S_{,i} d^4x, \tag{3.28}$$

where the f^i are arbitrary coefficients. [The group invariance of the second term follows from Eq. (3.7) together with the field equations.] We have

$$(A,B')=(A,B)+\int d^4x \int d^4x' \times \int d^4x'' A_{,i} G^{ij'} S_{,j'k''} f^{k''}, \tag{3.29}$$

in which terms in $S_{,i}$ have been dropped *after* the variational differentiations have been performed. In virtue of Eqs. (1.12), (3.8), and (3.17), however, this becomes

$$\begin{aligned}(A,B') &= (A,B)-\int d^4x \int d^4x' \int d^4x'' \int d^4z A_{,i} \times G^{ij'} R_{j'A} R_{k''}{}^{A} f^{k''} \\ &= (A,B)-\int d^4x \int d^4x'' \int d^4z \int d^4z' A_{,i} \times R^{i}{}_{B'} G^{B'}{}_{A} R_{k''}{}^{A} f^{k''},\end{aligned} \tag{3.30}$$

which reduces simply to (A,B) in view of the invariance condition (3.5).

4. YANG-MILLS FIELD

The invariance group of the Yang-Mills theory[5] is the infinite-dimensional group which is obtained by taking the direct product of the rotation group in three dimensions (isospin group) with itself an infinite number of times, once for each point of space time. The fact that the starting group is chosen to be the rotation group is unimportant as far as the mathematical structure of the theory is concerned, and we shall in this section take it to be an arbitrary finite dimensional Lie group, with structure constants $c^L{}_{MN}$.

The homogeneous linear representations of the infinite dimensional group are restricted by differentiability requirements to be in one to one correspondence with the linear representations of the starting group. The infinite dimensional group has, however, inhomogeneous linear representations which have no counterparts in the theory of the starting group, and which are introduced through the notion of "invariant differentiation." One begins with a field, represented by a column vector ψ, which provides an arbitrary linear representation of the group through an infinitesimal transformation law of the form

$$\delta\psi=G_L\psi\delta\xi^L, \tag{4.1}$$

where the G_L are matrices (infinitesimal generators) satisfying the commutation law

$$[G_M,G_N]=G_L c^L{}_{MN}. \tag{4.2}$$

One then introduces another auxiliary field $A^L{}_\mu$ (the Yang-Mills field) in terms of which the "invariant derivative" of ψ is defined:

$$\psi_{,\mu}\equiv\psi_{,\mu}+G_L A^L{}_\mu\psi. \tag{4.3}$$

(Here the comma followed by a Greek index denotes ordinary differentiation with respect to a space-time coordinate.) The group transformation law for the field $A^L{}_\mu$ is chosen in such a way as to make $\psi_{,\mu}$ have the same transformation law as ψ:

$$\delta\psi_{,\mu}=G_L\psi_{,\mu}\delta\xi^L. \tag{4.4}$$

By making use of Eq. (4.2) together with the relation $\delta\psi_{,\mu}=G_L(\psi\delta\xi^L)_{,\mu}$, it is not hard to see that $A^L{}_\mu$ must suffer the infinitesimal transformation

$$\delta A^L{}_\mu=-\delta\xi^L{}_{,\mu}+c^L{}_{MN}A^N{}_\mu\delta\xi^M. \tag{4.5}$$

That Eq. (4.5) provides a new type of representation of the group (linear inhomogeneous) may be verified in a straightforward manner by computing the commutator (3.2) of two successive infinitesimal transformations. If it were not for the presence of the inhomogeneous term $-\delta\xi^L{}_{,\mu}$ the transformation (4.5) would simply be that of the so-called "adjoint representation," the existence of which depends on the fact that the structure constants, when regarded as matrices in their first and last indices, themselves satisfy the commutation law (4.2) for infinitesimal generators [cf. Eq. (3.3)]. In the case of Abelian groups, for which the structure constants vanish, however, it is precisely the inhomogeneous term which renders the representation nontrivial. It is to be noted that in all questions of group representations we are here concerned only with the local group in the neighborhood of the unit element. If the starting group is compact the representation (4.5) is actually multivalued. For, owing to the finiteness of the group volume, any single valued function on the group must be periodic in the group parameters, whereas the inhomogeneous term of (4.5), even when integrated to give the finite transformation law, lacks this periodic property.

Indices induced by repeated invariant differentiation do not commute. We have, in fact,

$$\psi_{,\mu\nu}-\psi_{,\nu\mu}=-G_L F^L{}_{\mu\nu}\psi, \tag{4.6}$$

$$F^L{}_{\mu\nu}\equiv A^L{}_{\nu,\mu}-A^L{}_{\mu,\nu}+c^L{}_{MN}A^M{}_\mu A^N{}_\nu. \tag{4.7}$$

It is easily verified that the field $F^L{}_{\mu\nu}$ transforms according to the homogeneous adjoint representation. It therefore possesses an invariant derivative defined according to the law (4.3), which here takes the form

$$F^L{}_{\mu\nu,\sigma}\equiv F^L{}_{\mu\nu,\sigma}+c^L{}_{MN}A^M{}_\sigma F^N{}_{\mu\nu}. \tag{4.8}$$

Straightforward computation shows that this invariant derivative satisfies the identity

$$F^L{}_{\mu\nu,\sigma}+F^L{}_{\nu\sigma,\mu}+F^L{}_{\sigma\mu,\nu}\equiv 0. \tag{4.9}$$

Corresponding to every linear representation (4.1) there exists another given by the law

$$\delta\phi=-G_L{}^{\sim}\phi\delta\xi^L, \tag{4.10}$$

the tilde denoting the transpose. The field ϕ is said to transform "contragradiently" to the field ψ. For a compact starting group the two representations are always equivalent if they are real. This follows from the fact that a real matrix representation of a compact group is always equivalent to an orthogonal one (the proof of which involves the classic procedure of integrating over the whole group) and the fact that for an orthogonal representation the generators G_L are antisymmetric. It suffices to restrict our attention here to real representations, since a complex representation can always be regarded as a real representation of higher dimensionality.

It often happens that the two representations are equivalent even when the starting group is not compact. We shall now assume this to be the case, regardless of the compactness or noncompactness of the starting group. We shall also assume the equivalence to hold for the adjoint representation. We may then write

$$-G_L{}^{\sim}=\gamma G_L\gamma^{-1}, \tag{4.11}$$

$$-c^L{}_{MN}=g_{NJ}c^J{}_{MK}g^{KL},\quad g_{LK}g^{KM}=\delta_L{}^M, \tag{4.12}$$

where γ and (g_{MN}) are certain nonsingular real matrices.[15] The matrix g_{MN} and its inverse g^{MN} will be used to raise and lower group indices; Eq. (4.12) itself insures that this is an invariant process. The matrix γ must be either symmetric or antisymmetric whenever the matrices G_L are irreducible.[16] This follows from the fact that Eq. (4.11) implies

$$[\gamma^{-1}\gamma^{\sim},G_L]=0, \tag{4.13}$$

which, combined with the fact that $\det(\gamma^{-1}\gamma^{\sim})=1$, requires $\gamma^{-1}\gamma^{\sim}=\pm 1$. We may normalize γ so that $\det\gamma=\pm 1$.

If two fields ψ and ϕ transform contragradiently it is easy to see that the "scalar product" $\phi^{\sim}\psi$ is a group invariant. When Eq. (4.11) holds it is possible to construct quadratic forms which are invariant, for example $\bar{\psi}\psi$, where

$$\bar{\psi}\equiv\psi^{\sim}\gamma. \tag{4.14}$$

It will be noted, however, that if γ is antisymmetric, the quadratic form $\bar{\psi}\psi$ will vanish identically unless the field ψ satisfies Fermi statistics.

For the remainder of this section we shall consider the case in which γ and g_{MN} are symmetric. We shall take ψ to be a simple scalar field of mass m in a Lorentz invariant theory. The Lagrangian function of the Yang-Mills theory[5] then takes the form

$$L=-\tfrac{1}{4}F_{L\mu\nu}F^L{}_{\mu\nu}-\tfrac{1}{2}(\bar{\psi}_{,\mu}\psi_{,\mu}+m^2\bar{\psi}\psi), \tag{4.15}$$

for the combined fields ψ and $A^L{}_\mu$. For simplicity we here use an imaginary fourth coordinate in a Minkowski system and write all coordinate indices in the lower position. The stationary action principle based on the Lagrangian (4.15) leads to the field equations

$$0=\delta S/\delta A^L{}_\mu\equiv -F_{L\mu\nu,\nu}+\bar{\psi}G_L\psi_{,\mu}, \tag{4.16}$$

$$0=\delta S/\delta\bar{\psi}\equiv\psi_{,\mu\mu}-m^2\psi. \tag{4.17}$$

By means of these equations the field $A^L{}_\mu$ is dynamically coupled to the field ψ. When the invariance group is non-Abelian the field $A^L{}_\mu$ is also coupled to itself. A remark is in order concerning the strength of this coupling, which at first sight would seem to be fixed since no explicit coupling constants appear in the Lagrangian (4.15). The strength of the coupling is actually determined by the scale chosen for the "coordinate" or parameter mesh in the neighborhood of the unit element in the abstract group space, and is therefore completely flexible. A transformation from one parameter system to another would cause a change in the values of the structure constants and a rescaling of the field $A^L{}_\mu$, and "coupling constants" would then make an explicit appearance.

Under the change (2.4) in the action the fields $A^L{}_\mu$ and ψ suffer variations $\delta^\pm A^L{}_\mu$ and $\delta^\pm\psi$ satisfying

$$\begin{aligned}-\epsilon\delta B/\delta A^L{}_\mu&=\delta^\pm(\delta S/\delta A^L{}_\mu)\\&=-\delta^\pm A_{L\nu,\mu\nu}+\delta^\pm A_{L\mu,\nu\nu}-c_{LMN}F^N{}_{\mu\nu}\delta^\pm A^M{}_\nu\\&\quad+\bar{\psi}G_L G_M\psi\delta^\pm A^M{}_\mu+\bar{\psi}G_L\delta^\pm\psi_{,\mu}\\&\quad-\bar{\psi}_{,\mu}G_L\delta^\pm\psi,\end{aligned} \tag{4.18}$$

$$\begin{aligned}-\epsilon\delta B/\delta\bar{\psi}&=\delta^\pm(\delta S/\delta\bar{\psi})\\&=2G_L\psi_{,\mu}\delta^\pm A^L{}_\mu+G_L\psi\delta^\pm A^L{}_{\mu,\mu}\\&\quad+\delta^\pm\psi_{,\mu\mu}-m^2\delta^\pm\psi.\end{aligned} \tag{4.19}$$

As supplementary condition on these variations, it is convenient to choose

$$\delta^\pm A^L{}_{\mu,\mu}=0. \tag{4.20}$$

[15] With the aid of the identity (3.3) satisfied by the structure constants it is easy to show that one may choose $g_{MN}=-c^K{}_{ML}c^L{}_{NK}$ whenever the eigenvalues of the latter matrix are all different from zero (e.g., in the case of compact semisimple groups).

[16] We mean irreducible by means of real matrices. They may still be reducible to a complex form.

If this condition is not already satisfied, it may be imposed by performing a group transformation for which the parameters $\delta\xi^L$ satisfy[17]

$$\delta\xi^L{}_{,\mu\mu}=\delta^{\pm}A^L{}_{\mu,\mu}. \tag{4.21}$$

This equation is solved by

$$\delta\xi^L=-\int G^{\pm L}{}_{M'}\delta^{\pm}A^{M'}{}_{\mu',\mu'}d^4x', \tag{4.22}$$

where the Green's functions $G^{\pm L}{}_{M'}$ satisfy

$$G^{\pm L}{}_{M',\mu\mu}=-\delta^L{}_{M'}. \tag{4.23}$$

When condition (4.20) is satisfied, Eqs. (4.18) and (4.19) reduce to

$$\delta^{\pm}A_{L\mu,\nu\nu}-2c_{LMN}F^N{}_{\mu\nu}\delta^{\pm}A^M{}_{\nu}+\bar{\psi}G_LG_M\delta^{\pm}A^M{}_{\mu}+\bar{\psi}G_L\delta^{\pm}\psi_{,\mu}-\bar{\psi}_{,\mu}G_L\delta^{\pm}\psi=-\epsilon\delta B/\delta A^L{}_{\mu}, \tag{4.24}$$

$$2G_L\psi_{,\mu}\delta^{\pm}A^L{}_{\mu}+\delta^{\pm}\psi_{,\mu\mu}-m^2\delta^{\pm}\psi=-\epsilon\delta B/\delta\bar{\psi}. \tag{4.25}$$

These equations are solved with the aid of a set of Green's functions $G^{\pm}{}_{L\mu M'\nu'}$, $G^{\pm}{}_{L\mu\psi'}$, $G^{\pm}{}_{\psi M'\nu'}$, $G^{\pm}{}_{\psi\psi'}$ which satisfy the simultaneous equations

$$G^{\pm L}{}_{\mu M'\nu',\sigma\sigma}-2c^L{}_{KN}F^N{}_{\mu\sigma}G^{\pm K}{}_{\sigma M'\nu'}+\bar{\psi}G^LG_K\psi G^{\pm K}{}_{\mu M'\nu'}+\bar{\psi}G^LG^{\pm}{}_{\psi M'\nu',\mu}-\bar{\psi}_{,\mu}G^LG^{\pm}{}_{\psi M'\nu'}=-\delta_{\mu\nu}\delta^L{}_{M'}, \tag{4.26}$$

$$G^{\pm L}{}_{\mu\psi',\sigma\sigma}-2c^L{}_{KN}F^N{}_{\mu\sigma}G^{\pm K}{}_{\sigma\psi'}+\bar{\psi}G^LG_K\psi G^{\pm K}{}_{\mu\psi'}+\bar{\psi}G^LG^{\pm}{}_{\psi\psi',\mu}-\bar{\psi}_{,\mu}G^LG^{\pm}{}_{\psi\psi'}=0, \tag{4.27}$$

$$2G_L\psi_{,\mu}G^{\pm L}{}_{\mu M'\nu'}+G^{\pm}{}_{\psi M'\nu',\mu\mu}-m^2G^{\pm}{}_{\psi M'\nu'}=0, \tag{4.28}$$

$$2G_L\psi_{,\mu}G^{\pm L}{}_{\mu\psi'}+G^{\pm}{}_{\psi\psi',\mu\mu}-m^2G^{\pm}{}_{\psi\psi'}=-\delta(x-x'). \tag{4.29}$$

These Green's functions, in the symmetric combination (1.16) from which the retarded and advanced parts are easily re-extracted, can be shown[18] to have the following explicit structure:

$$\bar{G}_{LM'}=(4\pi)^{-1}[g_{LM'}\delta((x-x')^2)-v_{LM'}\theta(-(x-x')^2)], \tag{4.30}$$

$$\bar{G}_{L\mu M'\nu}=(4\pi)^{-1}[g_{LM'}\delta_{\mu\nu}\delta((x-x')^2)-v_{L\mu M'\nu'}\theta(-(x-x')^2)], \tag{4.31}$$

$$\bar{G}_{L\mu\psi'}=-(4\pi)^{-1}v_{L\mu\psi'}\theta(-(x-x')^2), \tag{4.32}$$

$$\bar{G}_{\psi M'\nu'}=-(4\pi)^{-1}v_{\psi M'\nu'}\theta(-(x-x')^2), \tag{4.33}$$

$$\bar{G}_{\psi\psi'}=(4\pi)^{-1}[\delta_{\psi\psi'}\delta((x-x')^2)-v_{\psi\psi'}\theta(-(x-x')^2)]. \tag{4.34}$$

Here, $\theta(-(x-x')^2)$ is the step function which vanishes outside the light cone and equals unity inside. The functions $g_{LM'}$ and $\delta_{\psi\psi'}$ are defined by the equations

$$g_{LM',\mu}(x_\mu-x'_\mu)\equiv(g_{LM',\mu}+c_{LK}{}^NA^K{}_{\mu}g_{NM'})(x_\mu-x'_\mu)=0, \tag{4.35}$$

$$\delta_{\psi\psi',\mu}(x_\mu-x'_\mu)\equiv(\delta_{\psi\psi',\mu}+G_LA^L{}_{\mu}\delta_{\psi\psi'})(x_\mu-x'_\mu)=0, \tag{4.36}$$

together with the boundary conditions

$$\lim_{x'\to x}g_{LM'}=g_{LM}, \tag{4.37}$$

$$\lim_{x'\to x}\delta_{\psi\psi'}=\text{unit matrix}. \tag{4.38}$$

The functions $v_{LM'}$, $v_{L\mu M'\nu'}$, $v_{L\mu\psi'}$, $v_{\psi M'\nu'}$, $v_{\psi\psi'}$ are given by infinite series of the general form

$$v=\sum_{n=0}^{\infty}v_n(x-x')^{2n}, \tag{4.39}$$

the coefficients of which satisfy the recurrence formulas

$$v_{0\ LM'}+v_{0\ LM',\mu}(x_\mu-x'_\mu)=-\tfrac{1}{4}g_{LM',\mu\mu}, \tag{4.40}$$

$$v_{n\ LM'}+(n+1)^{-1}v_{n\ LM',\mu}(x_\mu-x'_\mu)=-\tfrac{1}{4}n^{-1}(n+1)^{-1}v_{n-1\ LM',\mu\mu}, \tag{4.41}$$

$$v_{0\ L\mu M'\nu'}+v_{0\ L\mu M'\nu',\sigma}(x_\sigma-x'_\sigma)=-\tfrac{1}{4}(\delta_{\mu\nu}g_{LM',\sigma\sigma}-2c_{LKN}F^N{}_{\mu\nu}g^K{}_{M'}+\delta_{\mu\nu}\bar{\psi}G_LG_K\psi g^K{}_{M'}), \tag{4.42}$$

$$v_{n\ L\mu M'\nu'}+(n+1)^{-1}v_{n\ L\mu M'\nu',\sigma}(x_\sigma-x'_\sigma)=-\tfrac{1}{4}n^{-1}(n+1)^{-1}(v_{n-1\ L\mu M'\nu',\sigma\sigma}-2c_{LKN}F^N{}_{\mu\sigma}v_{n-1}{}^K{}_{\sigma M'\nu'}+\bar{\psi}G_LG_K\psi v_{n-1}{}^K{}_{\mu M'\nu'}+\bar{\psi}G_Lv_{n-1\ \psi M'\nu',\mu}-\bar{\psi}_{,\mu}G_Lv_{n-1\ \psi M'\nu'}), \tag{4.43}$$

$$v_{0\ L\mu\psi'}=0, \tag{4.44}$$

$$v_{n\ L\mu\psi'}+(n+1)^{-1}v_{n\ L\mu\psi',\sigma}(x_\sigma-x'_\sigma)=-\tfrac{1}{4}n^{-1}(n+1)^{-1}(v_{n-1\ L\mu\psi',\sigma\sigma}-2c_{LKN}F^N{}_{\mu\sigma}v_{n-1}{}^K{}_{\sigma\psi'}+\bar{\psi}G_LG_K\psi v_{n-1}{}^K{}_{\mu\psi'}+\bar{\psi}G_Lv_{n-1\ \psi\psi',\mu}-\bar{\psi}_{,\mu}G_Lv_{n-1\ \psi\psi'}), \tag{4.45}$$

$$v_{0\ \psi M'\nu'}=0, \tag{4.46}$$

$$v_{n\ \psi M'\nu'}+(n+1)^{-1}v_{n\ \psi M'\nu',\sigma}(x_\sigma-x'_\sigma)=-\tfrac{1}{4}n^{-1}(n+1)^{-1}(v_{n-1\ \psi M'\nu',\mu\mu}-m^2v_{n-1\ \psi M'\nu'}+2G_L\psi_{,\mu}v_{n-1}{}^L{}_{\mu M'\nu'}), \tag{4.47}$$

$$v_{0\ \psi\psi'}+v_{0\ \psi\psi',\mu}(x_\mu-x'_\mu)=-\tfrac{1}{4}(\delta_{\psi\psi',\mu\mu}-m^2\delta_{\psi\psi'}), \tag{4.48}$$

$$v_{n\ \psi\psi'}+(n+1)^{-1}v_{n\ \psi\psi',\mu}(x_\mu-x'_\mu)=-\tfrac{1}{4}n^{-1}(n+1)^{-1}(v_{n-1\ \psi\psi',\mu\mu}-m^2v_{n-1\ \psi\psi'}+2G_L\psi_{,\mu}v_{n-1}{}^L{}_{\mu\psi'}), \tag{4.49}$$

for $n=1, 2, 3\cdots$. Each of these equations may be integrated along each straight line emanating from the point x', and all the v's are thereby uniquely determined. The series (4.39) is everywhere convergent provided $A^L{}_{\mu}$ and ψ are bounded functions.

[17] The invariant differentiation law for the infinitesimals $\delta\xi^L$ is determined by the fact that they may be regarded as transforming according to the adjoint representation. This permits Eq. (4.5) to be written in the form $\delta A^L{}_{\mu}=-\delta\xi^L{}_{,\mu}$, which illustrates a special case of the rule that any variation in $A^L{}_{\mu}$, unlike $A^L{}_{\mu}$ itself, transforms according to the homogeneous adjoint representation.

[18] B. S. DeWitt and R. W. Brehme, Ann. Phys. **9**, 220 (1960).

The solutions of Eqs. (4.24) and (4.25) may be expressed in the form

$$\delta^{\pm}A^{L}{}_{\mu}=\epsilon\int[G^{\pm L}{}_{\mu M'\nu'}(\delta B/\delta A^{M'}{}_{\nu'})+G^{\pm L}{}_{\mu\psi'}(\delta B/\delta\bar{\psi}')]d^4x', \quad (4.50)$$

$$\delta^{\pm}\psi=\epsilon\int[G^{\pm}{}_{\psi M'\nu'}(\delta B/\delta A^{M'}{}_{\nu'})+G^{\pm}{}_{\psi\psi'}(\delta B/\delta\bar{\psi}')]d^4x', \quad (4.51)$$

from which the Poisson bracket immediately follows:

$$(A,B)=\int d^4x\int d^4x'\left(\frac{\delta A}{\delta A^{L}{}_{\mu}}G_{L\mu M'\nu'}\frac{\delta B}{\delta A^{M'}{}_{\nu'}}+\frac{\delta A}{\delta A^{L}{}_{\mu}}G_{L\mu\psi'}\frac{\delta B}{\delta\bar{\psi}'}+\frac{\delta A}{\delta\psi}G_{\psi M'\nu'}\frac{\delta B}{\delta A^{M'}{}_{\nu'}}+\frac{\delta A}{\delta\psi}G_{\psi\psi'}\frac{\delta B}{\delta\bar{\psi}'}\right). \quad (4.52)$$

In order to verify that the solution (4.50) satisfies the supplementary condition (4.20), we must establish some identities, analogous to Eq. (3.17), involving the Green's functions $G^{\pm L}{}_{\mu M'\nu'}$ and $G^{\pm L}{}_{\mu\psi'}$. By taking the invariant divergence of Eqs. (4.26) and (4.27), permuting the order of invariant differentiations, making use of Eqs. (4.2), (4.16), (4.17), (4.28), and (4.29), and taking note of the fact that c_{LMN} is completely antisymmetric in its indices [which follows from Eq. (4.12)], it is not hard to establish the following relations:

$$G^{\pm L}{}_{\mu M'\nu'.\mu\sigma}+\bar{\psi}G^{L}G_{K}\psi G^{\pm K}{}_{\mu M'\nu'}=\delta^{L}{}_{M'\nu'}, \quad (4.53)$$

$$G^{\pm L}{}_{\mu\psi'.\mu\sigma}+\bar{\psi}G^{L}G_{K}\psi G^{\pm K}{}_{\mu\psi'}=\bar{\psi}G^{L}\delta(x-x'). \quad (4.54)$$

From these relations it may be inferred that

$$G^{\pm L}{}_{\mu M'\nu'.\mu}=-\mathbf{G}^{\pm L}{}_{M'.\nu'}, \quad (4.55)$$

$$G^{\pm L}{}_{\mu\psi'.\mu}=-\mathbf{G}^{\pm L}{}_{M'}\bar{\psi}'G^{M'}, \quad (4.56)$$

where the $\mathbf{G}^{\pm L}{}_{M'}$ are Green's functions similar to the $G^{\pm L}{}_{M'}$ of Eq. (4.23) but satisfying the slightly more complicated equation[19]

$$\mathbf{G}^{\pm L}{}_{M'.\mu\mu}+\bar{\psi}G^{L}G_{K}\psi\mathbf{G}^{\pm K}{}_{M'}=-\delta^{L}{}_{M'}. \quad (4.57)$$

On taking the invariant divergence of Eq. (4.50) we, therefore, have

$$\delta^{\pm}A^{L}{}_{\mu.\mu}=-\epsilon\int[\mathbf{G}^{\pm L}{}_{M'.\nu'}(\delta B/\delta A^{M'}{}_{\nu'})+\mathbf{G}^{\pm L}{}_{M'}\bar{\psi}'G^{M'}(\delta B/\delta\bar{\psi}')]d^4x', \quad (4.58)$$

which, after an integration by parts, vanishes in virtue of the identity

$$(\delta B/\delta A^{L}{}_{\mu})_{.\mu}-\bar{\psi}G^{L}(\delta B/\delta\bar{\psi})=0 \quad (4.59)$$

which is necessarily satisfied by any group invariant, as may be readily inferred from the transformation laws (4.1) and (4.5). Consistency is, therefore, established.

---

[19] The reason for the difference between the two Green's functions here stems from the fact that the supplementary condition (4.20) actually renders the propagation equations (4.24) and (4.25) slightly nonself-adjoint. A less convenient choice of supplementary condition restores the self-adjoint property. The whole of the theory of Secs. 1 and 2 can be worked out for nonself-adjoint wave operators, but for the sake of clarity the author has refrained from doing this. Naturally any elements of nonself-adjointness which are arbitrarily inserted into the theory can have no effect on the identities satisfied by the Poisson brackets taken between group invariants.

Generalized Retarded Functions and Analytic Function in Momentum Space in Quantum Field Theory*

HUZIHIRO ARAKI†
Physics Department, University of Illinois, Urbana, Illinois
(Received May 2, 1960)

The analytic n-point function in momentum space in quantum field theory is studied. Its different boundary values for real value of the argument are determined, and a necessary and sufficient condition for them to be obtainable from the Wightman functions is given. The conditions are relativistic covariance, support properties in coordinate space (retardedness), two-term identities for momentum below threshold (corresponding to spectrum conditions) and four-term identities (Steinmann relations). The first three conditions are translatable into a statement about the domain of analyticity of the n-point function: it is analytic in a union of various extended tubes plus the points of contact of two neighboring tubes for real part of one momentum below threshold.

1. INTRODUCTION

THE retarded functions (the vacuum expectation values of retarded products of field operators) in quantum field theory are, as is well known, boundary values of an analytic function in momentum space. In this paper, we will attempt a systematic investigation of this analytic function and its boundary values. Such an investigation has also been made independently by Ruelle,[1] Steinmann,[2] and Burgoyne.[3] The present work puts emphasis on the geometrical nature of the problem in contrast with the algebraic method of Steinmann and Burgoyne. The method of Ruelle has some common features with the present work but we believe that ours is more explicit and detailed.

First we consider the analytic function in the energy component only, and we easily obtain all its boundary values which include all the conventional retarded and advanced functions. These boundary values will be called generalized retarded functions (r function). Their number is 6, 32, 370, and 10 932 for three-, four-, five-, and sixfold in contrast with 6, 24, 120, 720, for the Wightman functions.

By using a generalization of the θ function, we can express generalized retarded functions in terms of Wightman functions and the latter in terms of the former in a compact manner. Furthermore, we obtain necessary and sufficient conditions for generalized retarded functions to be obtainable from Wightman functions satisfying the usually considered conditions, namely, (W1) relativistic covariance, (W2) local commutativity or anticommutativity, and (W3) certain mass spectrum conditions. The resulting conditions on the r function are (R1) relativistic covariance, (R2) support properties in x space (retardedness or advancedness), (R3) two-term identities in momentum space for momentum below threshold, (R4) four-term identities. The four-term identities have first been found by Steinmann[4] for the four-point function.

The aforementioned analytic function can be extended to a covariant analytic function in all energy momentum components. The properties (R1)–(R3) are translatable into a statement about the domain of analyticity of this analytic function. Namely, it is analytic in the union of various extended tubes plus points of contact of two neighboring tubes for real parts of one momentum below threshold. We have not succeeded in translating (R4) into a statement about the domain of analyticity.

The time-ordered function can also be expressed as a boundary value of the same analytic function. The boundary values must then be approached from a direction which depends on the value of the real part of the momenta.

All the results are valid for arbitrary types of fields, bosons, and fermions.

In Sec. 2 we collect our main results (theorems 1–3), together with definitions of notations necessary for the statement of our results. In Sec. 3, the properties of generalized θ functions are studied and they are applied in Secs. 4–6 for the proof of our main results.

In Sec. 7 we make a few remarks about the class of functions for which our results hold. If the behavior of Wightman functions for large energy momentum is not sufficiently good, we have been unable to obtain our full results. As for the behavior at large coordinate separation, the truncated Wightman functions are expected to tend to zero in contrast to the Wightman functions themselves. Hence the truncated functions are used extensively in this work and their properties are studied in Appendix B.

The spectrum condition assumed in the main text is the existence of a single lowest positive mass. The case of more general mass spectrum conditions is treated in Appendix A. We obtain two-term identities for momentum below threshold and the corresponding analyticity. However, the sufficiency of this condition has

* Supported in part by the U. S. Office of Naval Research.
† Present address: Department of Nuclear Engineering, Kyoto University, Kyoto, Japan.
[1] D. Ruelle, thesis, Brussels, 1959.
[2] O. Steinmann, Helv. Phys. Acta 33, 347 (1960).
[3] N. Burgoyne (private communication); also see H. Araki and N. Burgoyne, Nuovo cimento 8, 342 (1960).
[4] O. Steinmann, Helv. Phys. Acta 33, 257 (1960).

not been fully established for a general mass spectrum condition.

In Appendix C, we collect definitions and known results concerning convex polyhedral cones which are extensively used in the main text.

2. NOTATIONS AND MAIN RESULTS

In this paper, we consider the quantum theory of several covariant fields $A_i(x)$ satisfying (1) the invariance under the inhomogeneous Lorentz group, (2) the local commutativity or anticommutativity, and (3) spectrum conditions. As spectrum conditions, we assume (3a) the existence of the vacuum (the nondegenerate invariant state), (3b) the positiveness of energy, and (3c) the existence of a lowest positive mass m. In Appendix A, we treat the case where (3c) is replaced by more complicated mass spectrum conditions.

These conditions can be used in a most compact way[5] for the truncated vacuum expectation values, as we shall see in the following. The Wightman functions are denoted by

$$w_P(x)=\sigma_P(\Psi_0,A_{P(1)}(x_{P(1)})\cdots A_{P(n+1)}(x_{P(n+1)})\Psi_0), \quad (2.1)$$

where P denotes the permutation of $1\cdots(n+1)$, σ_P is the signature of the permutation of anticommuting fields,[6] and

$$x=(x_1,\cdots,x_{n+1}). \quad (2.2)$$

Throughout this paper we shall take x_i as the argument of the field A_i. The truncated Wightman functions are defined recursively by[7]

$$\begin{aligned}(\Psi_0,A_{i1}(x_{i1})\cdots A_{im}(x_{im})\Psi_0)&=[A_{i1}(x_{i1})\cdots A_{im}(x_{im})]_T\\&\quad+\Sigma\sigma[A_{i1}(x_{i1})\cdots]_T[A_{ik}(x_{ik})\cdots]_T\cdots, \quad (2.3)\end{aligned}$$

$$w_P{}^T(x)=\sigma_P[A_{P(1)}(x_{P(1)})\cdots A_{P(n+1)}(x_{P(n+1)})]_T, \quad (2.4)$$

where the summation extends over all grouping of points $x_1\cdots x_m$, the A's in each $[\ ]_T$ of Eq. (2.3) are in the same order as on the left-hand side, σ is the signature of the permuation of anticommuting fields which brings $A_{i1}\cdots A_{im}$ to the order of the A in that term, and σ_P is as in Eq. (2.1). The purpose of this definition is to subtract from the Wightman functions in a symmetric manner the contributions from the vacuum intermediate states.

Because of the translational invariance of the theory, x can be taken modulo $(1,\cdots,1)$. The $4n$-dimensional vector space formed by x modulo $(1,\cdots,1)$ is denoted by X.

The Fourier transform of a Wightman function is denoted by

$$(2\pi)^4\delta(\sum_{i=1}^{n+1}q_i)\tilde{w}_P(q)=\int e^{i(q,x)}w_P(x)dx_1\cdots dx_{n+1}, \quad (2.5)$$

where

$$(q,x)=\sum_{i=1}^{n+1}(q_i,x_i), \quad (2.6)$$

and (q_i,x_i) is the conventional inner product in Minkowski space.[8] The $4n$-dimensional vector space formed by

$$q=(q_1,\cdots,q_{n+1}), \quad (2.7)$$

such that $\Sigma q_i=0$ is denoted by Q. The $\tilde{w}_P(q)$ are functions of q in Q.

The $\tilde{w}_P{}^T(q)$ are defined in a similar manner, namely,

$$\tilde{w}_P{}^T(q)=\int e^{i(q,x)}w_P{}^T(x)dx, \quad q\in Q \quad (2.8)$$

$$w_P{}^T(x)=(2\pi)^{-4n}\int e^{-i(q,x)}\tilde{w}_P{}^T(q)dq, \quad x\in X, \quad (2.9)$$

where dx and dq are the volume elements of X and Q,

$$dx=dx_1\cdots dx_n, \quad dq=\delta(\Sigma q_i)dq_1\cdots dq_{n+1}. \quad (2.10)$$

In order to control the combinatorical difficulties for large n, it is essential to introduce a compact, though somewhat involved notation. A set of integers is generally denoted by I, in particular the set $\{1,\cdots,n+1\}$ by $I(n+1)$ and

$$\{P(1),\cdots P(k)\}=I(P,k). \quad (2.11)$$

The set (of sets) $\{I(P,k);k=1,\cdots n\}$ will be called $\mathcal{I}_P$. We define

$$q(I)=\sum_{\nu\in I}q_\nu. \quad (2.12)$$

Note that

$$q[I(n+1)]=0, \quad q[I(n+1)-I]=-q(I). \quad (2.13)$$

The energy momentum vectors of intermediate states in the Wightman function w_P are $q(I)$, with $I\in\mathcal{I}_P$.

The properties of $w_P{}^T$ which follow from the assumptions (1)–(3) on the theory are (see Appendix B):

(W1) The $w_P{}^T(x)$ are covariant functions of $x\in X$.

(W2) If P' results from P by the interchange of the indices $P(k)$ and $P(k+1)$, and if $x_{P(k)}-x_{P(k+1)}$ is spacelike, then $w_P{}^T(x)=w_{P'}{}^T(x)$.

(W3) $\tilde{w}_P{}^T(q)=0$ unless $q(I)^2\geqq m$, $q^0(I)>0$ for all $I\in\mathcal{I}_P$.

[5] The mass spectrum condition for Wightman function is stated in (W2″) of Appendix B. It is more complicated because of the presence of the vacuum intermediate state. Also see discussion of Sec. 7.

[6] It is meant that σ_P is the sign change which one obtains if one changes the order of the fields from the natural order $1, 2, \cdots n+1$ to (the order) $P(1), P(2), \cdots P(n+1)$ for totally spacelike configuration of x_i. See Appendix B.

[7] R. Haag, Phys. Rev. **112**, 669 (1958). See also Appendix B for more detail. Equation (2.3) corresponds to Ursell's expansion in statistical mechanics. H. P. Ursell, Proc. Cambridge Phil. Soc. **23**, 685 (1927).

[8] The signature of the metric is $(1,-1,-1,-1)$.

We now turn to the main subject of the paper, the analytic function in momentum space. This function will be defined by Eq. (2.27) or in explicitly covariant form by Eq. (2.39). To show the equivalence of this definition with conventional usuage, let us start from the customary definition of a retarded function for Bose fields:

$$\begin{aligned} &r(x_1; x_2\cdots x_{n+1}) \\ &\quad=(-i)^n\Sigma\theta(x_1{}^0-x_{P(2)}{}^0)\cdots\theta(x_{P(n)}{}^0-x_{P(n+1)}{}^0) \\ &\qquad\times(\Psi_0,[\cdots[[A_1(x_1),A_{P(2)}(x_{P(2)})],A_{P(3)}(x_{P(3)})]\cdots \\ &\qquad\qquad\times A_{P(n+1)}(x_{P(n+1)})]\Psi_0), \end{aligned} \tag{2.14}$$

where the summation is over all permutations P of $2, \cdots, n+1$. On expanding the multiple commutators, this can be written as[9]

$$\begin{aligned} &r(x_1; x_2\cdots x_{n+1}) \\ &=\sum_j(-1)^{j-1}(-i)^n\sum_{P'(j)=1}\prod_{\nu=1}^{j-1}\theta(x_{P'(\nu+1)}{}^0-x_{P'(\nu)}{}^0) \\ &\qquad\times\prod_{\nu=j}^{n+1}\theta(x_{P'(\nu)}{}^0-x_{P'(\nu-1)}{}^0)w_{P'}(x). \end{aligned} \tag{2.15}$$

Because the time components appear explicitly in Eq. (2.15), we consider the n-dimensional vector space T formed by the time component of $x\in X$,

$$x^0=(x_1{}^0\cdots x_{n+1}{}^0) \mod(1\cdots1), \tag{2.16}$$

and the n-dimensional vector space S formed by the energy component of $q\in Q$. We use the following inner products:

$$q\cdot t=\sum_{i=1}^{n+1}q_it_i, \quad s\cdot x=\sum_{i=1}^{n+1}s_ix_i, \tag{2.17}$$

$$s\cdot t=\sum_{i=1}^{n+1}s_it_i \tag{2.18}$$

where $x\in X$, $q\in Q$, $t\in T$, and $s\in S$. The inner products in Eq. (2.17) are Minkowski vectors while the inner product in Eq. (2.18) is a number. The space Q is the dual of X relative to the inner product (2.6), and S is the dual of T relative to the inner product in Eq. (2.18). The complex vector spaces corresponding to X, Q, T, and S are denoted by Z, Z', U, and V, respectively. Equations (2.6), (2.17), and (2.18) are used also for these spaces.

[9] To prove Eq. (2.15), we note that the vacuum expectation value of the multiple commutator for each P in Eq. (2.14) contains a fixed $w_{P'}(x)$ in Eq. (2.15), if, and only if,

$$\begin{aligned} &P^{-1}P'(1)>P^{-1}P'(2)>\cdots>P^{-1}P'(j-1), \\ &P^{-1}P'(j+1)<\cdots<P^{-1}P'(n+1). \end{aligned}$$

Since $(j-1)$ A's always come to the left of $A_1(x_1)$, the $w_{P'}(x)$ in all these terms have a common sign $(-1)^{j-1}$. Summing up θ functions over all P satisfying the previous equation, we get Eq. (2.15).

If we define $t(I)$ by

$$\begin{aligned} t(I)_\nu&=1 \quad \text{if } \nu\in I \\ &=0 \quad \text{if } \nu\notin I, \end{aligned} \tag{3.19}$$

the $q(I)$ can be written as $q\cdot t(I)$. In a similar manner, we define

$$s(ij)_\nu=\delta_{i\nu}-\delta_{j\nu}, \tag{2.20}$$

which will be used to express x_i-x_j as $s(ij)\cdot x$. By using the notation of Eq. (2.19), we can write the Fourier-Laplace transform of r as

$$\begin{aligned} \tilde{r}(v,\mathbf{q})=\sum_P\int dq^0\tilde{w}_P(q)(2\pi)^{-n} \\ \times\prod_{I\in\mathfrak{I}(P)}[(v-q^0)\cdot t(I)]^{-1}. \end{aligned} \tag{2.21}$$

Here dq^0 is defined in an analogous way to dq in Eq. (2.10). The w_P in Eqs. (2.15) and (2.21) can be replaced with the $w_P{}^T$ as will be seen in Appendix B. Because of (W3), $\tilde{r}(v,\mathbf{q})$ is analytic everywhere except at the cuts

$$\operatorname{Im}v\cdot t(I)=0, \quad \operatorname{Re}v\cdot t(I)\geqslant\{m^2+[\mathbf{q}\cdot t(I)]^2\}^{\frac{1}{2}}. \tag{2.22}$$

If we fix the sign of every $\operatorname{Im}v\cdot t(I)$, and let $\operatorname{Im}v$ tend to zero, then $r(v,\mathbf{q})$ approaches to one boundary value. Geometrically speaking, the family $H_{n-1}{}^R$ of hyperplanes (in the space S of $\operatorname{Im}v$) defined by

$$\begin{aligned} H_{n-1}{}^R&=\{h(I):I\subset I(n+1)\}, \\ h(I)&=\{s; s\cdot t(I)=0\} \end{aligned} \tag{2.23}$$

divides the entire space S into several convex polyhedral cones which we shall call C_i. If $\operatorname{Im}v$ stays in the interior of one cone C_i, then the sign of $\operatorname{Im}v\cdot t(I)$ stays constant, while if it moves from one cone to another the sign of some $\operatorname{Im}v\cdot t(I)$ changes. Thus as $\operatorname{Im}v$ tends to zero from inside each cone C_i, $\tilde{r}(v,\mathbf{q})$ approaches to one of its boundary values which we shall call $r_i(q)$. The $r_i(q)$ exhaust all boundary values of $\tilde{r}(v,\mathbf{q})$. In particular, we obtain the Fourier transform of the retarded function (2.15) as the boundary value corresponding to the cone $\operatorname{Im}v\cdot t(I)\leqslant0$ for $I=\{2\},\{3\},\cdots\times\{n+1\}$, i.e., for $\operatorname{Im}v_i\leqslant0$ for $i=2,\cdots,n+1$.

We shall use the generalized θ function

$$\begin{aligned} \theta(t:c)&=1 \quad \text{if } t\in C \\ &=0 \quad \text{if } t\notin C \end{aligned} \tag{2.24}$$

If C is a pointed convex polyhedral cone,[10] the Fourier-Laplace transform of θ,

$$\tilde{\theta}(v;C)=\int e^{iv\cdot t}\theta(t;C)dt, \tag{2.25}$$

is a rational function of v. Its boundary value (considered as a distribution), as $\operatorname{Im}v$ tends to zero from within a cone C' of the space of $\operatorname{Im}v$, is denoted by

[10] For the definition, see Appendix C.

$\tilde{\theta}(s;C/C')$ and its inverse Fourier transform is denoted by $\theta(t;C/C')$. If C' is the positive polar[10] of C, then $\theta(t;C/C')$ is equal to $\theta(t;C)$. Similar definitions hold for $\theta(s;C)$, $\tilde{\theta}(u;C)$, $\tilde{\theta}(t;C/C')$, and $\theta(s;C/C')$. The properties of these functions will be studied in Sec. 3.

As an example, let us consider the cones C_P in T defined by

$$C_P=\{x^0;x_{P(1)}{}^0\geqslant x_{P(2)}{}^0\geqslant\cdots\geqslant x_{P(n+1)}{}^0\}. \quad (2.26)$$

Then Eq. (2.20), with w_P replaced by $w_P{}^T$, can be written as

$$\tilde{r}(v,\mathbf{q})=\sum_P\int dq^0\tilde{w}_P{}^T(q)\tilde{\theta}(v-q^0;C_P)(2\pi i)^{-n}. \quad (2.27)$$

We remark that although the starting Eq. (2.14) referred to the Bose case, Eq. (2.27) is the appropriate definition of the retarded function for an arbitrary collection of local Bose and Fermi fields, i.e., theorems 1 and 2 following are true always.

Our first main theorem lists the necessary and sufficient condition for the r_i to be obtainable from the $w_P{}^T$ satisfying (W1)–(W3).

Theorem 1.[11] If $w_P{}^T(x)$ satisfies (W1)–(W3), then $r_i(x)$ defined by

$$r_i(x)=(-i)^n\sum_P\theta(x;C_P/C_i)w_P{}^T(x) \quad (2.28)$$

satisfies

(R1) $r_i(x)$ is a covariant function of $x\in X$.

(R2) $r_i(x)=0$, if $x^0\notin C_i{}^+$. (C^+ is the positive polar[10] of C.)

(R3) $\tilde{r}_i(q)=\tilde{r}_j(q)$, if $\dim[C_i\cap C_j\cap h(I)]=n-1$[12] and $q(I)^2<m^2$.

(R4) $r_{++}(x)-r_{+-}(x)-r_{-+}(x)+r_{--}(x)=0$, if $\dim[C_{++}\cap C_{+-}\cap C_{-+}\cap C_{--}\cap h(I)\cap h(I')]=n-2$,[12] $\sigma I\cap\sigma' I'\neq$ empty, $C_{\sigma\sigma'}\subset C[t(I)]^\sigma\cap C[t(I')]^{\sigma'}$ $(\sigma,\sigma'=+\text{ or }-)$.

Conversely, if $r_i(x)$ satisfies (R1)–(R4), then $w_P{}^T(x)$, defined by

$$\tilde{w}_P{}^T(q)=(i)^n\sum_i\theta(q^0;C_i/C_P)\tilde{r}_i(q), \quad (2.29)$$

satisfies (W1)–(W3), and the original $r_i(x)$ is given by Eq. (2.28) in terms of this $w_P{}^T$.

Remarks. (1) Note that the conditions (R1), (R2), and (R3) in this theorem are almost dual to the conditions (W1), (W3), and (W2). In fact, (W2) can be rewritten in our notation as (W2′) $W_P{}^T(x)=W_{P'}{}^T(x)$, if $\dim[C_P\cap C_{P'}\cap h(ij)]=n-1$[13] and if $[s(ij)\cdot x]^2<0$, where $h(ij)$ is the hyperplane orthogonal to $s(ij)$.

(2) The support condition in x space, (R2), expresses the retardedness in certain variables. Namely, if we denote the 1-facets[10] of C_i by $C(s_i{}^\lambda)$, then (R2) is equivalent to (R2′) $r_i(x)=0$, unless $s_i{}^\lambda\cdot x\in\bar{V}_+$ (the future light cone) for all λ. Actually, r_i has in general more retardedness than (R2′), which, however, invariably contains alternative statements. This retardedness is, of course, implied by (R1)–(R4), but is not immediately apparent.[14]

(3) The condition (R4) has been first noted by Steinmann[4] for the four-point function ($n=3$). The intersection of two $(n-1)$-planes[10] $h(I)$ and $h(I')$ $(I\neq I')$ is a $(n-2)$-plane.[10] This intersection is not contained in any other $h(I'')$, if, and only if, $\pm I$ and $\pm I'$ has nonempty intersection for any combination of the signs where we have denoted $I(n+1)-I$ by $-I$. If this is the case, the $(n-2)$-plane $h(I)\cap h(I')$ is divided into several polyhedral convex cones by $h(I'')(I''\neq I,I')$ and corresponding to each of these cones, there are exactly four cones C_i which have that cone as a $(n-2)$-facet[10] and which are on different sides of $(n-1)$-planes $h(I)$ and $h(I')$. The condition (R4) gives a linear relation among the corresponding four r_i which are denoted by $r_{\sigma\sigma'}$ $(\sigma,\sigma'=+\text{ or }-)$.

Our second main task is to convert conditions (R1)–(R4) on r_i to a condition on the domain of analyticity of the analytic function in p space. We have succeeded in this only for (R1)–(R3).

To state our result, we need further definitions. We define open convex cones $V_i{}^Q$ in Q by

$$V_i{}^Q=\{q;q\cdot t(I)\in V_+,I\in\mathfrak{I}_i\}, \quad (2.30)$$

where V_+ is the interior of the future light cone and $\mathfrak{I}_i$ is the set of $I\subset I(n+1)$ such that $C[t(I)]$, $I\in\mathfrak{I}_i$ constitute the one-facets of $C_i{}^+$. [The $h(I)$, $I\in\mathfrak{I}_i$ are boundary planes of C_i.] If C_i and C_j are neighboring cones across the $(n-1)$-plane $h(I_0)$, namely,

$$\dim[C_i\cap C_j\cap h(I_0)]=n-1,$$

the interior of the set $(\bar{V}_i{}^Q\cap\bar{V}_j{}^Q)$ is denoted by $S^Q(ij)$:

$$S^Q(ij)=\{q;q\cdot t(I_0)=0,\quad q\cdot t(I)\in V_+\text{ for }I\in\mathfrak{I}_i\text{ or }\mathfrak{I}_j\text{ and }I\neq I_0\}. \quad (2.31)$$

The tube $T(V_i{}^Q)$ is the subset of Z' defined by

$$T(V_i{}^Q)=\{\zeta\in Z';\operatorname{Im}\zeta\in V_i{}^Q\}. \quad (2.32)$$

The extended tube $T'(V_i{}^Q)$ is the union of images of $T(V_i{}^Q)$ under all complex proper Lorentz transforma-

[11] To be precise, we have to specify the class of distributions to which $w_P(x)$ and $r_i(x)$ belong. The point is that a product like $\theta(x;C_P/C_i)w_P(x)$ or $\theta(q;C_i/C_P)r_i(q)$ has to be well defined and the integral over dq^0 or dx^0 has to be convergent. In this paper we do not attempt any thorough discussion of this point, although we shall make a few remarks in Sec. 7. See also footnote 16.

[12] This means C_i and C_j are neighboring cones with their common $(n-1)$-facet lying on $h(I)$. The cones in (R4) will be explained.

[13] C_P and $C_{P'}$ are neighboring cones with their common $(n-1)$-facet lying on $h(ij)$.

[14] For example, take $r_{12}(x_1\cdots x_4)$ for the fourfold case. (See H. Araki and N. Burgoyne, footnote 3.) This vanishes unless x_1 is advanced over x_3 and x_4 and x_2 is advanced over *either* x_3 *or* x_4. (R2) says that it vanishes unless (x_1-x_3), (x_1-x_4), and $(x_1+x_2-x_3-x_4)\in\bar{V}_+$. Of course, the latter and (R4) imply the former.

tions. The corresponding definitions in X are

$$V_P{}^X=\{x\in X; s[P(k), P(k+1)]\cdot x\in V_-, \quad k=1, \cdots, n\}, \tag{2.33}$$

$$S(P,k)=\{x\in X; s[P(k), P(k+1)]\cdot x=0, \quad s[P(m), P(m+1)]\cdot x\in V_- \text{ for } m\neq k\} \tag{2.34}$$

$$T(V_P{}^X)=\{z\in Z; \operatorname{Im} z\in V_P{}^X\}. \tag{2.35}$$

If the two cones C_P and $C_{P'}$ are neighboring, namely, if $P(i)=P'(i)$ for $i\neq k$, $k+1$ and $P(k)=P'(k+1)$, $P(k+1)=P'(k)$, then

$$S(P,k)=S(P',k)=\text{the interior of } \bar{V}_P{}^X\bigcap\bar{V}_{P'}{}^X. \tag{2.36}$$

We are now ready to state our second main theorem.

Theorem 2. The $\tilde{r}_i(q)$ satisfying (R1)–(R3) are boundary values of one analytic function $\tilde{r}(\zeta)$ as ζ tends to q from inside the tube $T(V_i{}^Q)$. $\tilde{r}(\zeta)$ is analytic in the union of $T'(V_i{}^Q)$ for all possible i and in the sets

$$\Sigma(ij,m)=\{\zeta\in Z'; \operatorname{Im}\zeta\in S^Q(ij), [\operatorname{Re}\zeta\cdot t(I)]^2<m^2\} \tag{2.37}$$

for all i, j, I, such that C_i and C_j is neighboring across $h(I)$. $\tilde{r}(\zeta)$ is analytic at a real point $\zeta=q$, if all $q(I)^2$ are smaller than m^2. Conversely, if $r(\zeta)$ is analytic in this region and has a certain boundedness property,[15] then its boundary values $\tilde{r}_i(q)$ satisfy (R2) and (R3).

This will be proved in Sec. 6. For the sake of comparison, we mention the corresponding theorem for $w_P{}^T$.

Theorem 3. The $w_P{}^T(x)$ satisfying (W1)–(W3) are boundary values of one analytic function $w^T(z)$ as z tends to x from inside the tube $T(V_P{}^X)$. $w^T(z)$ is analytic in the union of $T'(V_P{}^X)$ for all possible P and in the sets

$$\Sigma(P,k)=\{z\in Z; \operatorname{Im} z\in S(P,k), \quad (\operatorname{Re} s[P(k), P(k+1)]\cdot z)^2<0\}, \tag{2.38}$$

$w^T(z)$ is analytic at a real point $z=x$, if all $s(ij)\cdot x$ are spacelike. Conversely, if $w^T(z)$ is analytic in the above region and satisfies a certain boundedness condition,[15] then its boundary values $w_P{}^T(x)$ satisfy (W2) and (W3).

Covariant formulas which express $\tilde{r}(\zeta)$ and $w(z)$ in terms of boundary values of the other are given by

$$\tilde{r}(\zeta)=(-i)^n\sum_\alpha\sum_\nu\int dx e^{i(\zeta,x)}\theta(x^0; C_{\alpha\nu}{}^X/\operatorname{Im}\zeta^0) \times\theta(x;\Delta_\alpha{}^X)w_{\alpha\nu}{}^T(x), \tag{2.39}$$

$$w^T(z)=(i)^n\sum_\beta\sum_\nu\int dq e^{-i(q,z)}\theta(q^0; C_{\beta\nu}{}^Q/\operatorname{Im} z^0) \times\theta[q;\Delta_\beta{}^Q(m)]\tilde{r}_{\beta\nu}(q). \tag{2.40}$$

Here $\Delta_\alpha{}^X$ and $\Delta_\beta{}^Q(m)$ designate various regions in X or Q where $w(z)$ or $\tilde{r}(\zeta)$ have different number of boundary values. Namely, we divide the space X into several $\Delta_\alpha{}^x$ according to whether each $s(ij)\cdot x$ is space- or timelike and the different regions are distinguished by subscript α. A similar definition holds for $\Delta_\beta{}^Q(m)$:

$$\Delta_\alpha{}^X=\{x\in X; \sigma_\alpha(ij)[s(ij)\cdot x]^2>0\}, \tag{2.41}$$

$$\Delta_\beta{}^Q(m)=\{q\in Q; \sigma_\beta(I)[(q\cdot t(I))^2-m^2]>0\}, \tag{2.42}$$

where σ_α and σ_β are $+$ or $-$. For each region $\Delta_\alpha{}^X$ vectors $[s(kl)\cdot x]$ with $\sigma_\alpha(kl)>0$ can be either positive or negative timelike. To distinguish such possibilities, we use the cones $C_{\alpha\nu}{}^X$ in T which are defined by

$$C_{\alpha\nu}{}^X=\{t\in T; [s(kl)\cdot t]\sigma_{\alpha\nu}(kl)>0 \text{ for all } k, l \text{ such that } \sigma_\alpha(kl)>0\}, \tag{2.43}$$

where as ν varies $\sigma_{\alpha\nu}(kl)$ exhaust all possibilities for consistent assignments of signs to $s(kl)\cdot t$. For example, if all $\sigma_\alpha(kl)>0$, then $\{C_{\alpha\nu}\}$ coincides with $\{C_P\}$. In general, $C_{\alpha\nu}$ is a union of several C_P. $C_\beta{}^Q$ are similarly defined and coincides with $\{C_i\}$, if $\sigma_\beta(I)>0$ for all I. The summation over α in Eq. (2.39) extends over α such that the C_α are pointed. [In other words, if the $s(kl)$ for which $\sigma_\alpha(kl)>0$ span S.] For each α, the summation over ν extends over all possibilities. Similar prescription applies for the summations in Eq. (2.40). $\theta(x^0; C_{\alpha\nu}{}^x/\operatorname{Im}\zeta^0)$ is the $\theta(x^0; C_\alpha{}^x/C')$ where C' is determined by $\operatorname{Im}\zeta^0\in C'$. It is invariant if $x\in\Delta_\alpha{}^X$ and all $\operatorname{Im}\zeta\cdot t(I)$ are time- or lightlike. $\theta(q^0; C_{\beta\nu}{}^Q/\operatorname{Im} z^0)$ is similarly defined.

$w_\alpha{}^T$ is the $w_P{}^T$ with P such that $C_P\subset C_{\alpha\nu}{}^X$. Owing to (W2), if $x\in\Delta_\alpha{}^X$, then the $w_P{}^T(x)$ are all equal for different P as long as C_P stays in one $C_\alpha{}^X$, $r_{\beta\nu}$ is the r_i with i such that $C_i\subset C_{\beta\nu}{}^Q$.

Finally, we note that the vacuum expectation value of time ordered product, $\tau(x)$, and its Fourier transform $\tilde{\tau}(q)$ can be expressed as

$$\tau(x)=\lim_{z\to x,\ \operatorname{Im} z\in V_T(x)} w(z), \tag{2.44}$$

$$\tilde{\tau}(q)=\lim_{\zeta\to q,\ \operatorname{Im}\zeta\in V_T(q)} i^n r(\zeta) \tag{2.45}$$

where V_T and $\tilde{V}_T$ are defined by

$$V_T(x)=V_P{}^X, \quad \text{if } x^0\in C_P; \qquad \tilde{V}_T(q)=V_i{}^Q, \quad \text{if } q^0\in C_i. \tag{2.46}$$

3. PROPERTIES OF GENERALIZED θ FUNCTION

First let us consider the generalized θ function defined by Eq. (2.24) for the special case of a simplex cone C.[10] Suppose 1-facets of C and C^+ are $t_1\cdots t_n$ and $s_1\cdots s_n$ where $s_i\cdot t_j=\delta_{ij}$, $[\det(t_i)\neq 0]$. Then we have

$$\theta(t;C)=\prod_{i=1}^n\theta(s_i\cdot t), \tag{3.1}$$

$$\tilde{\theta}(v;C)=i^n|\det(t_i)|\prod_{i=1}^n(v\cdot t_i)^{-1}. \tag{3.2}$$

[15] Compare L. Schwartz, Medd. Lunds Mat. Sem. Suppl. 196 (1952).

If we define associated simplex cones σC by

$$\sigma C = C(\sigma_1 t_1 \cdots \sigma_n t_n), \tag{3.3}$$

where the σ_i are ± 1, then, we have the formulas

$$\tilde{\theta}(v; \sigma C) = (\prod_{i=1}^{n} \sigma_i)\tilde{\theta}(v; C), \tag{3.4}$$

$$\theta(t; C/\sigma C^+) = (\prod_{i=1}^{n} \sigma_i)\theta(t; \sigma C).^{16} \tag{3.5}$$

We note that the poles of $\tilde{\theta}(v; C)$ appear at $v \cdot t_i = 0$, $i = 1 \cdots n$ and the discontinuity of $\theta(t; C/\sigma C^+)$ appears at $s_i \cdot t = 0$, $i = 1 \cdots n$.

We now turn to the case of general convex polyhedral cones C.

Lemma 1. Let C be a pointed polyhedral convex cone.[10] The integral in Eq. (2.25) defines an analytic function of v in the tube $T(C^+) = \{v;\ \mathrm{Im}\, v \in \text{interior of } C^+\}$ (which is nonempty). The analytic function is a rational function with simple poles at $v \cdot t = 0$, for $t \in F_1(C)$ (the set of all 1-facets of C).

Lemma 2. (Addition theorem.) Let C and C_α be convex polyhedral cones such that C is the union of C_α and the C_α are mutually almost disjoint [$C = \bigcup_\alpha C_\alpha$, $\dim(C_\alpha \cap C_\beta) < n$ for $\alpha \neq \beta$]. Then

$$\sum_\alpha \tilde{\theta}(v; C_\alpha) = \tilde{\theta}(v; C) \quad \text{if } \mathrm{lin}\, C = 0 \tag{3.6}$$

$$= 0 \quad \text{if } \mathrm{lin}\, C \neq 0 \quad \mathrm{lin}\, C_\alpha = 0 \tag{3.7}$$

$$\sum_\alpha \tilde{\theta}(u; C_\alpha^+) = \tilde{\theta}(u; C^+) \quad \text{if } \dim C^+ = n \tag{3.8}$$

$$= 0 \quad \text{if } \dim C_\alpha^+ = n \text{ and } \dim C^+ \neq n. \tag{3.9}$$

For the proof, we first note that if $v \in T(C^+)$, then $\mathrm{Im}\, v \cdot t > 0$ for $t \in C$ and as $t \to \infty$ within C, the integrand of Eq. (2.25) tends to zero exponentially. Hence it defines an analytic function of v. Next, we obviously have

$$\theta(t; C) = \sum_\alpha \theta(t; C_\alpha) \text{ almost everywhere.} \tag{3.10}$$

Because $C_\alpha^+ \supset C^+$ and C^+ is nonempty, the integral representation Eq. (2.25) can be applied to all $\tilde{\theta}(v; C_\alpha)$ and $\tilde{\theta}(v; C)$, if $v \in T(C^+)$. Hence we obtain Eq. (3.6) from Eq. (3.10) as a relation between analytic functions. To prove that $\tilde{\theta}(v; C)$ is rational, we invoke the simplexial decomposition of C: $C = \bigcup_\alpha C_\alpha$. We already know that, for simplex cones C_α, $\tilde{\theta}(v; C_\alpha)$ is rational. Hence $\tilde{\theta}(v; C)$ is also rational by Eq. (3.6). Moreover, because $F_1(C_\alpha) \subset F_1(C)$ for *standard* simplexial decomposition and the latter is possible, if $\mathrm{lin}\, C = 0$,[17] we see that the singularities of $\tilde{\theta}(v; C)$ occur only at $v \cdot t = 0$, $t \in F_1(C)$.

To prove Eq. (3.7), we first consider a special case where $C = \bigcup_\sigma C_\sigma$, $C_\sigma = C(\sigma_1 t_1 \cdots \sigma_m t_m, t_{m+1} \cdots t_n)$, $\dim C_\sigma = n$, and $\sigma_i = \pm 1$. Since C_σ is simplex, we easily obtain Eq. (3.7) from Eq. (3.4). By using this result, we make generalizations in two steps. First consider the case where $C = \bigcup_\sigma C_\sigma$, $C_\sigma = C(T_0 \bigcup T_\sigma)$, $\dim C(T_0) = n - m$, $\mathrm{lin}\, C(T_0) = 0$, $T_\sigma = \{\sigma_1 t_1 \cdots \sigma_m t_m\}$, $\sigma_i = \pm 1$ and $\dim C_\sigma = n$. We make a simplexial decomposition of $C(T_0)$ in $h(T_0)$: $C(T_0) = \bigcup_\beta C(T_\beta)$. On setting $C_{\beta\sigma} = C(T_\beta \bigcup T_\sigma)$ and $C_\beta = \bigcup_\sigma C_{\beta\sigma}$, and using Eq. (3.6) for $C_\sigma = \bigcup_\beta C_{\beta\sigma}$ and Eq. (3.7) for $C_\beta = \bigcup_\sigma C_{\beta\sigma}$, we have $\Sigma_\sigma \tilde{\theta}(v; C_\sigma) = \Sigma_\beta [\Sigma_\sigma \tilde{\theta}(v; C_{\beta\sigma})] = 0$. Finally, for the most general case, let $C = \bigcup_\alpha C_\alpha$, $\mathrm{lin}\, C = m$, $L(C) = h(\Sigma)^\perp$, and $\Sigma = \{s_1 \cdots s_m\}$. Let $\Sigma_\sigma = \{\sigma_1 s_1 \cdots \sigma_m s_m\}$, $C_\sigma = C(\Sigma_\sigma)^+ \bigcap C$, and $C_{\alpha\sigma} = C_\alpha \bigcap C_\sigma$. Since $\mathrm{lin}\, C_\sigma = 0$ by construction, we have $\tilde{\theta}(v; C_\sigma) = \Sigma_\alpha \tilde{\theta}(v; C_{\alpha\sigma})$ as a result of Eq. (3.6). Since $\mathrm{lin}\, C_\alpha = 0$ by assumption, we have $\tilde{\theta}(v; C_\alpha) = \Sigma_\sigma \theta(v; C_{\alpha\sigma})$. By Eq. (3.7) for the previously proved case, we have $\Sigma \tilde{\theta}(v; C_\sigma) = 0$. On combining these, we obtain Eq. (3.7) for the most general case.

Equations (3.8) and (3.9) can be proved at the same time. [If $\dim C \neq n$, $\tilde{\theta}(u; C) = 0$.] First, consider a special case where $C = C_1 \bigcup C_2$, $C_1 = C \bigcap C(-s)^+$, $C_2 = C \bigcap C(s)^+$, and s, $-s \notin C^+$. The $(n-1)$-planes in $H_{n-1}(C^+)$[18] divide C_1^+ and C_2^+ into several convex cones. Let this decomposition be $C_1^+ = C^+ \bigcup (\bigcup_\alpha C_\alpha^+)$ and $C_2^+ = C^+ \bigcup (\bigcup_\beta C_\beta^+)$. Since C_1^+ and C_2^+ are pointed, we have from Eq. (3.6) $\tilde{\theta}(u; C_1^+) = \tilde{\theta}(u; C^+) + \Sigma \tilde{\theta}(u; C_\alpha^+)$ and $\tilde{\theta}(u; C_2^+) = \tilde{\theta}(u; C^+) + \Sigma \tilde{\theta}(u; C_\beta^+)$. Since $C_1^+ \bigcup C_2^+$ is not pointed, we have from Eq. (3.7) $\tilde{\theta}(u; C^+) + \Sigma \tilde{\theta}(u; C_\alpha^+) + \Sigma \tilde{\theta}(u; C_\beta^+) = 0$. Hence we obtain Eqs. (3.8) and (3.9) for this case. Next, consider the case where C is cut into several C_α by a family of planes $h(s)^\perp$, $s \in S_0$. By applying the previous result, every time one cuts C by a $h(s)^\perp$, one obtains Eq. (3.8) or (3.9) for $C = \bigcup C_\alpha$. Finally, consider the most general case $C = \bigcup C_\alpha$. The $(n-1)$ planes in $\bigcup_\alpha H_{n-1}(C_\alpha)$ cut C and C_α into several convex cones. Let this decomposition be $C_\alpha = \bigcup_i C_{\alpha i}$ and $C = \bigcup_{\alpha i} C_{\alpha i}$. Then by applying the previous result for C_α and C, we obtain Eqs. (3.8) and (3.9). This completes the proof of lemmas 1 and 2.

Next we investigate the residue of $\tilde{\theta}$ at its pole. We define

$$R(v; C/t) = \lim_{v' \to v} v' \cdot t \tilde{\theta}(v'; C), \quad v \cdot t = 0 \tag{3.11}$$

$$R(v; C/t_1 \cdots t_m) = \lim_{v' \to v} v' \cdot t_m R(v'; C/t_1 \cdots t_{m-1}),$$
$$v \cdot t_i = 0, \quad i = 1 \cdots m. \tag{3.12}$$

Lemma 3.

$$R(v; C/t_1) = i\epsilon(C; f_1)\tilde{\theta}_1(v; C_1'), \tag{3.13}$$

$$R(v; C/t_1 \cdots t_n) = i^n |\det(t_i)| \prod_{m=1}^{n} \epsilon(C_{m-1}; f_m), \tag{3.14}$$

[16] Note that $\theta(t; C/\sigma C^+)$ is defined only almost everywhere. Equation (3.5) should be taken in this sense. The product like $\theta(t)w(t)$ is meaningful only when $w(t)$ belongs to a certain class of distribution. See L. Schwartz, Seminaire Schwartz-Levy, No. 3, Faculté des Sciences de Paris, 1956–57.

[17] Compare lemma C2 in Appendix C.

[18] $F_m(C)$ is the set of all m-facets of C and $H_m(C)$ is the set of dimensionality spaces of all m-facets of C:

$$H_m(C) = \{h(f);\ f \in F_m(C)\}.$$

where

$$C_m=C+h(t_1\cdots t_m),\quad C_0=C,\quad f_m=C(t_m)+h(t_1\cdots t_{m-1}).$$
$$\epsilon(C;f)=1 \text{ if } f\in F_m(C)^{18} \quad \text{for some } m$$
$$=-1 \text{ if } -f\in F_m(C) \quad \text{for some } m, \quad (3.15)$$
$$=\text{otherwise},$$

$\tilde{\theta}_1$ is the $\tilde{\theta}$ where the space T mod $h(t_1)$ is used instead of T and $h(t_1)^\perp$ instead of V. The volume element of T mod $h(t_1)$ in the definition of $\tilde{\theta}_1$ is so chosen that, if t_1, $t_2'\cdots t_n'$ span a parallelepiped of unit volume, $t_2'\cdots t_n'$ span the same in T mod $h(t_1)$.

To prove Eq. (3.13), we note that $R(v;C/t_1)$ is a rational function of v in $h(t_1)^\perp$ in V. We can calculate R by

$$-2\pi i\delta(v\cdot t_1)R(v;C/t_1)$$
$$=\lim_{\epsilon\to 0}[\tilde{\theta}(v+i\epsilon;C)-\tilde{\theta}(v-i\epsilon;C)] \quad (3.16)$$

where $\text{Im } v\cdot t_1=0$ and $\epsilon\cdot t_1>0$. From lemma 1, we have

$$R(v;C/t_1)=0, \text{ unless } t_1 \text{ or } -t_1\in F_1(C). \quad (3.17)$$

Suppose $C(\sigma t_1)\in F_1(C)(\sigma=\pm)$. Because of Eq. (3.17) and the addition theorem (3.6), we can adjoint to C or cut off from C any convex cones whose 1-facets do not contain $\pm t_1$ without changing $R(v;C/t_1)$. By this process, we can shift all $(n-1)$-facets of C not containing $\pm t_1$, to one facet f. Suppose $f\perp s$ and $s\cdot t_1>0$. On denoting

$$C_1=C+h(t_1),\quad C'=C_1\cap C(s)^\sigma,\quad C''=C_1\cap C(-s)^\sigma,$$

we have

$$R(v;C/t_1)=R(v;C'/t_1).$$

On the other hand, we know from Eq. (3.7) that

$$\tilde{\theta}(v-i\sigma\epsilon;C')=-\tilde{\theta}(v-i\sigma\epsilon;C'').$$

From these we obtain

$$-2\pi i\delta(v\cdot t_1)R(v;C/t_1)$$
$$=\lim_{\epsilon\to 0}\sigma[\tilde{\theta}(v+i\sigma\epsilon;C')+\tilde{\theta}(v-i\sigma\epsilon;C'')]. \quad (3.18)$$

Since R is rational function, we can easily find an open set $\mathcal{O}$ (relative to $h(t_1)^\perp$) in domain of analyticity of R and ϵ satisfying $\epsilon\cdot t_1>0$, such that $\sigma\epsilon+\text{Im } v\in C'$ and $-\sigma\epsilon+\text{Im } v\in C''^+$ when $v\in\mathcal{O}$. We can use the integral representation for both $\tilde{\theta}$ in Eq. (3.18) for such v and ϵ, and we obtain

$$-2\pi i\delta(v\cdot t_1)R(v;C/t_1)=\int e^{iv\cdot t}\theta(t;C_1)dt.$$

Thus Eq. (3.13) is true for $v\in\mathcal{O}$. Since both sides of Eq. (3.13) are rational, it holds everywhere.

By repeated application of Eq. (3.13), we obtain

$$[\prod_{m=1}^{n}2\pi\delta(s\cdot t_m)]R(v;C/t_1\cdots t_n)$$
$$=i^n\prod_{m=1}^{n}\epsilon(C_{m-1}/v_m)\int e^{is\cdot t}dt.$$

which implies Eq. (3.14). This completes the proof of lemma 3.

We now discuss the boundary values of $\tilde{\theta}$.

Lemma 4. The boundary value of $\tilde{\theta}$

$$\tilde{\theta}(s;C/s')=\lim_{k\to+0}\tilde{\theta}(s+iks';C) \quad (3.19)$$

is the same for all s' in the interior of any one cone C' of $\Gamma[H_{n-1}(C^+)]$.[19]

This is obvious if one recalls that $\tilde{\theta}(v;C)$ is rational and its poles appear only when $\text{Im } v$ is on one of planes in $H_{n-1}(C^+)=H_1(C)^\perp$.

This justifies the notation $\tilde{\theta}(s;C/C')$ instead of $\tilde{\theta}(s;C/s')$, as long as C' is a cone of $\Gamma[H_{n-1}(C^+)]$ or contained in such a cone.

Lemma 5. The Fourier transform of $\tilde{\theta}(s;C/C')$,

$$\theta(t;C/C')=\int e^{-is\cdot t}\tilde{\theta}(s;C/C')ds(2\pi)^{-n} \quad (3.20)$$

is a function taking integral values (almost everywhere) and with discontinuities only at planes belonging to $H_{n-1}(C)$. Furthermore,

$$\theta(t;C/C')=0 \quad \text{if } t\notin C'^+ \quad (3.21)$$

$$\theta(t;C/s')=0 \quad \text{if } t\in \text{ interior of } C \text{ and } s'\notin C^+. \quad (3.22)$$

To prove the first part of the lemma, we note that this is true for simplex C [cf. Eq. (3.5)]. For arbitrary C, we see by a simplexial decomposition $C=\Sigma C_\alpha$, that discontinuities occur on $(n-1)$-planes. Furthermore, if a $(n-1)$-plane $h\notin H_{n-1}(C)$, then by lemma C2, we can make this decomposition in such a way that $h\notin H_{n-1}(C_\alpha)$ for any α. Hence discontinuities occur only on planes of $H_{n-1}(C)$.

To prove Eq. (3.21),[20] we note that if $t\notin C'^+$ then there is a $s_1\in C'$ such that $s_1\cdot t<0$. On using a basis $s_1\cdots s_n$ in S,

$$\theta(t;C/C')=\int e^{-i\Sigma\rho_j s_j\cdot t}|\det(s_j)|\,\tilde{\theta}(\Sigma\rho_j s_j;C/C')\prod d\rho_j.$$

Since $\tilde{\theta}$ is analytic for $\text{Im } \rho_1>0$ with fixed real ρ_j, $j\geqslant 2$, we have Eq. (3.21) by contour deformation in the ρ_1 integration.

[19] $\Gamma(H)$ is the set of all convex polyhedral cones obtained by division of the whole space by $(n-1)$ planes belonging to H. See Appendix C.

[20] Another proof can be obtained by using a standard simplexial decomposition of C: $C=\cup C_\alpha$. Then $\theta(t;C/C')=\Sigma_\alpha\theta(t;C_\alpha/C')$. Since $F_1(C_\alpha)\subset F_1(C)$ and $C'\in\Gamma(H_1(C)^\perp)$, C' is contained in one of $\sigma C_\alpha{}^+$ [defined by Eq. (3.3)]. By Eq. (3.5), if $t\notin C'^+\supset\sigma C_\alpha$, then $\theta(t;C_\alpha/C')=\theta(t;C_\alpha/\sigma C_\alpha{}^+)=0$.

To prove Eq. (3.22), we make a simplexial decomposition of C^+: $C^+=\bigcup C_\alpha{}^+$. Obviously $s'\notin C_\alpha{}^+$. Since $\tilde\theta(v;C)=\Sigma\tilde\theta(v;C_\alpha)$, because of Eq. (3.8), we have $\theta(t;C/s')=\Sigma\theta(t;C_\alpha/s')$. If s' happens to be on some plane of $H_{n-1}(C_\alpha{}^+)$, there is always another s'' near s' which is not on any plane of $H_{n-1}(C_\alpha{}^+)$ nor in C^+ and satisfies $\theta(t;C/s')=\theta(t;C/s'')$. [$s'\notin H_{n-1}(C^+)$.] For simplex C_α, we see from Eq. (3.5) that $\theta(t;C_\alpha/s')=0$, if $t\in C\subset C_\alpha$ and $s''\notin C_\alpha{}^+$. Hence we have Eq. (3.22).

Finally we prove the following inversion formula.

Lemma 6. If $\dim C=n$, $\lin C=0$, and $H\supset H_{n-1}(C^+)$, then

$$\sum_{C'\in\Gamma(H)}\tilde\theta(v;C')\theta(t;C/C')=\tilde\theta(v;C^+). \tag{3.23}$$

To prove this, we first consider the case where C is simplex. Since $H\subset H_{n-1}(C^+)$, each $C'\in\Gamma(H)$ is contained in some $C_\sigma{}^+$. By lemma 4, Eqs. (3.6), (3.4), and (3.5), we obtain

$$\begin{aligned}\sum_{C'}\tilde\theta(v;C')\theta(t;C/C')&=\sum_\sigma\Big[\sum_{C'\subset C_\sigma{}^+}\tilde\theta(v;C')\theta(t;C/C_\sigma{}^+)\Big]\\&=\sum_\sigma\tilde\theta(v;C_\sigma{}^+)\theta(t;C/C_\sigma{}^+)\\&=\sum_\sigma\tilde\theta(v;C^+)\theta(t;C_\sigma)=\tilde\theta(v;C^+).\end{aligned}$$

For general C, we make a standard simplexial decomposition $C=\bigcup_\alpha C_\alpha$. Since $H_{n-1}(C_\alpha{}^+)\subset H_{n-1}(C^+)$, we can use Eq. (3.23) for every C_α. Hence by using Eqs. (3.6) and (3.8), we obtain Eq. (3.23) for the general case.

4. NECESSITY PROOF OF THEOREM 1

To prove (R1), we rewrite definition (2.28) in a form similar to Eq. (2.39). Namely, using the notation (2.41)–(2.43), we see that $C_{\alpha\nu}{}^X$ is sum of several C_P. Moreover, because of (W2), if $x\in\Delta_\alpha{}^X$, then the $w_P(x)$ are equal for various P, as far as C_P stays in one $C_\alpha{}^X$. Hence using Eq. (3.6), we obtain

$$r_i(x)=(-i)^n\Sigma_\alpha\theta(x;\Delta_\alpha{}^X)\Sigma_\nu\theta(x^0;C_{\alpha\nu}{}^X/C_i)w_{\alpha\nu}{}^T(x). \tag{4.1}$$

$\theta(x^0;C_{\alpha\nu}{}^X/C_i)$ is invariant, as long as $x\in\Delta_\alpha{}^X$ because its discontinuity occurs only at $s(kl)\cdot x^0=0$ with k,l such that $[s(kl)\cdot x]^2>0$; otherwise it stays constant. Since $w_{\alpha\nu}{}^T(x)$ is covariant because of (W1) for $w_P{}^T$, we have (R1).

(R2) is an obvious consequence of Eq. (3.21).

To prove (R3), we note that the difference $\tilde\theta(s;C_P/C_i)-\tilde\theta(s;C_P/C_j)$ for neighboring C_i and C_j is the boundary value of $R[v;C_P/t(I)]$ multiplied by $\pm2\pi i\delta(s\cdot t(I))$ [compare Eq. (3.16)]. Hence, because of Eq. (3.13), only terms with those P for which $\pm C[t(I)]\in F_1(C_P)$ survive and, because of the presence of the above δ function and (W3), $w_P{}^T$ vanishes, if $\pm q\cdot t(I)$ is one of its intermediate momentum. (Note that $[q(I)]^2<m^2$.) Since $\pm C[t(I)]\in F_1(C_P)$ implies that $\pm q(I)$ is an intermediate momentum of $w_P{}^T$, we have (R3).

To prove (R4), we first note that, since $I(P,k)$, $k=1\cdots n$ is totally ordered by set inclusion, if $\sigma I\bigcap\sigma' I'\neq$ empty then $\pm q(I)$ and $\pm q(I')$ cannot be intermediate state for one $w_P{}^T$ simultaneously. Thus by lemma 4

$$\begin{aligned}&\theta(x;C_P/C_{+\sigma'})=\theta(x;C_P/C_{-\sigma'})\quad\text{if }\pm C[t(I)]\notin F_1(C_P)\\&\theta(x;C_P/C_{\sigma+})=\theta(x;C_P/C_{\sigma-})\quad\text{if }\pm C[t(I')]\notin F_1(C_P).\end{aligned}$$

Since one of these equalities is true for each C_P, we have (R4).

5. SUFFICIENCY PROOF OF THEOREM 1

First let us show that if r_i is obtained from $w_P{}^T$ as in Eq. (2.28), then we obtain the $w_P{}^T$ by Eq. (2.29). Namely, we define

$$\tilde w^T(q/t)=(i)^n\sum_i\theta(q^0;C_i/t)\tilde r_i(q). \tag{5.1}$$

Then by substituting Eq. (2.28) into Eq. (5.1), we have

$$\tilde w^T(q/t)=\sum_P\int e^{iq\cdot x}w_P{}^T(x)[\Sigma_i\theta(q^0;C_i/t)\theta(x^0;C_P/C_i)dx].$$

By Eq. (3.23) the summation within brackets is equal to $\theta(q^0;C_P{}^+/t)$. [Note that $\{C_i\}=\Gamma(H_{n-1}{}^R)$ and $H_{n-1}{}^R=\bigcup_P H_{n-1}(C_P{}^+)\supset H_{n-1}(C_P{}^+)$.] We now have

$$\tilde w^T(q/t)=\sum_P\theta(q^0;C_P{}^+/t)\tilde w_P{}^T(q).$$

By (W2) $\tilde w_P{}^T(q)=0$, if q^0 is not in the interior of $C_P{}^+$. If q^0 belongs to the interior of $C_P{}^+$ and $t\notin C_P$, then by Eqs. (3.22) $\theta(q^0;C_P{}^+/t)=0$. If q^0 is in the interior of $C_P{}^+$ and $t\in C_P$, then $\theta(q^0;C_P{}^+/t)=\theta(q^0;C_P{}^+)=1$. Thus we have

$$\tilde w^T(q/t)=\tilde w_P{}^T(q)\quad\text{if } t\in C_P. \tag{5.2}$$

We now assume (R1)–(R4) for $r_i(x)$ and define

$$w^T(u;\mathbf{x})=(-2\pi i)^{-n}\int dx^0\Sigma_i\tilde\theta(u-x^0;C_i)r_i(x) \tag{5.3}$$

$$w^T(x/t)=\lim_{\lambda\to+0}w^T(x^0+i\lambda t;\mathbf{x}). \tag{5.4}$$

We denote $\bigcup_i H_1(C_i)^\perp=\bigcup_i H_{n-1}(C_i{}^+)$ by $H_{n-1}{}^{RW}$ and $\bigcup_P H_{n-1}(C_P)$ by $H_{n-1}{}^W$. We easily see that $H_{n-1}{}^W\subset H_{n-1}{}^{RW}$ and in fact $H_{n-1}{}^{RW}$ is much larger set than $H_{n-1}{}^W$ in general.

If we denote the cones in $\Gamma(H_{n-1}{}^{RW})$ by $C_{P\gamma}$, where $C_P=\bigcup_\gamma C_{P\gamma}$ and the $w^T(x/t)$ with t in the interior of $C_{P\gamma}$ by $w_{P\gamma}{}^T(x)$, then by lemma 4, $W_{P\gamma}{}^T(x)$ is independent of the choice of t in $C_{P\gamma}$. However, it depends on γ in general.

By lemma 1, $w^T(u,\mathbf{x})$ has singularities for $\operatorname{Im}u\in h\subset H_{n-1}{}^{RW}$ in general. Hence, in order to be able to define $w_P{}^T$ from $w^T(u,\mathbf{x})$, we have to show that the jump across the cut on $\operatorname{Im}u\in h$ for $w^T(u,\mathbf{x})$ vanishes, if $h\in H_{n-1}{}^{RW}$ and $h\notin H_{n-1}{}^W$. [The $w^T(u,\mathbf{x})$ constructed

from r_i of the form (2.28) is regular there.] This follows from (R4) in the following way.

By Eq. (3.16), what we have to show is

$$\sum_i \int R(u-x^0; C_i/s)\delta[s\cdot(u-x^0)]r_i(x)dx^0=0, \quad (5.5)$$

for $\mathrm{Im}\, s\cdot u=0$, $h(s)^1\in H_{n-1}{}^{RW}$ and $h(s)^1\notin H_{n-1}{}^{W}$. This is equivalent to

$$\sum_i \prod_{m=1}^{n} \epsilon[(C_i)_{m-1}; f_m]r_i(x)=0 \quad (5.6)$$

for all $s_2\cdots s_n$, where $(C_i)_m=C_i+h_m$, $f_m=C(s_m)+h_{m-1}$, and $h_m=h(s,s_2\cdots s_m)$. The necessity follows from Eq. (3.14). For the sufficiency proof, we expand the rational function $R(u; C_i/s)$ into partial fractions first with respect to u_1 (the first component of u). Each expansion coefficient is the residue of R at the pole of that partial fraction and is a rational function of u given by some $R(u; C/s,s_2)$. On repeating this process, we arrive at a formula of the type[21]

$$R(u; C_i/s)=\text{const}\,\Sigma \prod_{m=1}^{n} \epsilon[(C_i)_{m-1}; f_m]R(u)$$

where f_m and $(C_i)_{m-1}$ are defined as in lemma 3, the summation is over $s_2\cdots s_n$ and $R(u)$ is a rational function of u depending on s, $s_2\cdots s_n$. By substituting this into Eq. (5.5) and using Eq. (5.6), we see the sufficiency of Eq. (5.5).

Next we prove Eq. (5.6) from (R4). In Eq. (5.6), if $h_m\notin H_m(C_i)$, for all i and for one fixed m, then all ϵ vanishes and the equation is satisfied. Hence we now assume that $h_m\in H_m(C_i)$ for some i, namely, that h_m is a m-dimensional intersection of planes $h(I_1)\cdots h(I_{n-m})$.

We first show that there is one and only one C_i for a given σ_m, $m=1\cdots n$ such that

$$\epsilon[(C_i)_{m-1}; f_m]=\sigma_m, \quad (5.7)$$

where $\sigma_m=\pm 1$. If this is true, then denoting the corresponding r_i by r_σ, we can rewrite Eq. (5.6) as

$$\sum_{\sigma_1\cdots\sigma_n} \sigma_1\cdots\sigma_n r_\sigma=0. \quad (5.8)$$

To prove this statement, we note that each h_m is divided by planes $h(I)$ not containing h_m into several (closed) convex polyhedral cones, say $C_\alpha{}^{(m)}$. For each $C_\alpha{}^{(m)}$, there is at least one cone C_i for which $C_i\cap h_m=C_\alpha{}^{(m)}$. Furthermore, each h_m is divided by h_{m-1} into two sides: $h_m=h_m{}^+\cup h_{m-1}\cup h_m{}^-$, where $\pm s_m\in h_m{}^\pm$. For each $C_\alpha{}^{(m-1)}$ there are just two $C_\beta{}^{(m)}$ containing $C_\alpha{}^{(m-1)}$ (in its boundary), one on each side of h_{m-1}. Hence by induction we obtain Eq. (5.8). (Note that $C_\alpha{}^{(n)}$ coincides with C_i.) We also see that $C_\alpha{}^{(m)}$ can be characterized by the value σ_k, $k\leqslant m$. Hence we use the notation $C^{(m)}(\sigma_1\cdots\sigma_m)$.

Next let us investigate h_m more closely. If I_a and I_b are proper nonempty subsets of $I(n+1)$, and if $\sigma_a I_a\neq\sigma_b I_b$ for σ_a, $\sigma_b=\pm$, then there are five mutually exclusive possibilities: (α1) $I_a\cap I_b=$ empty, (α2) $I_a\subset I_b$, (α3) $I_a\supset I_b$, (α4) $I_a\cup I_b=I(n+1)$, or (β) $\sigma_a I_a\cap\sigma_b I_b$ =nonempty for σ_a, $\sigma_b=\pm$. We now prove that there exists integers k and λ ($\lambda<k<n$) and the set $\{I_\nu{}^{(m)}; \nu\leqslant m\leqslant k\}$ satisfying the conditions:

(A1) $$h_{n-m}=\bigcap_{\nu=1}^{m} h(I_\nu{}^{(m)}),$$

(A2) $I_\nu{}^{(m)}$ is a partial sum of $I_\mu{}^{(m')}$, $\mu=\nu\cdots m'$ where $m<m'$, (A3) $I_a=I_\mu{}^{(m)}$ and $I_b=I_\nu{}^{(m)}$ satisfy (α1) for $u,\nu<k$ and (α1), or (β) for $m=\nu=k$. In the latter case, (β) holds for $\mu=\lambda$.

Suppose $I_\nu{}^{(m)}$ has been defined for $m<M$ satisfying (A1), (A2), and the condition (A3′): $I_\mu{}^{(m)}$ and $I_\nu{}^{(m)}$ fulfil (β). Then we will construct $I_\nu{}^{(M)}$ which satisfy (A1), (A2), and either (A3) or (A3′). If this can be done, then by induction there is some $M=k$ for which (A3) is true for the first time or else we find mutually disjoint $I_\nu{}^{(n-1)}$ such that $h_1=\bigcap_\nu h[I_\nu{}^{(n-1)}]$. This latter possibility contradicts $h_1{}^1\notin H_{n-1}{}^W$. To construct $I_\nu{}^{(M)}$, let $h_{n-M}=h_{n-M+1}\cap h(I)$. If $I\supset I_\mu{}^{(M-1)}$, we replace I by $I'=I-I_\mu{}^{(M-1)}$. After doing this replacement for each μ, I', and $I_\mu{}^{(M-1)}$ never satisfy (α3) nor (α4). [If $M=2$, (α4) may happen, but then we replace I by $-I$ without harming other conditions.] Now if $I'\subset I_\mu{}^{(M-1)}$ (which happens only for one μ), we define $I_\nu{}^{(M)}=I_\nu{}^{(M-1)}$ for $\nu\neq\mu$, $I_\mu{}^{(M)}=I_\mu{}^{(M-1)}-I'$, and $I_M{}^{(M)}=I'$ and they will satisfy (A1), (A2), and (A3′). Otherwise, we define $I_\mu{}^{(M)}=I_\mu{}^{(M-1)}$ and $I_M{}^{(M)}=I'$, and they will satisfy (A1), (A2), and (A3) or (A3′).

We now claim that, for the smallest l satisfying $I_l{}^{(l)}\supset I_\lambda{}^{(k)}$,

$$\sum_{\sigma_{n-k+1},\,\sigma_{n-l+1}} \sigma_{n-k+1}\sigma_{n-l+1}r_\sigma=0. \quad (5.9)$$

To prove this, we consider an inner point P of $C^{(n-k)}{}_{(\sigma_1\cdots\sigma_{n-k})}$ in h_{n-k}. In the neighborhood of P, there are no planes $h(I)$ except those containing h_{n-k}. We define the point

$$P(\epsilon_{n-k+1}\cdots\epsilon_m)=P(\epsilon_{n-k+1}\cdots\epsilon_{m-1})+\epsilon_m s_m', \quad (5.10)$$

where $s_m'=s_m$ except s'_{n-l+1} is chosen to satisfy $s'_{n-l+1}\cdot t(I_\mu{}^{(k)})=0$ for $\mu\neq\lambda$ and $s'_{n-l+1}\in h_{n-l+1}{}^+$. Obviously, $P(\epsilon_{n-k+1}\cdots\epsilon_m)\in h_m$. If we choose ϵ_m successively smaller enough, and if (sign ϵ_m)$=\sigma_m$, then $P(\cdots\epsilon_m)$ will be in the relative interior of $C^{(m)}{}_{(\sigma_1\cdots\sigma_m)}$. We now fix ϵ_m so that the point

$$P(\rho,\rho')=P(\rho\epsilon_{n-k+1}\cdots\rho'\epsilon_{n-l+1}\cdots\epsilon_n) \quad (5.11)$$

is in the interior of $C^{(n)}{}_{(\sigma_1\cdots\sigma_n)}$ for $\rho=\sigma_{n-k+1}$, and $\rho'=\sigma_{n-l+1}$. We also define $C(\rho\rho')=C_i$, $r(\rho\rho')=r_i$, if

[21] We are only interested in the coefficients.

$P(\rho\rho')$ is in the interior of C_i. We now prove that $r_{\rho+}-r_{\rho-}$ is constant in ρ. This will prove Eq. (5.9).

For this purpose, we consider the segment

$$L_+=\{P(\rho,+1);\,|\rho|\leqslant 1\} \text{ and } L_-=\{P(\rho,-1);\,|\rho|\leqslant 1\}$$

and consider the question: Where do L_+ and L_- meet the boundaries of C_i? Since L_+ and L_- are parallel to $s_{n-k+1}\in h_{n-k+1}$, they will never meet planes containing h_{n-k+1}, namely, planes $h(I)$ where I is any partial sum of $I_\mu^{(k-1)}$. On the other hand, if the ϵ are sufficiently small, $L_\pm$ are near P and will never meet with planes not containing h_{n-k}. Thus, the only planes $h(I)$ which $L_\pm$ meets are for $I=I_k^{(k)}+\Sigma I_\mu^{(k)}$,[22] where the summation is any partial sum of $I_\mu^{(k)}$ such that $I_k^{(k)}$ and $I_\mu^{(k)}$ have the property (α1). L_+ and L_- may meet more than one planes $h(I)$ at one time. In such a case we change the choice of ϵ slightly and then $I_\mu^{(k)}$ will meet only one plane at a time. Since $s_{n-l+1}\cdot t(I_\mu^{(k)})=0$ for $\mu\neq\lambda$, and $\mu=\lambda$ does not appear in the summation in the definition of I, L_+ and L_- meet $h(I)$ at the same time.

For each fixed I, we fix ρ^+ and ρ^- such that $P(\rho^\sigma,\sigma')$ is on the same side of $h(I)$ as $P(\sigma,\sigma')$ and sufficiently near to $h(I)$. We now prove

$$r(\rho^+,+1)-r(\rho^+,-1)=r(\rho^-,+1)-r(\rho^-,-1),$$

by proving that $r(\rho^+,\rho')-r(\rho^-,\rho')$ is constant in $\rho'(|\rho'|\leqslant 1)$.

Let the segment $\{P(\rho^\sigma,\rho');\,|\rho'|\leqslant 1\}$ be L_σ'. We investigate planes $h(I')$ which L_σ' meets. Since the L_σ' are near P, $h(I')$ should contain h_{n-k}. Since L_σ' are parallel to s_{n-l+1}, and since $s_{n-l+1}\cdot t(I_\mu^{(k)})=0$ for $\mu\neq\lambda$, I cannot be a partial sum of $I_\mu^{(k)}$, $\mu\neq\lambda$. Hence $I'=I_\lambda^{(k)}+\Sigma I_\mu^{(k)}$, where summation is any partial sum of $I_\mu^{(k)}$ with $\mu\neq\lambda,k$. Suppose $P(\rho^\sigma,\rho'^\sigma)$ is sufficiently near to $h(I')$ and on the same side of $h(I')$ as $P(\rho^\sigma,\sigma')$. Then what we would like to prove is

$$r(\rho^+,\rho'^+)-r(\rho^-,\rho'^+)=r(\rho^+,\rho'^-)-r(\rho^-,\rho'^-).$$

Because I and I' satisfies (β), this is nothing but (R4). Thus we have succeeded in proving that $w^T(u;\mathbf{x})$ has no cut across the plane $\operatorname{Im} s\cdot u=0$, unless $h(s)^\perp\in H_{n-1}^W$.

We now prove the properties (W1)–(W3) for w^T. First (W1) becomes obvious if we write $\tilde{w}_P{}^T(q)$ as

$$\tilde{w}_P{}^T(q)=i^n\Sigma_\beta\theta[q;\Delta_\beta{}^Q(m)]\times[\Sigma_\nu\theta(q^0;C_{\beta\nu}{}^0/t)\tilde{r}_{\beta\nu}(q)]\,t\in C_P, \quad (5.12)$$

where notations are as in Eq. (2.40) and the proof is similar to that of Eq. (4.1).

To prove (W2) or equivalently (W2′), we calculate by lemma 3 the jump of $w^T(u;\mathbf{x})$ across the cut $\operatorname{Im} s\cdot u=0$,

$$i^n\sum_i\int R[u-x^0;C_i/s(\mu\nu)]\delta[s(\mu\nu)\cdot(u-x^0)]\times r_i(x)dx^0,\quad \operatorname{Im} s(\mu\nu)\cdot u=0. \quad (5.13)$$

If $\pm s(\mu\nu)$ is not a 1 facet of C_i, then R vanishes. If $\pm s(\mu\nu)$ is a 1 facet of C_i, and if $[\operatorname{Re} s(\mu\nu)\cdot x]^2<0$, $r_i(x)$ vanishes because of (R2′). Thus Eq. (5.13) vanishes, if $[\operatorname{Re} s(\mu\nu)\cdot z]^2<0$, which proves (W2). ($z^0=u$, $\mathbf{z}=\mathbf{x}$.)

To prove (W3), we first note that if $q^0\notin C_P{}^+$, then $q^0\notin C_{P\gamma}{}^+$ for at least one γ, and, therefore, $\tilde{w}_P{}^T(q)=\tilde{w}_{P\gamma}{}^T(q)$ vanishes because each $\theta(q^0;C_i/C_{P\gamma})$ vanishes as a result of Eq. (3.21). Suppose $q^0\in C_P{}^+$ and $(q\cdot t(I))^2<m^2$ for at least one $I\in\mathcal{I}_P$. We will prove $\tilde{w}_P{}^T(q)=0$ for this case by using the following lemma.

Lemma 7. If $[q\cdot t(I)]^2<m^2$ for one $I\in\mathcal{I}_P$ and $\Delta_\beta{}^Q(m)$ contains q, then each cone $C_{\beta\nu}{}^Q$ contains points outside of $C_P{}^+$.

If this lemma is true, then for any point $q^0\in C_P{}^+$ there is a point $q^{0\prime}$ outside the cone $C_P{}^+$ which can be connected with q^0 by a continuous line without crossing boundary planes of any $C_{\beta\nu}{}^Q$. For such a $q^{0\prime}$,

$$\theta(q^0;C_{\beta\nu}{}^Q/C_{P\gamma})=\theta(q^{0\prime};C_{\beta\nu}{}^Q/C_{P\gamma})$$

by lemma 4. Since $\theta(q^{0\prime};C_{\beta\nu}{}^Q/C_{P\gamma})$ is a sum of $\theta(q^{0\prime};C_i/C_{P\gamma})$ by Eq. (3.6), and the latter vanishes, we have $\tilde{w}_P{}^T(q)=0$.

To prove lemma 7, it suffices to prove that if $q(I)^2<m^2$ for at least one $I\in\mathcal{I}_P$ and a polyhedral convex cone

$$C=\bigcap_{I'\in\mathcal{I}}C[t(I')]^+$$

is contained in $C_P{}^+$, then there is at least one $I'\in\mathcal{I}$ for which $q(I')^2<m^2$. To prove this, we note that $C\subset C_P{}^+$ implies (lemma C1) that

$$\lambda(I)t(I)=\sum_{I'\in\mathcal{I}}\lambda(I,I')t(I')\quad\text{for } I\in\mathcal{I}_P, \quad (5.14)$$

where $\lambda(I)$ and $\lambda(I,I')$ are positive integers. By comparing any fixed component on both sides of Eq. (5.15), we easily see

$$\lambda(I)\leqslant\Sigma_{I'}\lambda(I,I'). \quad (5.15)$$

If $q(I')^2\geqslant m^2$ for all $I'\in\mathcal{I}$, and if $q^0\in C$, then each $q(I')$ is positive timelike and we have[23]

$$[q(I)^2]^{\frac{1}{2}}\geqslant\Sigma_{I'}\lambda(I)^{-1}\lambda(I,I')[q(I')^2]^{\frac{1}{2}}\geqslant m^2,$$

which contradicts with the assumption. This completes the proof of lemma 7.

Finally we show that $w_P{}^T(x)$ satisfies Eq. (2.28). Since $w_P{}^T(x)=w_{P\gamma}{}^T(x)$ for any γ, we obtain, because of Eq. (3.6),

$$\sum_P\theta(x^0;C_P/C_i)w_P{}^T(x)=\sum_{P\gamma}\theta(x^0;C_{P\gamma}/C_i)w_{P\gamma}{}^T(x).$$

[22] Since $h(I)$ should contain h_{n-k}, $q(I)=0$ should be derived from $q[I_\mu^{(k)}]=0$, $\mu=1\cdots k$. (Compare lemma C1.) One can easily find that I should contain the whole or no part of $I_\mu^{(k)}$ for each $\mu\neq k$, and I cannot contain $I_k^{(k)}$ and $I_\mu^{(k)}$ at the same time if they fulfill (β). Furthermore, since $h(I)\pitchfork h_{n-k+1}$ and since $I_\mu^{(k)}=I_\mu^{(k-1)}$ for this case, I should contain $I_k^{(k)}$. Thus we have this result.

[23] If a_i is positive timelike $[(\Sigma a_i)^2]^{\frac{1}{2}}\geqslant\Sigma(a_i^2)^{\frac{1}{2}}$. This is easily seen in the rest system of Σa_i.

Substituting the definition of $w_P{}^T(x)$ into this equation and using Eq. (3.23), we obtain

$$\sum_P \theta(x^0; C_P/C_i) w_P{}^T(x) = \sum_{i'} \theta(x^0; C_{i'}{}^+/C_i) r_{i'}(x).$$

By using (R2) and Eq. (3.22), the term with $i' \neq i$ vanishes. By $\theta(x^0; C_i{}^+/C_i) = \theta(x^0; C_i{}^+)$, the remaining term is identical with $r_i(x)$,

6. PROOF OF THEOREM 2

The Fourier transform of $r_i(x)$,

$$\tilde{r}_i(\zeta) = \int e^{i(\zeta,x)} r_i(x) dx, \tag{6.1}$$

is analytic for $\zeta \in T(V_i{}^Q)$ because of (R2). Conversely, if $\tilde{r}_i(\zeta)$ is analytic in $T(V_i{}^Q)$ and satisfy certain boundedness condition, then its boundary value $\tilde{r}_i(q)$ has the property (R2). Since $\tilde{r}_i(\zeta)$ is covariant, because of (R1), it is analytic in the extended tube $T'(V_i{}^Q)$ by the theorem of Hall and Wightman.[24]

We shall now prove from the property (R3), that

$$\lim_{\epsilon \to +0} \tilde{r}_i(\zeta + i\epsilon q) = \lim_{\epsilon \to +0} \tilde{r}_j(\zeta - i\epsilon q), \tag{6.2}$$

where C_i and C_j are neighboring across the plane $h(I)$,

$$\zeta \in \Sigma(ij,m) \tag{6.3}$$

$q \in Q, [q \cdot t(I)]^2 > 0$ and $q^0 \cdot t(I) > 0$. If this is proved, then by the edge of wedge theorem,[25] $\tilde{r}_i$ and $\tilde{r}_j$ are analytic at $\Sigma(ij,m)$[26] and identical with each other, and, therefore, theorem 2 is proved.

To prove Eq. (6.2), we denote the boundary values in Eq. (6.2) by $\tilde{r}_i(\zeta)$ and $\tilde{r}_j(\zeta)$. By taking the Fourier transform of Eqs. (3.16) and (3.13), we obtain

$$\theta(x^0; C/C_i) - \theta(x^0; C/C_j) = \sigma \epsilon\{C; C[t(I)]\} \theta_1[(x^0)_I; C_I/C_{ij}], \tag{6.4}$$

where $C_I = C + h[t(I)]$ (as a set in $T \bmod h[t(I)]$), $C_{ij} = C_i \cap C_j (\subset h(I) = h[t(I)]^+)$, $(x^0)_I$ is x^0 taken $\bmod h[t(I)]$. θ_1 is as described in lemma 3, and σ is defined by $C[t(I)]^\sigma \supset C_i$. By using the addition theorem of Eq. (3.6) for the left-hand side of Eq. (6.4) we easily see

$$\epsilon(\bigcup C_P; C[t(I)]) \theta_1[(x^0)_I; (\bigcup C_P)_I/C_{ij}] = \Sigma \epsilon(C_P; C[t(I)]) \theta_1[(x^0)_I; (C_P)_I/C_{ij}], \tag{6.5}$$

where $\bigcup C_P$ is any partial sum of C_P and is assumed to be a polyhedral convex cone.

On using the integral representation Eq. (6.1) for $\tilde{r}_i(\zeta)$ and $\tilde{r}_j(\zeta)$ with $\zeta \in \Sigma(ij,m)$, we obtain by Eq. (6.4)

$$\tilde{r}_i(\zeta) - \tilde{r}_j(\zeta) = \int e^{i(\zeta,x)} dx \sum_{\alpha\nu} \epsilon\{C_{\alpha\nu}; C[t(I)]\} \times \theta_1[(x^0)_I; (C_{\alpha\nu})_I/C_{ij}] \theta(x; \Delta_\alpha{}^X) w_{\alpha\nu}{}^T(x). \tag{6.6}$$

Since $C_{\alpha\nu}$ is a partial sum of C_P, we can rewrite Eq. (6.6) using Eq. (6.5) as

$$\tilde{r}_i(\zeta) - \tilde{r}_j(\zeta) = \int e^{i(\zeta,x)} dx \sum_P \theta_1[(x^0)_I; (C_P)_I/C_{ij}] \times \epsilon\{C_P; C[t(I)]\} w_P{}^T(x). \tag{6.7}$$

We now introduce a basis $t(I), t_2 \cdots t_n$ in T and make the transformation of variables $x \to y$, through

$$x = t(I) \otimes y_1 + \sum_{i=2}^n t_i \otimes y_i.$$

(y_1 and y_i are Minkowski vectors.) Then θ_1 in Eq. (6.7) is independent of y_1 and if $t(I) \in F_1(C_P)$, the Fourier transform of $w_P{}^T(x)$ in y_1 with fixed y_i, $i \geqslant 2$,

$$w_P{}^T(p; y_2 \cdots y_n) = \int e^{i(p,y_1)} w_P{}^T(y_1; y_2 \cdots y_n) dy_1 = (2\pi)^{n-1} \int \exp[-\sum_{i=2}^n (q, y_i) \cdot t_i] \times \delta[p - q \cdot t(I)] \tilde{w}_P{}^T(q) dq$$

vanishes for $p^2 < m^2$ because of (W3). On the other hand, if $t(I) \notin F_1(C_P)$, then $\epsilon\{C_P; C[t(I)]\}$ vanishes by the definition of Eq. (3.15) and hence we have $\tilde{r}_i(\zeta) = \tilde{r}_j(\zeta)$ for $\zeta \in \Sigma(ij,m)$.

We note that Eq. (2.39) is obtained from Eq. (6.1) because if $\zeta \in T(V_i{}^Q)$ then Im $\zeta^0 \in C_i$. Unlike Eq. (6.1), Eq. (2.39) holds in all $T(V_i{}^Q)$.

Finally we add the proof of Eq. (2.45). By definition

$$\tau(x) = \Sigma_P \theta(x^0; C_P) w_P(x).$$

If we use $\theta(x^0; C_P) = \theta(x^0; C_P/C_P{}^+)$, we obtain

$$\tilde{\tau}(q) = (2\pi)^{-n} \Sigma_P \int \tilde{\theta}(s - q^0; C_P/C_P{}^+) \tilde{w}_P(q) dq^0.$$

We now assume that $s (= q^0) \in C_i$. If $C_P{}^+ \supset C_i$, the replacement of $C_P{}^+$ by C_i can be done trivially. On the other hand, if $C_P{}^+ \not\supset C_i$, then $s \cdot t(I) < 0$ for at least one $I \in \mathcal{I}_P$ and as a result of (W3), at $v = s$,

$$\int \tilde{\theta}(v - q^0; C_P) \tilde{w}_P(q) dq^0$$

will be analytic. Hence we can again replace $C_P{}^+$ by C_i. Thus we have the formula (2.45).

[24] D. Hall and A. Wightman, Kgl. Danske Videnskab. Selskab. Mat.-fys. Medd. **31**, No. 5 (1957).

[25] H. Bremmermann, R. Oehme, and J. G. Taylor, Phys. Rev. **109**, 2178 (1958); J. G. Taylor, Ann. Phys. **5**, 391 (1958); F. J. Dyson, Phys. Rev. **110**, 579 (1958); L. Garding and A. Beurling (to be published).

[26] Compare H. Epstein, "Generalization of the edge of wedge theorem" (preprint).

7. ADDITIONAL REMARKS

To make theorem 1 of Sec. 2 precise, one has to state the class of distributions to which $w_P{}^T$ and r_i belong.[16] We do not attempt to make a precise statement as to the class of distributions for which our proof holds, but we would like to make some remarks pertinent to this point.

The behavior for large value of space-time coordinate can be estimated by physical arguments and it is expected that $w_P{}^T$ decreases exponentially in spacelike directions and according to a power law in timelike directions. This behavior will be inherited by r_i. Hence the assumption that the multiplication of $\tilde{r}_i(q)$ by $\theta(q^0; C_i/t)$ is well defined is a reasonable one.

We have shown that $w_P{}^T$ and w_P yield the same r_i. We have also shown that $w_P{}^T$ can be obtained from r_i by an inversion formula. The reason why w_P cannot be obtained by the same inversion formula is the following. w_P will (in general) approach to nonzero values for large separation of its coordinates because of the vacuum intermediate state. Because of this, expressions like $\theta(q^0; C_i/t)\tilde{w}_P(q)$ have ambiguity and, especially, the formula (3.23) cannot be used when multiplied by $\tilde{w}(q)$.[16] Thus, if we substitute Eq. (2.28) with $w_P{}^T$ replaced by w_P into Eq. (5.1), we cannot change the order of summation over P and multiplication by $\theta(q^0; C_i/t)$, and, hence we do not get $w_P(x)$. On the other hand, if $w_P{}^T$ behaves as we conjectured, then we will obtain $w_P{}^T$ by Eq. (2.29). This is one of the reasons for using $w_P{}^T$ instead of w_P.

We do not know much about the behavior of $\tilde{w}_P(q)$ for large energy momentum. If $\tilde{w}_P(q)$ does not decrease for large q, we have to use the subtraction method. It seems to be a nontrivial problem to extend our results to this case.

APPENDIX A. CASE OF MORE COMPLEX SPECTRUM CONDITIONS

We define $m(P,k)$ by the lowest upper bound of m such that

$$(\Psi_0, A_P(1)(x_{P(1)})\cdots A_{P(k)}(x_{P(k)})(P_m-P_0) \times A_{P(k+1)}(x_{P(k+1)})\cdots A_{P(n+1)}(x_{P(n+1)})\Psi_0) \quad \text{(A.1)}$$

vanishes identically where P_m is the projection into states with mass below m and P_0 is the projection into the vacuum Ψ_0. We first prove

$$m(P,k)=m(P',k) \quad \text{if } I(P,k)=I(P',k). \quad \text{(A.2)}$$

Suppose (A.1) vanishes identically for P and $m<m(P,k)$. Then (A.1) vanishes for P' and $m<m(P,k)$, if the points $x_{P(1)}\cdots x_{P(k)}$ and $x_{P(k+1)}\cdots x_{P(n+1)}$ are spacelike to each other within each group. We now note that (A.1) for P' as a distribution in the difference variables $\xi_j=x_{P'(j)}-x_{P'(j+1)}$ is a boundary value of a function which is analytic for Im $\xi_j\in V_-$. Hence[27] Eq. (A.1) for P' also vanishes identically for $m<m(P,k)$.

[27] By the theorem of Hall and Wightman, the analytic function in question is analytic in a Jost point [R. Jost, Helv. Phys. Acta 30, 409 (1957)], where we have proved that Eq. (A.1) vanishes. Hence it vanishes identically. We could use also edge of wedge theorem (instead of Jost points), taking 0 as the analytic function approaching the same boundary value from the other side.

Because of Eq. (A.2) we can define

$$m(I)=m(P,k) \quad \text{if } I=I(P,k). \quad \text{(A.3)}$$

We now assume the following:

Assumption A. $m(I)$ with fixed I is the same for all n such that $w_P(x)\neq 0$. In addition, for integers $\lambda_j>0$,

$$m[I(n+1)-I]=m(I), \quad \text{(A.4)}$$

$$\Sigma_j\lambda_j m(I_j)\geqslant m(I) \quad \text{if } t(I)=\Sigma_j\lambda_j t(I_j). \quad \text{(A.5)}$$

Equation (A.4) is obviously true for Hermitian fields. The idea behind Eq. (A.5) is the following. The state

$$\Psi=\prod_j\prod_{\nu=1}^{\lambda_j}\prod_{i\in I_j} A_i(x_{j\nu i})^*\Psi_0$$

will have the same quantum numbers (which is associated with fields, additive, and zero for vacuum state)[28] as the states

$$\Psi'=\prod_{i\in I} A_i(x_i)^*\Psi_0 \quad \text{and} \quad \Psi''=\prod_{i\in -I} A_i(x_i)\Psi_0.$$

By definition of $m(I_j)$ there is a state Ψ_j with mass around $m(I_j)$ such that

$$(\Psi_j, \prod_{i\in I_j} A_i(x_i)^*\Psi_0)\neq 0.$$

On assuming asymptotic conditions, we write Ψ_j in the form $\Psi_j=F_j(A^{in})\Psi_0$. Then the state

$$\Psi'''=\Pi_j[F_j(A^{in})]^{\lambda_j}\Psi_0$$

will have the same quantum number as Ψ and the mass around $\Sigma_j\lambda_j m(I_j)$. Then, assuming no accidental cancellation, (Ψ,Ψ') and (Ψ,Ψ'') will not vanish identically, and we see that Eq. (A.5) is a reasonable assumption.

We note that for $n=2$, Eq. (A.5) takes the form

$$m_i+m_j\geqslant m_k\geqslant |m_i-m_j|, \quad \text{(A.6)}$$

where (ijk) is any permutation of (123).[29]

As will be proved in Appendix B, $w_P{}^T$ will satisfy (under the assumption B)

(W3′) $\tilde{w}_P{}^T(q)=0$ unless $q\cdot t(I)\in V_+$ and $[q\cdot t(I)]^2\geqslant m(I)^2$ for all $I\in\mathcal{I}_P$.

By the same proof as for (R3), we obtain

(R3′) $\tilde{r}_i(q)=\tilde{r}_j(q)$ if C_i and C_j are neighboring across $h(I)$ and if $[q\cdot t(I)]^2<m(I)^2$.

[28] For multiplicative quantum numbers of the form $(-1)^n$, one can take n mod 2.

[29] We thank Professor A. S. Wightman for an illuminating explanation of the relevance of Eq. (A.6) for the sufficiency of the condition of the type (R3).

We also obtain the analyticity of $\tilde{r}(\zeta)$ at

$$\Sigma(ij;\{m(I)\})=\{\zeta\in Z';\ \mathrm{Im}\,\zeta\in S(ij),\ [\mathrm{Re}\,q\cdot t(I)]^2<m(I)^2\}. \quad (A.7)$$

The sufficiency of (R3′) for (W3′) will be established in the same way as in Sec. 5, if the following is true (compare lemma 7).

(M1) If $[q\cdot t(I)]^2<m(I)^2$ for at least one $I\in\mathcal{I}_P$ and $\Delta_\beta{}^Q(\{m(I)\})$ contains q, then each cone $C_{\beta\nu}{}^Q$ contains points outside of $C_P{}^+$, where

$$\Delta_\beta{}^Q(\{m(I)\})=\{q\in Q;\ \sigma_\beta(I)([q\cdot t(I)]^2-m(I)^2)>0\}. \quad (A.8)$$

This lemma follows from Eq. (A.5) in the same way as the proof of lemma 7, if the following statement is true:

(M2) The $\lambda(I)$ in Eq. (5.13) can be taken as 1. Namely, if

$$C=\bigcap_{I'\in\mathcal{I}} C[t(I')]^+ \quad \text{and} \quad C\subset C_P,$$

then

$$t(I)=\sum_{I'\in\mathcal{I}}\lambda(I,I')t(I') \quad \text{for } I\in\mathcal{I}_P, \quad (A.9)$$

where $\lambda(I,I')$ is an integer.

We have been unable to prove this for general n, but for $n\leqslant 4$ ($n=4$ corresponds to the five-point function) (M2) can be verified easily.

Summing up we have the following theorem:

Theorem A. If $w_P{}^T$ satisfies (W1), (W2), and (W3′), then r_i satisfies (R1), (R2), and (R3′), and (R4). The converse is true, if ($M1$) holds (which is the case for $n\leqslant 4$). $\tilde{r}(\zeta)$ is analytic in the union of extended tubes $T'(V_i{}^Q)$ and at the points of $\Sigma(ij;\{m(I)\})$.

APPENDIX B. TRUNCATED VACUUM EXPECTATION VALUES

First we prove a lemma which will be used in later discussion. Let $B(x_1\cdots x_n)$ and $C(y_1\cdots y_m)$ be products of fields $B_i(x_i)$ and $C_i(y_i)$, respectively. If the theory satisfies (2) in Sec. 2, $B(x_1\cdots x_n)$ and $C(y_1\cdots y_m)$ either commute or anticommute if all the x_i-y_j are spacelike.

Lemma B. If $B(x_1\cdots x_n)$ and $C(y_1\cdots y_m)$ anticommute for spacelike x_i-y_j, then the vacuum expectation value of either $B(x_1\cdots x_n)$ or $C(y_1\cdots y_m)$ vanishes identically.[30]

For the proof, by theorem 3 of our previous paper[31] which has been proved there under the assumption of (1), (3a), and (3b) [but not (2)] we have

$$\lim_{\lambda\to\infty}(\Psi_0,BU(\lambda a,1)C\Psi_0)=(\Psi_0,B\Psi_0)(\Psi_0,C\Psi_0),$$

$$\lim_{\lambda\to\infty}(\Psi_0,CU(-\lambda a,1)B\Psi_0)=(\Psi_0,B\Psi_0)(\Psi_0,C\Psi_0),$$

where $U(\lambda a,1)$ is the unitary operator for the translation by λa, and a is any spacelike vector. If B and C anticommute for spacelike x_i-y_j, then for sufficiently large λ

$$(\Psi_0,BU(\lambda a)C\Psi_0)=-(\Psi_0,CU(-\lambda a)B\Psi_0).$$

Hence we have

$$(\Psi_0,B\Psi_0)\cdot(\Psi_0,C\Psi_0)=0. \quad (B.1)$$

We now consider the truncated vacuum expectation values defined recursively by Eq. (2.3). We note that, although the definition of sign σ of each term in Eq. (2.3) refers to the order of the factors in that term, σ is actually independent of their order or else that term vanishes identically due to lemma B.

We define

$$w(i_1\cdots i_k)=(\Psi_0,A_{i_1}(x_{i_1})\cdots A_{i_k}(x_{i_k})\Psi_0)\sigma i_1\cdots(i_k), \quad (B.2)$$

$$[i_1\cdots i_k]_T=[A_{i_1}(x_i)\cdots A_{i_1k}(x_{i_k})]_T\sigma(i_1\cdots i_k). \quad (B.3)$$

$\sigma(i_1\cdots i_k)$ is the sign which one obtains if one commutes fields from the natural order to the order $i_1\cdots i_k$ for totally spacelike configuration of x_i. σ_P of Eq. (2.1) is $\sigma[P(1)\cdots P(n+1)]$.

The definition of Eq. (2.3) now becomes

$$w(i_1\cdots i_m)=[i_1\cdots i_m]_T+\Sigma_G\sigma_G[i_1\cdots]_T[i_k\cdots]_T\cdots, \quad (B.4)$$

where the order of the i in $[\ ]_T$ is as in w, and the summation is over all groupings G of $i_1\cdots i_m$. σ_G is

$$\sigma_G=\sigma\cdot\sigma(i_1\cdots i_m)\sigma(i_1\cdots)\sigma(i_k\cdots)\cdots=\sigma(i_1\cdots,i_k\cdots,\cdots)\sigma(i_1\cdots)\sigma(i_k\cdots)\cdots \quad (B.5)$$

In this form we see that σ_G depends only on the grouping and not on the order of $i_1\cdots i_m$ on w. Note that, by lemma B, $\sigma(i_1\cdots,i_k\cdots,\cdots)$ is independent of the order of the groups $(i_1\cdots)$, $(i_k\cdots)$, $\cdots$, unless that term vanishes identically.

The spectrum condition of Appendix A can be written as (W3″) $\tilde{w}(i_1\cdots i_m)=0$ unless $q(i_1\cdots i_k)\in V_+$,

$$q(i_1\cdots i_k)^2\geqslant m(i_1\cdots i_k)^2$$

for all $k\leqslant m$ or $q(i_1\cdots i_k)=0$ for some $k\leqslant m$,

$$\tilde{w}(i_1\cdots i_m)=\sigma(i_1\cdots i_m)\sigma(i_1\cdots i_k)\sigma(i_{k+1}\cdots i_m)\times\tilde{w}(i_1\cdots i_k)\tilde{w}(i_{k+1}\cdots i_m) \text{ if } q(i_1\cdots i_k)=0.$$

The notations are

$$\tilde{w}(i_1\cdots i_m)=\int \exp i[\Sigma(q_i,x_i)]\times w(i_1\cdots i_m)dx_{i_1}\cdots dx_{i_m}, \quad (B.6)$$

$$q(i_1\cdots i_m)=q_{i_1}+\cdots+q_{i_m}. \quad (B.7)$$

[30] We assume (1), (2), (3a), (3b), and (3c) for the theory. However, we do not make assumptions about the connection between commutation relation among different fields and the type of fields. See H. Araki, J. Math. Phys. (to be published).

[31] H. Araki, Ann. Phys. **11**, 260 (1960). Theorem 3 in that paper is expressed in terms of w^T. However, the properties used for w^T in the proof are the covariance and the existence of lowest positive mass in that intermediate state where $U(\lambda a,1)$ is inserted. $(\Psi_0,BU(\lambda a,1)C\Psi_0)-(\Psi_0,B\Psi_0)(\Psi_0,C\Psi_0)$ clearly has these properties.

Note that $\tilde{w}$ contains δ function, in contrast to our former definition of $\tilde{w}_P$.

We now strengthen assumption A a little.

Assumption B. If $t(I)=\Sigma_j\lambda_j t(I_j)$ for integers $\lambda_j>0$,

$$\begin{aligned} m(I) &\leqslant \Sigma_j\lambda_j m'(I_j) \text{ unless } m'(I_j)=0 \text{ for all } j, \\ &\leqslant \min_j m(I_j) \text{ if } m'(I_j)=0 \text{ for all } j, \end{aligned} \tag{B.8}$$

where

$$\begin{aligned} m'(\{i_1\cdots i_m\}) &= m(\{i_1\cdots i_m\}) && \text{if } w(i_1\cdots i_m)=0 \\ &= 0 && \text{otherwise.} \end{aligned} \tag{B.9}$$

The idea behind this assumption is the same as for assumption A.

We now prove the following theorem.

Theorem B. If $w(i_1\cdots i_m)$ satisfies conditions (W1), (W2), and (W3''), then $(i_1\cdots i_m)_T$ satisfies (W1), (W2), and (W3'). The converse is also true. (We make the assumption B.)

For the proof, the equivalence of (W1) for $w(i_1\cdots i_m)$ and $[i_1\cdots i_m]_T$ is obvious, because the defining equation has a unique solution in both directions. In addition, because (W2) is the requirement of symmetry in i and j, when x_i-x_j is spacelike, and because Eq. (B.4) is a completely symmetric definition, the equivalence of (W2) for $w(i_1\cdots i_m)$ and $[i_1\cdots i_m]_T$ is also obvious. (It is important here that σ_G is independent of the order of $i_1\cdots i_m$.)

We now prove the equivalence of (W3'') and (W3'). First suppose $q(i_1\cdots i_l)^2<m(i_1\cdots i_l)^2$. Then by assumption B, for any grouping of $i_1\cdots i_l$, either there is a group for which $q(i_k\cdots)^2<m'(i_k\cdots)^2$ or else $q(i_k\cdots)^2<m(i_k\cdots)^2$ for all groups. From this we easily see that (W3') implies (W3''). To prove the converse, we define

$$\begin{aligned} (i_1\cdots i_m)_0=\sigma(i_1\cdots i_m)(\Psi_0, &A_{i_1}(x_{i_1})(1-P_0)\cdots \\ &\times(1-P_0)A_{i_m}(x_{i_m})\Psi_0). \end{aligned} \tag{B.10}$$

In the same way as in our previous paper,[32] we can derive

$$[i_1\cdots i_n]_T=(i_1\cdots i_n)_0-\sum_{\text{con}}\sigma_G[i_1\cdots]_T\cdots, \tag{B.11}$$

where the summation is over all connected groupings.[33] We can now apply the previous argument to Eq. (B.11) and easily see that (W3'') implies (W3').

Finally we prove that r_i defined from $w_P{}^T$ and w_P are the same. We show that the term from the summation over G in Eq. (2.3) cancels out in Eq. (2.28). Consider one fixed grouping $(i_1\cdots i_k)$, $(j_1\cdots j_l)$, $\cdots$. We note that there are several w_P which contribute to the same term of the form $(x_{i_1}\cdots x_{i_k})_T(x_{j_1}\cdots x_{j_l})_T\cdots$. The union of the C_P for such P is the cone

$$C_G=\{t\in T;\ t_{i_1}\geqslant\cdots\geqslant t_{i_k},\quad t_{j_1}\geqslant\cdots\geqslant t_{j_l},\cdots\}.$$

This cone is obviously not pointed. Since σ_G is independent of P, we see from Eq. (3.7) that the contributions from various P cancels out.

[32] H. Araki[31]; see Eqs. (2.11)-(2.16).

[33] If each group of a grouping G occupies consecutive positions in $(i_1\cdots i_n)$, then G is called a division of $(i_1\cdots i_n)$. If a grouping is a subgrouping of a proper division, then it is called a disconnected grouping. Otherwise, a grouping is called a connected grouping. Thus for a connected grouping, numbers in one group are interlocked in $(i_1\cdots i_n)$, with those in another group.

APPENDIX C. CONVEX POLYHEDRAL CONES[34]

Consider a real n-dimensional vector space T and its dual S. A k-dimensional linear subspace is called k-plane. The linear subspace generated by a subset T_1 is denoted by $h(T_1)$. For example,

$$h(\{t_1,\cdots t_m\})=\{\sum_{i=1}^{m}\rho_i t_i;\ \rho_i \text{ real}\}.$$

The orthogonal compliment of h is denoted by $h^\perp$. (If $h\in T$, then $h^\perp\in S$. If H is a family of planes h, then $H^\perp$ means the family of planes $h^\perp$.) The convex polyhedral cone generated by $t_1,\cdots t_m$ is denoted by

$$C(t_1,\cdots t_m)=\{\sum_{i=1}^{m}\lambda_i t_i;\ \lambda_i\geqslant 0\}. \tag{C.1}$$

The positive polar C^+ and the negative polar C^- of a convex cone C is defined by

$$\begin{aligned} C^+&=\{s\in S;\ s\cdot t\geqslant 0,\ t\in C\}, \\ C^-&=\{s\in S;\ s\cdot t\leqslant 0,\ t\in C\}. \end{aligned} \tag{C.2}$$

The polars of a polyhedral convex cone in T are again polyhedral convex cones in S. The positive polar of the positive polar is the original cone. Note that

$$C(t_1\cdots t_m)^+=\{s\in S;\ s\cdot t_i\geqslant 0,\quad i=1,\cdots,m\}, \tag{C.3}$$

$$h(t_1\cdots t_m)=C(\pm t_1\cdots\pm t_m),\quad h^+=h^-=h^\perp. \tag{C.4}$$

We call $h(C)$ the dimensionality space of the cone C and its dimension the dimension of the cone C. A polyhedral convex cone C has nonempty interior, if, and only if, $\dim C=n$. The maximum linear subspace contained in C is called the linearity space of C and its dimension is called the linearity of C. [Notation: $L(C)$ and $\operatorname{lin} C$.] If $\operatorname{lin} C=0$, C is called pointed. C is pointed, if, and only if, there is a $(n-1)$-plane intersecting with the cone C only at the origin. We have the following relations:

$$h(C^+)=h(C^-)=L(C)^\perp,\quad L(C^+)=L(C^-)=h(C)^\perp, \tag{C.5}$$

$$\dim C+\operatorname{lin} C^+=\dim C^++\operatorname{lin} C=n. \tag{C.6}$$

By Eq. (C.6) C is pointed, if, and only if, C^+ has nonempty interior.

An extremum subset X of C is the set such that t_1, $t_2\in C$ and $\alpha t_1+\beta t_2\in X$ for some positive α and β with $\alpha+\beta=1$ necessarily imply t_1, $t_2\in X$. Any convex extremum subset of C is again a polyhedral convex cone

[34] Compare M. Gerstenhaber in *Activity Analysis of Production and Allocation*, edited by T. C. Koopmans (John Wiley & Sons, Inc., New York, 1951), Chap. 18.

and is called k-facet where k is its dimension. If $\dim C=n$, the $(n-1)$-facets of C form the boundary of C. If $\lin C=0$, the 1-facets of C generate C. If $k+1<\dim C$, a k-facet F is a k-facet of some $(k+1)$-facet G and the intersection of such G is F. If f^+ is a k-facet of C^+, f is called k-corner of C. One-facet is sometimes called extreme half-line and 1-corner is sometimes called supporting half-space. We denote the set of all k-facets of C by $F_k(C)$, the set of all $h(f)$ with f in $F_k(C)$ by $H_k(C)$, and the set of all k-corners by $F_k{}^+(C)$.

The sum $C+C'$ is the set of all sums $t+t'$ for $t\in C$ and $t'\in C'$. It is again a polyhedral convex cone. Note that $C(T_1\bigcup T_2)=C(T_1)+C(T_2)$ where T_i are subsets of T. The intersection $C\bigcap C'$ is also a polyhedral convex cone. The C's form a lattice with the operations $+$ and $\bigcap$. C^+'s form its dual. Namely,

$$\begin{aligned}(C_1\bigcap C_2)+C_3&=(C_1+C_3)\bigcap(C_2+C_3),\\ (C_1+C_2)\bigcap C_3&=(C_1\bigcap C_3)+(C_2\bigcap C_3),\end{aligned}\tag{C.7}$$

$$(C+C')^+=C^+\bigcap C'^+,\quad (C\bigcap C')^+=C^++C'^+.\tag{C.8}$$

[Note that C can be replaced by h because of Eq. (C.4).] The set of $-t$ for all $t\in C$ is denoted by $-C$.

If every element s of a set Σ is expressible as a positive linear combination $s=\Sigma\lambda(\nu)s(\nu)$ $(\lambda(\nu)\geqslant 0)$ of elements $s(\nu)$ of a subset Σ', then Σ' is called a positive basis of Σ. If every s in Σ is expressible as $s=\pm\Sigma\lambda(\nu)s(\nu)$ $(\lambda(\nu)\geqslant 0)$, then Σ' is called a c basis of Σ. A c basis of Σ which does not contain any sub-c basis is called c minimal. If $C(\Sigma)$ for a finite set Σ is pointed, Σ has a unique c minimal positive basis. If a finite set Σ is c minimal, $C(\Sigma)$ is pointed and $F_1[C(\Sigma)]$ consists of $C(s)$, $s\in\Sigma$.

We now state a lemma which is equivalent to the statement $(C^+)^+=C$.

Lemma C1. If $s\cdot t_1\geqslant 0,\ \cdots s\cdot t_m\geqslant 0$ imply $s\cdot t\geqslant 0$, then $t=\Sigma_i\lambda_i t_i$ with some nonnegative λ_i.

Given a family of $(n-1)$ planes, $H=\{h(s)^\perp; s\in\Sigma\}$. If $h(\Sigma)$ is the total space S, then the planes in H will divide the entire space T into several pointed polyhedral convex cones with nonempty interior. We denote the set of all these convex cones by $\Gamma(H)$. Let

$$\Sigma_0=\{\pm s; s\in\Sigma\}$$

and Σ_α be distinct c-minimal c basis of Σ_0. Then $\Gamma(H)=\{C(\Sigma_\alpha)^+\}$. If we denote the set of k-planes generated by a subset of Σ by $\Pi_k(\Sigma)$, then $H_k[C(\Sigma_\alpha)]\subset\Pi_k(\Sigma)$ and $H_k(C)\subset\Pi_{n-k}(\Sigma)^\perp$ for any $C\in\Gamma(H)$.

A cone $C(t_1\cdots t_n)$ with dimension n and linearity 0 is called a simplex cone. Its polar is also a simplex cone. If $s_i\cdot t_j=\delta_{ij}$, then $C(t_1\cdots t_n)^+=C(s_1\cdots s_n)$. Any polyhedral convex cone with dimension n can be decomposed into a union of almost disjoint simplex cones C_α.

$$C=\bigcup_\alpha C_\alpha,\ C_\alpha\text{: simplex, }\dim C_\alpha\bigcap C_\beta<n\text{ for }\alpha\neq\beta.\tag{C.9}$$

If $F_1(C_\alpha)\subset F_1(C)$ for all α, this decomposition is called a standard simplexial decomposition. We now prove the following lemma:

Lemma C2. If $\dim C=n$ and $\lin C=0$, C has a standard simplexial decomposition. Furthermore, for any given plane h not belonging to $H_{n-1}(C)$, there is a standard simplexical decomposition (C.9) for which $h\notin H_{n-1}(C_\alpha)$ for any α.

For the proof of the first half, take any 1-facet f_1 and consider all polyhedral convex cones C_α generated by f_1 and any $(n-1)$-facet $f_{n-1}{}^\alpha$ not containing f_1. We easily see that $C=\bigcup C_\alpha$, $\dim(C_\alpha\bigcap C_\beta)<n$ for $\alpha\neq\beta$, and $F_1(C_\alpha)\subset F_1(C)$. Hence by induction on the number of 1-facets, we get the first half. Moreover, we get the second half by always taking a 1 facet f_1 not containing the given plane h. Note that if $f_1\not\subset h$ and if there is only one facet not containing f_1, then any standard simplexial decomposition after that stage will have the property that $h\notin H_{n-1}(C_\alpha)$. Note also that if there is only one $(n-1)$-facet not containing f_1 for every 1 facet f_1, then the cone is simplex.

Note added in proof. The definition (2.30) of $V_i{}^Q$ should be replaced by

$$V_i{}^Q=C(V_+\otimes\text{interior of }C_i)$$

where Q is considered as a direct product of a Minkowski space and S. A similar definition can be given for $V_P{}^X$ but is equivalent to Eq. (2.33). The author is indebted to Dr. O. Steinmann for pointing out the unfitness of (2.30) for the proof of analyticity in $T(V_i{}^Q)$ in Sec. 6.

ACKNOWLEDGMENTS

The author would like to thank Professor A. S. Wightman and Dr. N. Burgoyne for discussions and Dr. D. Ruelle and Dr. O. Steinmann for correspondence. The author would also like to thank Professor R. Haag for critical reading of the manuscript.

Integral Representation for the Nonrelativistic Coulomb Green's Function*

EYVIND H. WICHMANN AND CHING-HUNG WOO
University of California, Berkeley, California

(Received August 26, 1960)

Although the radial Green's function for the Schrödinger equation in a Coulomb field can be obtained in the usual way in terms of the two linearly independent solutions to the radial equation for a particular angular momentum state, the sum over angular momentum states does not seem to have been carried out. In this note this sum is carried out and a "closed form" obtained in the form of a double integral. The result is believed to be useful for perturbation calculations where the "intermediate states" involve many angular momentum states.

THE nonrelativistic radial wave equation for a Coulomb potential, corresponding to angular momentum state l, is of the form

$$\frac{d^2}{dx^2}u(x)+\left[-1+\frac{2i\epsilon}{x}-\frac{l(l+1)}{x^2}\right]u(x)=0.$$

We will first consider solutions to the radial equation in this simple form, so that the formulas will look more transparent. The transformations which bring this to the more familiar form of the Schrödinger equation will be introduced later at the appropriate place [see Eq. (14)].

Consider the function $u_1(x)$, which is the radial solution corresponding to angular momentum l and which is finite at the origin,

$$u_1(x)=\frac{x^{l+1}e^{-x}}{\Gamma(2l+2)}\Gamma(l+1-i\epsilon)_1F_1(l+1-i\epsilon,\,2l+2,\,2x)$$
$$=\frac{x^{l+1}e^{-x}}{\Gamma(l+1+i\epsilon)}\int_0^1 dt(1-t)^{l+i\epsilon}t^{l-i\epsilon}e^{2xt} \qquad (1)$$
$$\mathrm{Re}(l+1-i\epsilon)>0;\quad \mathrm{Re}(l+1+i\epsilon)>0.$$

Trivially, this can be transformed into another integral which contains a spherical Bessel function of integral order:

$$u_1(x)=\frac{x(2i)^{-l}}{\Gamma(i\epsilon)}\int_0^1 dt(1-t)^{i\epsilon-1}t^{-i\epsilon}e^{-x(1-t)}j_l(ixt) \qquad (2)$$
$$\mathrm{Re}(l+1-i\epsilon)>0;\quad \mathrm{Re}(i\epsilon)>0.$$

On integrating by parts and using the fact that $j_l(ixt)$ is an entire function of its argument,

$$u_1(x)=\frac{x(2i)^{-l}}{\Gamma(1+i\epsilon)}\int_0^1 dt(1-t)^{i\epsilon}t^{-i\epsilon}\frac{\partial}{\partial t}[te^{-x(1-t)}j_l(ixt)] \qquad (3)$$
$$\mathrm{Re}(l+1-i\epsilon)>0;\quad \mathrm{Re}(i\epsilon)>-1.$$

Thus, one sees that $u_1(x)$ as given by Eq. (3) is an analytic function of ϵ in the complex ϵ plane in a region larger than that given in Eq. (2). All we will need is the fact that Eq. (3) will be an analytic function of ϵ inside the circle $|\epsilon|=1$ for every l.

An analogous result for the second solution $u_2(x)$ which is finite at infinity does not seem to be available in the literature, and will be derived in the following in some detail.

Consider $u_2(x)$ defined by

$$u_2(x)=\frac{x^{l+1}e^{x}}{\Gamma(l+1-i\epsilon)}\int_1^\infty ds(s-1)^{l-i\epsilon}s^{l+i\epsilon}e^{-2xs} \qquad (4)$$
$$\mathrm{Re}(x)>0;\quad \mathrm{Re}(l+1-i\epsilon)>0.$$

It will be shown that

$$u_2(x)=\frac{-x(-2i)^{-l}}{2\Gamma(-i\epsilon)}\int_1^\infty ds(s-1)^{-i\epsilon-1}s^{i\epsilon}h_l^{(1)}(ixs)e^{-x(s-1)} \qquad (5)$$
$$\mathrm{Re}(-i\epsilon)>0.$$

The proof will be simpler if one proceeds from Eq. (5) to Eq. (4). On using the modified Poisson's integral representation for $h_l^{(1)}(ixs)$, $u_2(x)$ as given by Eq. (5) can be written as follows:

$$u_2(x)=\frac{x^{l+1}}{\Gamma(-i\epsilon)\Gamma(l+1)}\int_1^\infty ds(s-1)^{-i\epsilon-1}s^{l+i\epsilon}$$
$$\times\int_1^\infty dp(p-1)^lp^le^{x(1-2sp)} \qquad (6)$$
$$=\frac{x^{l+1}}{\Gamma(-i\epsilon)\Gamma(l+1)}\int_1^\infty ds(s-1)^{-i\epsilon-1}s^{l+i\epsilon}$$
$$\times\int_1^\infty dp\left\{\frac{\partial^l}{\partial\beta^l}\left[\frac{(p-1)^le^{x(1-2ps\beta)}}{(-2xs)^l}\right]\right\}_{\beta=1}. \qquad (7)$$

On making use of the identity[1]

$$\frac{\partial^{l+1}}{\partial s^{l+1}}[s^{i\epsilon}(s-1)^{l-i\epsilon}]=\frac{\Gamma(l+1-i\epsilon)}{\Gamma(-i\epsilon)}s^{i\epsilon-l-1}(s-1)^{-i\epsilon-1}, \qquad (8)$$

* Research supported in part by the National Science Foundation.

[1] This identity is easily proved by induction on l.

one obtains

$$u_2(x)=\frac{x^{l+1}}{\Gamma(l+1-i\epsilon)}\int_1^\infty ds\left\{\frac{\partial^{l+1}}{\partial s^{l+1}}[s^{i\epsilon}(s-1)^{l-i\epsilon}]\right\}\times\left\{\frac{\partial^l}{\partial\beta^l}\left[\frac{e^{x(1-2s\beta)}}{(-2x)^l(2x\beta)^{l+1}}\right]\right\}_{\beta=1} \quad (9)$$

$$=\frac{x^{l+1}}{\Gamma(l+1-i\epsilon)}\int_1^\infty ds(s-1)^{l-i\epsilon}s^{i\epsilon}\times\left\{\frac{\partial^l}{\partial\beta^l}\left[\frac{e^{x(1-2s\beta)}}{(-2x)^l}\right]\right\}_{\beta=1} \quad (10)$$

$$=\frac{x^{l+1}e^x}{\Gamma(l+1-i\epsilon)}\int_1^\infty ds(s-1)^{l-i\epsilon}s^{l+i\epsilon}e^{-2xs}\quad \text{Q. E. D.}$$

As before, upon integrating by parts in Eq. (5) and upon using the regularity at $s=1$ and boundedness at $s=\infty$ of the function $h_l^{(1)}(ixs)$,

$$u_2(x)=\frac{x(-2i)^{-l}}{2\Gamma(1-i\epsilon)}\int_1^\infty ds(s-1)^{-i\epsilon}s^{i\epsilon}\frac{\partial}{\partial s}\times[se^{-x(s-1)}h_l^{(1)}(ixs)]. \quad (11)$$

Thus $u_2(x)$ as given by Eq. (11) is also an analytic function of ϵ for $|\epsilon|<1$ for every l, in common with $u_1(x)$ as given by Eq. (3).

We now consider the radial solutions to the Schrödinger equation for a Coulomb potential, corresponding to angular momentum state l. They are usually denoted by $w_1(r)$ and $w_2(r)$, and are related to $u_1(x)$ and $u_2(x)$ by[2]

$$w_1(r)=C_lu_1(x) \quad (12)$$

$$w_2(r)=D_lu_2(x), \quad (13)$$

with

$$\epsilon=(\gamma/p);\quad x=-ipr;\quad p=(2\mathfrak{z})^{\frac{1}{2}} \quad (14)$$

$$\mathfrak{z}=\frac{E(\text{energy})m_0}{\hbar^2};\quad \gamma=\frac{e^2Zm_0}{4\pi\epsilon_0\hbar^2}. \quad (15)$$

These functions are defined in a cut $\mathfrak{z}$ plane, the cut being along the positive real axis. The square root is defined to have a positive imaginary part in the cut plane. The C_l and D_l are constant normalization factors. Since the Green's function constructed from these functions will be independent of their normalizations, we will for simplicity set $C_l=D_l=1$.

The Green's function is obtained, following the usual procedure,[3] as

$$G_l(r',r'',\mathfrak{z})=[J(\mathfrak{z})]^{-1}[\theta(r'-r'')w_2(r')w_1(r'')+\theta(r''-r')w_1(r')w_2(r'')], \quad (16)$$

where

$$\theta(r)=\begin{cases}1 & r>0\\ 0 & r<0\end{cases}\quad\text{for}$$

and the Wronskian

$$J(\mathfrak{z})=w_1(r)w_2'(r)-w_2(r)w_1'(r)=ip(2)^{-2l-1}. \quad (17)$$

Thus, using Eqs. (3) and (11), for $r''>r'$, one gets

$$G_l(r',r'',\mathfrak{z})=\frac{ipr'r''}{\Gamma(1+i\epsilon)\Gamma(1-i\epsilon)}\int_0^1 dt\int_1^\infty ds[s(1-t)]^{i\epsilon}[t(s-1)]^{-i\epsilon}\times\frac{\partial^2}{\partial t\partial s}[ste^{ip[r'(1-t)-r''(1-s)]}j_l(pr't)h_l^{(1)}(pr''s)]. \quad (18)$$

We note that in this form the radial Green's function is an analytic function of ϵ for $|\epsilon|<1$ for every l. In view of Eq. (14), this corresponds to analyticity in the cut $\mathfrak{z}$ plane outside the circle $|\mathfrak{z}|=(\gamma^2)/2$, i.e., the circle passing through the lowest eigenvalue. However, since we know that the Green's function is an analytic function of $\mathfrak{z}$ in the cut $\mathfrak{z}$ plane (with the cut along positive real axis and the square root having a positive imaginary part in the cut plane), except for poles along the negative real axis extending to

$$\mathfrak{z}=-(\gamma^2)/2,$$

the expression can be continued analytically inside the circle by integrating by parts as many times as is necessary.

We now obtain the three-dimensional Green's function by summing over angular momentum states,[4]

$$G(\mathbf{r}',\mathbf{r}'',\mathfrak{z})=(4\pi r'r'')^{-1}\sum_{l=0}^\infty(2l+1)P_l(\cos\theta)G_l(r'',r'',\mathfrak{z})$$
$$r'=|\mathbf{r}'|;\quad r''=|\mathbf{r}''|;\quad \cos\theta=\mathbf{r}'\cdot\mathbf{r}''/r'r''. \quad (19)$$

In order to interchange the order of summation and the integrations in $G_l(r',r'',\mathfrak{z})$ in Eq. (19), it is sufficient to show that $\sum(\partial^2/\partial t\partial s)[j_l(pr't)h_l^{(1)}(pr''s)]$ converges uniformly and absolutely in some region in the cut $\mathfrak{z}$ plane (i.e., for $\mathrm{Im}p>0$) for all s and t in the ranges of integrations. We first show the uniform and absolute convergence for the series $\sum[j_l(pr't)h_l^{(1)}(pr''s)]$. (The series actually converges to an analytic function of $\mathfrak{z}$.)

On using the familiar expansion

$$\frac{e^{ip|s\mathbf{r}''-t\mathbf{r}'|}}{ip|s\mathbf{r}''-t\mathbf{r}'|}=\sum_{l=0}^\infty(2l+1)P_l(\cos\theta)\times j_l(pr't)h_l^{(1)}(pr''s) \quad (20)$$
$$\mathrm{Im}(p)>0,$$

[2] See, for example, N. F. Mott and H. S. W. Massey, *Theory of Atomic Collisions* (Clarendon Press, Oxford, England, 1949), 2nd ed., Chap. III.

[3] B. Friedman, *Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1956), Chap. 3.

[4] Remembering that

$$\delta(\mathbf{r}''-\mathbf{r}')=(4\pi r'r'')^{-1}\delta(r''-r')\Sigma(2l+1)P_l(\cos\theta).$$

one sees by inspection of the left-hand side that the expansion converges in the complex $\cos\theta$ plane inside an ellipse with foci at ± 1, and with semimajor axis

$$M=[(sr'')^2+(tr')^2]/(2str'r'')>1. \tag{21}$$

On the other hand, a function $f(\cos\theta,tr',sr'')$ which is analytic on and inside an ellipse E with foci at ± 1 can be expanded in a Legendre series,

$$f(\cos\theta,tr',sr'')=(2\pi i)^{-1}\sum_{l=0}^{\infty}(2l+1)P_l(\cos\theta) \times\int_E f(y,tr',sr'')Q_l(y)dy, \tag{22}$$

which is uniformly convergent for all $\cos\theta$ lying in any domain wholly inside E.[5] Furthermore, the function $Q_l(y)$, which is Legendre's function of the second kind and which is single valued in the cut y plane with the cut along real axis from -1 to $+1$, satisfies the inequality[6]

$$|Q_l(y)|<(\pi/l)^{\frac{1}{2}}(1-|\omega|^{-2})^{-\frac{1}{4}}|\omega|^{-(l+1)}, \tag{23}$$

where $\omega=y+(y^2-1)^{\frac{1}{2}}$, and the real part of the square root is understood to have the same sign as the real part of y. The formula (23) is valid for $|\omega|>1$, which is certainly the case for y on any ellipse with foci at ± 1.

If one chooses the semimajor axis to be

$$M_E=[(r'')^2+(\beta r')^2]/(2\beta r'r'')>1, \tag{24}$$

with

$$(r''/r')>\beta>1, \tag{25}$$

then

$$M_E<[(r'')^2+(r')^2]/(2r'r'')<M, \tag{26}$$

and

$$|\omega|=r''/(\beta r'). \tag{27}$$

One therefore obtains, for $l>0$,

$$|j_l(ptr')h_l^{(1)}(psr'')|=\left|\frac{1}{2\pi i}\int_E\frac{\exp\{ip[(sr'')^2-2str'r''y+(tr')^2]^{\frac{1}{2}}\}}{ip[(sr'')^2-2str'r''y+(tr')^2]^{\frac{1}{2}}}Q_l(y)dy\right|<N_\beta[(\beta r')/r'']^{l+1}. \tag{28}$$

The existence of such a constant N_β, which is independent of l, t, and s, follows from the boundedness of

$$\frac{\exp\{ip[(sr'')^2-2str'r''y+(tr')^2]^{\frac{1}{2}}\}}{ip[(sr'')^2-2str'r''y+(tr')^2]^{\frac{1}{2}}}, \tag{29}$$

as a function of s and t, $(1\geqslant t\geqslant 0, s\geqslant 1)$, for ϑ in any bounded region in some sector in the cut plane, and y on the ellipse E.

Since differentiating Eq. (29) does not change either its analytic properties or its boundedness, one concludes that the series $\sum(\partial^2/\partial t\partial s)[j_l(ptr')h_l^{(1)}(psr'')]$ is also uniformly and absolutely convergent for $r''>r'$, $1\geqslant t\geqslant 0$, $s\geqslant 1$, and for ϑ in some region in a sector in the cut plane. Furthermore, because of the presence of the term $\exp(ipsr'')$ in front of $j_l(ptr')h_l^{(1)}(psr'')$ in Eq. (18), the restrictions on the sector of ϑ can be relaxed, and one actually obtains, somewhat as a by-product, an estimate on the product of two confluent hypergeometric functions *as a function of* l, viz.,

$$|G_l(r',r'',\vartheta)|<N_\beta'(r',r'')[(\beta r')/r'']^{l+1}. \tag{30}$$

This is valid for ϑ in any region in the whole cut plane outside the circle $|\vartheta|=(\gamma^2)/2$.

Now, interchanging the order of summation and the integrations in $G_l(r',r'',\vartheta)$ of Eq. (19), and carrying out the summation, one obtains the principal result:

$$G(\mathbf{r}',\mathbf{r}'',\vartheta)=\frac{1}{4\pi\Gamma(1+i\epsilon)\Gamma(1-i\epsilon)}\int_0^1 dt\int_1^\infty ds[s(1-t)]^{i\epsilon}[t(s-1)]^{-i\epsilon} \times\frac{\partial^2}{\partial t\partial s}(ts|s\mathbf{r}''-t\mathbf{r}'|^{-1}\exp\{ip[r'(1-t)+r''(s-1)+|s\mathbf{r}''-t\mathbf{r}'|]\}), \tag{31}$$

where $p=(2\vartheta)^{\frac{1}{2}}$, and $r''>r'$. The formula is valid in the cut ϑ plane (with the cut along positive real axis) outside the circle $|\vartheta|=(\gamma^2)/2$, the square root being defined to have a positive imaginary part in the cut plane. We treated the case where the potential is attractive. In the case the potential is repulsive, the sign of ϵ must be reversed.

One notes that in the limit $\epsilon\to 0$, the expression correctly goes over to the free-particle Green's function,

$$\lim_{\epsilon\to 0}G(\mathbf{r}',\mathbf{r}'',\vartheta)=-\frac{e^{ip|\mathbf{r}''-\mathbf{r}'|}}{4\pi|\mathbf{r}''-\mathbf{r}'|}. \tag{32}$$

ACKNOWLEDGMENT

One of the authors (E.H.W.) wishes to thank Dr. Hiroshi Suura for helpful discussions.

[5] E. W. Hobson, *Spherical and Ellipsoidal Harmonics* (Cambridge University Press, New York, 1931), p. 62.

[6] See footnote 5, p. 61.

Analytical Properties of S Matrix and Uniqueness of the Scattering Potential

A. O. BARUT AND K. H. RUEI
Physics Department, Syracuse University, Syracuse, New York
(Received August 12, 1960)

The Schrödinger equation with the complex momentum k leads to an S matrix with very simple analytical properties. It differs from the conventional S matrix as little as one wishes on the real k axis, but it has, in general, completely different analytical behavior outside the real axis. The present formulation removes some of the unsatisfactory features of the conventional formalism in the sense that no redundant poles can occur and a phase shift determines the scattering potential uniquely. The complete analytical behavior of the S matrix, in particular at infinity, is discussed and the theory is extended to Klein-Gordon and Dirac equations with central potential.

1. INTRODUCTION

ANALYTICAL properties of the S matrix as a function of the momentum k or energy E has been the subject of extensive study. This is a central problem in the S matrix approach and in the theory of dispersion relations. The fundamental question is whether or not the S matrix (or the dispersion relations) contain enough physical information that it can replace the dynamics of the system, i.e., the Hamiltonian or the equations of motion.

Although information about the bound states and the decaying and capture states can be obtained from the S matrix, it has not been possible to answer the above question completely in the affirmative. For example, in the case of potential scattering the S matrix may have redundant poles which do not correspond to bound states,[1] or one can give examples showing that the phase shifts (or S matrix) do not determine the potential uniquely.[2] Furthermore, the dispersion relations for the scattering amplitude have extra solutions which do not correspond to the solutions of the Schrödinger equation.[3] In most cases even the complete analytical properties of the S matrix in the whole complex plane are not known, since one can continue S analytically only in a region of the complex plane.[4]

In this paper we give a slightly modified form of the S matrix which has very simple analytical properties in the whole complex k plane and avoids all the difficulties just mentioned. In particular, it has no redundant poles and allows a unique determination of the potential. The basic idea is to start from a Schrödinger (or Dirac) equation with complex k (hereafter referred to as the complex Schrödinger equation) rather than perform an analytic continuation on the solutions of the Schrödinger equation with real k. It is shown that in order for the complex Schrödinger equation to have an asymptotic solution proportional to $\sin[kr-(l\pi/2)+\eta_l(k)]$, the potential must have a cutoff at arbitrarily large distances. This condition changes the S matrix on the real axis extremely slightly, but it has very large effects outside. This is a well-known situation in the theory of analytic functions. Two functions may differ only slightly in some domain, but may have completely different behavior outside this domain. Alternatively, the preceding condition may be taken as the criterion under which all the previously mentioned simplifications are achieved.

We discuss the complete analytical properties of the S matrix for all l. In particular, we show that $S_l(k)$ has an essential singularity at infinity in the lower half-plane, the number of poles at infinity being infinite. Thus, it is shown that a representation of the S matrix as a product of its zeros and poles[5] is not meaningful, and that dispersion relations for the S matrix in the lower k plane which have been recently proposed[6] are not valid.

We first discuss the analytical properties of the functions $F_l(k)$ introduced by Levinson.[7] In addition, we introduce the functions $F_l^*(k)$ (Sec. II). The S matrix is expressed in terms of $F_l(k)$ and $F_l^*(k)$. In Sec. III we investigate the analytical properties of $S_l(k)$, especially at infinity. In Sec. IV we show how the redundant poles are eliminated; in Sec. V we discuss the uniqueness of the potential for a given phase shift. Finally in Sec. VI the theory is extended to the Klein-Gordon and Dirac equations with central potential. Most of our conclusions about finite range potentials agree with what is known about them in special cases.[8]

We use units such that $\hbar^2/2m=1$.

II. ANALYTICAL PROPERTIES OF $F_l(k)$

It is known that the radial Schrödinger equation

$$u_l''(r,k)+\left[k^2-\frac{l(l+1)}{r^2}-V(r)\right]u_l(r,k)=0, \qquad (1)$$

where we assume k to be complex from the beginning,

[1] S. T. Ma, Phys. Rev. **69**, 668 (1946); **71**, 195 (1947).
[2] V. Bargmann, Revs. Modern Phys. **21**, 488 (1949).
[3] S. Gasiorowicz and H. A. Ruderman, Phys. Rev. **107**, 868 (1957).
[4] R. Jost (unpublished); J. Bowcock and D. Walecka, Nuclear Phys. **12**, 371 (1959); A. Martin, Nuovo cimento **15**, 98 (1960).
[5] N. G. van Kampen, Phys. Rev. **89**, 1072 (1953); **90**, 1267 (1953). N. Hu, Phys. Rev. **74**, 131 (1948).
[6] B. W. Lee, Phys. Rev. **112**, 2122 (1958).
[7] N. Levinson, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **25**, No. 9 (1949).
[8] S. T. Ma, Phys. Rev. **71**, 195 (1947).

$k=\lambda+i\chi$, can be written in the form of an integral equation[7]

$$\mathcal{U}_l(r,k)=\frac{j_l(kr)}{k^{l+1}}-\frac{1}{k}\int_0^r g_l(r,\xi,k)V(\xi)\mathcal{U}_l(\xi,k)d\xi, \quad (2)$$

with the boundary condition

$$\mathcal{U}_l(0,k)=0, \quad (3)$$

where

$$g_l(r,\xi,k)=j_l(kr)n_l(k\xi)-j_l(k\xi)n_l(kr), \quad (4)$$

and $j_l(z)$, $n_l(z)$ are Riccatti Bessel functions.[9]

Equation (1) has, under the boundary condition in Eq. (3) and for vanishing potential at infinity, the asymptotic form

$$\lim_{r\to\infty}\mathcal{U}_l(r,k)=\frac{A_l(k)}{k^{l+1}}\sin\left[kr-\frac{\pi l}{2}+\eta_l(k)\right], \quad (5)$$

where $A_l(k)$ and $\eta_l(k)$ are continuous functions of k. Note that $\mathcal{U}_l(k)=\mathcal{U}_l(-k)$; hence $A_l(k)$ is an even, $\eta_l(k)$ is an odd function of k (see Appendix II).

We define a function $F_l(k)$ by[7]

$$\begin{aligned}F_l(k)&=1+ik^l\int_0^\infty \mathcal{U}_l(\xi,k)V(\xi)h_l(k\xi)d\xi\\&=F_l^{(1)}(k)+iF_l^{(2)}(k),\end{aligned} \quad (6)$$

where

$$\begin{aligned}F_l^{(1)}(k)&=1-k^l\int_0^\infty \mathcal{U}_l(\xi,k)V(\xi)n_l(k\xi)d\xi\\F_l^{(2)}(k)&=k^l\int_0^\infty \mathcal{U}_l(\xi,k)V(\xi)j_l(k\xi)d\xi,\end{aligned} \quad (7)$$

since

$$h_l(k\xi)=j_l(k\xi)+in_l(k\xi). \quad (8)$$

Note that $F_l^{(1)}(k)$ and $F_l^{(2)}(k)$ are complex, since k is. $F_l^{(1)}(k)$ is an even function of k, $F_l^{(2)}(k)$ is an odd function of k.

The asymptotic form of Eq. (2), compared with Eq. (5), gives (generalizing a result of Levinson[7] for $l=0$ to arbitrary l and complex k) the following relations:

$$\begin{aligned}A_l^2(k)&=F_l^{(1)2}(k)+F_l^{(2)2}(k),\\\tan\eta_l(k)&=-[F_l^{(2)}(k)/F_l^{(1)}(k)].\end{aligned} \quad (9)$$

This is shown in Appendix I. Note that for real k, $k=\lambda$, $\eta_l(\lambda)$ are the phase shifts and we have

$$\begin{aligned}A_l(\lambda)&=|F_l(\lambda)|,\\\eta_l(\lambda)&=-\arg F_l(\lambda).\end{aligned} \quad (9')$$

Thus, the function $F_l(k)$ determines the asymptotic form of the problem completely. S matrix, partial wave amplitudes $f_l(k)$, and asymptotic wave function can all be expressed in terms of it (see Appendix I):

$$\lim_{r\to\infty}\mathcal{U}_l(k)=\frac{1}{2(-ik)^{l+1}}[F_l(k)e^{-ikr}+(-1)^{l+1}F_l(k)e^{ikr}] \quad (10)$$

$$f_l(k)=\frac{1}{2ik}(2l+1)\left[\frac{F_l^*(k)}{F_l(k)}-1\right] \quad (11)$$

$$S_l(k)=(-1)^l[F_l^*(k)/F_l(k)], \quad (12)$$

where

$$F_l^*(k)=F_l^{(1)}(k)-iF_l^{(2)}(k)=[F_l(k^*)]^*, \quad (13)$$

and

$$F_l^*(k)=F_l(-k). \quad (14)$$

First we investigate whether or not $F_l(k)$ is defined in the whole complex plane. Considering the asymptotic behavior of the $h_l(k)$ function, i.e.,

$$\lim_{\substack{\xi\to\infty\\k\neq 0}} h_l(k\xi)\sim e^{ik\xi}, \quad \lim_{\substack{\xi\to\infty\\k\neq 0}} h_l^*(k\xi)\sim e^{-ik\xi}, \quad (15)$$

we see that the integral in the defining Eq. (6) for $F_l(k)$ exists in the upper half-plane, but in general not in the lower half-plane, where we will have for large ξ a factor $e^{-\chi\xi}$ with $\chi<0$, and the integral does not exist. Similarly, $F_l^*(k)$ will exist in the lower half-plane and not in the upper half-plane. Hence, in order for the S matrix (12) to be defined, or equivalently, in order to have the asymptotic solution (5) for all k, the potential must be such that it compensates the exponential increase in $h_l(k)$. Potentials increasing faster than the exponential, say a Gaussian potential, satisfy this condition. One can, however, impose much weaker conditions, by putting a cutoff to the potential at arbitrarily large distances. One may say, therefore, that a cutoff in the potential at arbitrarily large distances is a consequence of the solubility of the complex Schrödinger equation in the form of Eq. (5). Of course, such a cutoff does not change the physics of the problem, but as we shall see it will simplify the analytical properties of $S_l(k)$ and will eliminate the unsatisfactory features of the usual formalism mentioned in the introduction.

We study now the analytic properties of $F_l(k)$ and $F_l^*(k)$ in the complex k plane. Levinson[7] has shown that for $l=0$, $\mathcal{U}_0(r,k)$ and hence $F_0(k)$ are analytic in the upper half-plane and continuous in the region $\chi\geqslant 0$ and

$$\lim_{|k|\to\infty} F_0(k)=1$$

in the upper half-plane, provided

$$\int_0^\infty r|V(r)|dr<\infty. \quad (16)$$

[9] Note that these $j_l(z)$ and $n_l(z)$ differ from the spherical Bessel functions by a factor z, also represented by the same symbols; see, for example, L. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), p. 77.

One can generalize this theorem to arbitrary l, provided, in addition to (16), another condition is satisfied:

$$\int_0^\infty |V(r)|dr<\infty. \tag{17}$$

The proof is given in Appendix II, where we also show the following asymptotic expression:

$$\lim_{\chi\to-\infty}|F_l(k)-1|=\frac{Q}{|k|}\int_0^\infty e^{-2\chi\xi}|V(\xi)|d\xi, \tag{18}$$

with Q some constant. Similarly $F_l^*(k)$ is analytic and

$$\lim_{\chi\to+\infty}|F_l^*(k)-1|=\frac{Q}{|k|}\int_0^\infty e^{2\chi\xi}|V(\xi)|d\xi. \tag{18'}$$

Therefore, $F_l^*(k)$ has an essential singularity at infinity in the upper half-plane.

Using the Schwartz' reflection theorem in the theory of analytic continuation, it is seen that

$$F_l^*(k)=[F_l(k^*)]^*, \quad \text{and} \quad [F_l^*(k^*)]^*=F_l(k)$$

are also analytic in the lower-half k plane, with the difference that now $F_l^*(k)$ approaches 1 as $|k|\to\infty$ and $F_l(k)$ has an essential singularity at infinity as it is seen from Eqs. (18) and (18'). As a matter of fact, from Eq. (14) we see that $F_l^*(k)$ in the lower half-plane is equal to $F_l(k)$ in the upper half-plane.

The wave function $v_l(r,k)$ is bounded (Appendix II); so is the function $h_l(k\xi)$, except for the essential singularity at infinity. Hence from its definition, Eq. (6), we see that $F_l(k)$ and $F_l^*(k)$ have no other singularities in the finite k plane.

We conclude, therefore, that both $F_l(k)$ and $F_l^*(k)$ are analytic in the whole complex k plane with no singularities, except the essential singularity at infinity in the upper half plane for $F_l^*(k)$, and in the lower half plane for $F_l(k)$.

III. SCATTERING MATRIX

The scattering matrix in the momentum representation is given by Eq. (12). From the symmetry of $F_l(k)$ and Eq. (14), we derive the unitarity and the symmetry of the S matrix

$$S_l(k)S_l^*(k)=1, \quad S_l(k)=S_l^*(-k). \tag{19}$$

We remark at this point that the previous results will hold even for a complex, i.e., non-Hermitian potential, except for Eq. (14); hence the unitarity of the S matrix.

Being essentially the ratio of $F_l^*(k)$ and $F_l(k)$, the scattering matrix will be analytic in the whole complex plane with an essential singularity at infinity in the upper half-plane, and with poles corresponding to the zeros of $F_l(k)$. Since $F_l^*(k)$ has no poles in the finite k plane, there will be no other singularities.

One may, therefore, attempt to write dispersion relations for $S_l(k)$ in the lower half-plane,[6] since there the numerator of Eq. (14) approaches 1 at infinity and the denumerator has an essential singularity. This would be possible, if $F_l(k)$ did not have any zeros at infinity. We will show, however, that $F_l(k)$ will have zeros at infinity in the lower half-plane. The proof is very simple and utilizes the following theorem: An analytic function comes arbitrarily close to any complex value in every neighborhood of its essential singularities.[10] $F_l(k)$ takes the value 1 as $|k|\to\infty$ in the upper half-plane; hence it must take the value zero at infinity somewhere in the lower half-plane.

As an example, we consider the square-well potential for which

$$S_0(k)=\frac{F_0^*(k)}{F_0(k)}=\frac{[\cos ak'+i(k/k')\sin ak']e^{-iak}}{[\cos ak'-i(k/k')\sin ak']e^{iak}}, \tag{20}$$

with $k'=+(k^2+V_0)^{\frac{1}{2}}$, where a is the range and V_0 is the depth of the potential. All the previous theorems can easily be verified on this example. The zeros of $F_l(k)$, i.e., the poles of $S_l(k)$, are given by the solutions of the equation

$$\cos ak'-i(k/k')\sin ak'=0. \tag{21}$$

On putting $k=\lambda+i\chi$, $k'=x+iy$ and expanding $k'=k+(1/2k)V_0+\cdots$, it is easy to see that Eq. (21) is satisfied at infinity only if χ is negative (lower half-plane) and both λ and χ tending to infinity where λ is of the order of $e^{-\chi}$. The number of such zeros is infinite.

We next prove quite generally that the number of poles of the S matrix in the lower half-plane is infinite. To see this, we first show in Appendix III that if $S_l(k)$ had a finite number of poles, it could be written in the following form:

$$S_l(k)=\pm e^{ick}\prod_\alpha\frac{k-k_\alpha^*}{k-k_\alpha}\prod_\beta\frac{(k-k_\beta^*)(k+k_\beta)}{(k-k_\beta)(k+k_\beta^*)}, \tag{22}$$

where c is a negative number, k_α are the poles corresponding to the bound states, and k_β, k_β^* poles corresponding to the decaying and capture states which are symmetrical around the imaginary axis (see Sec. IV). This is the expression also derived by Hu and others[5] in the conventional theory which assumes, we emphasize, a finite number of poles.

The expression (22) gives correctly the essential singularity at infinity in the upper half-plane, but it contradicts our previous result that $S_l(k)$ has poles at infinity in the lower half-plane; it does not take into account the infinite number of poles at infinity. Moreover, for real k and $|k|\to\infty$, Eq. (22) would give $S_l(k)\to\pm e^{ick}$ or phase shifts $\eta_l(k)\to\frac{1}{2}ck$, or cross sections which fluctuate as $k=\lambda\to\infty$, whereas Eq. (12) gives $S_l(k)\to(-1)^l$, or $\eta_l(k)\to n\pi$, as $|k|\to\infty$ which is correct from a physical point of view.

[10] N. Ahlfors, *Complex Analysis* (McGraw-Hill Book Company, Inc., New York, 1953), p. 114.

IV. ELIMINATION OF REDUNDANT POLES

In the conventional S matrix theory, one shows that the poles of the S matrix on the positive imaginary axis correspond to the bound states, and the poles in the lower half-plane symmetric to and off the imaginary axis correspond to the decaying and capture states.[11] These are the "true" poles of the S matrix. In addition, there are the so-called "redundant" poles[1] which do not correspond to bound states.

It will be shown now that in our formulation the true poles will correspond to the zeros of $F_l(k)$ and the redundant poles to the poles of $F_l^*(k)$ in Eq. (12). Since $F_l(k)$ and $F_l^*(k)$ are finite everywhere, no redundant poles will occur, nor any other singularity as found by Regge.[12]

To see the meaning of the zeros of $F_l(k)$, we use Eq. (10). Clearly, if $F_l(k)=0$ for $k=i\chi$, $\chi>0$, we get exponentially decreasing wave functions, i.e., true bound states.

It remains to see the relation of the present formulation with the conventional analytical continuation and the origin of the redundant poles in the latter theory. For this purpose we write the solution of Eq. (1) for real k under the boundary condition (3) in the following form:

$$\mathcal{U}_l(k,r)=N_l(k)[f_l(-k)f_l(k,r)e^{-ikr} -f_l(k)f_l(-k,r)e^{ikr}], \quad (23)$$

where $f_l(k,r)e^{-ikr}$, $f_l(-k,r)e^{ikr}$ are two independent solutions of Eq. (1), $N_l(k)$ a normalization factor, and

$$f_l(-k)=f_l(-k,0); \quad f_l(k)=f_l(k,0).$$

If we compare the asymptotic form of Eq. (23) with Eq. (10) we get

$$S_l(k)=(-1)^l\frac{F_l^*(k)}{F_l(k)}=(-1)^l\frac{f_l(k)f_l(-k,\infty)}{f_l(-k)f_l(k,\infty)}, \quad (24)$$

and

$$\mathcal{U}_l(k,r) =N_l'(k)\left[\frac{f_l(k,r)}{f_l(k,\infty)}e^{-ikr}-S_l(k)\frac{f_l(-k,r)}{f_l(-k,\infty)}e^{ikr}\right]. \quad (23')$$

In this form the S matrix depends on the values of the solutions $f_l(k,r)$, $f_l(-k,r)$, both at zero and at infinity. Since both F_l^* and F_l approach 1 as $|k|\to\infty$ for real k, we can actually identify

$$F_l^*(k)=\frac{f_l(k,0)}{f_l(k,\infty)}, \quad F_l(k)=\frac{f_l(-k,0)}{f_l(-k,\infty)}. \quad (25)$$

The poles of $F_l^*(k)$, which are the zeros of $f_l(k,\infty)$, are redundant as one can see from Eq. (23). At such zeros, both terms with e^{ikr} and e^{-ikr} do not vanish. This is the connection with the conventional formulation for real k. The fact that the analytic continuation of Eqs. (23)–(25) does not necessarily agree with our $S_l(k)$ results from the fact that the asymptotic form of the wave function does not have the required form for complex k. Consider, for example, the following potential used by Jost[13]:

$$V(r)=2\alpha e^{-r}/(1-\alpha e^{-r})^2, \quad |\alpha|<1. \quad (26)$$

Assuming a solution of the form

$$\frac{f_0(k,r)}{f_0(k,\infty)}=\sum_{\nu=0}^{\infty}c_\nu(k)e^{-\nu r}, \quad c_0=1, \quad (27)$$

we find from the Schrödinger equation $c_\nu(k)=2\alpha^\nu/(1+2ik)$, and from Eq. (24)

$$S_0(k)=\frac{\{k-i[(1+\alpha)/2(1-\alpha)]\}(-k-\frac{1}{2}i)}{(k-\frac{1}{2}i)\{-k-i[(1+\alpha)/2(1-\alpha)]\}}. \quad (28)$$

The limit of Eq. (27) as $r\to\infty$ should be 1. This condition is satisfied for $|\chi|<\frac{1}{2}$ but not for other complex k values, for

$$\frac{f_0(k,r)}{f_0(k,\infty)}=\left(1+\frac{2\alpha}{1+2ik}e^{-r}+\frac{2\alpha^2}{(1+2ik)^2}e^{-2r}+\cdots\right);$$

hence if $|\chi|>1$, the second term cannot be neglected. Indeed this example leads to a true pole at

$$k=-i[(1+\alpha)/2(1-\alpha)],$$

which is a zero of $F_l(k)$. In addition, there is a redundant pole at

$$k=i/2,$$

which is a pole of $F_0^*(k)$.

We conclude, therefore, that the solution (27) does not satisfy the asymptotic condition for complex k as we have required from the beginning. This holds true for all other examples in which redundant poles occur.

It has been shown by Ma[8] that in the example of the exponential potential, a cutoff at arbitrary large distances removes the redundant poles. Our treatment shows that this procedure is quite general and gives its foundation and justification.

We remark finally that the Heisenberg relations

$$\int_{-\infty}^{\infty}S_l(k)e^{ik(r+r')}dk=\sum_n|c_n|^2e^{-k_n(r+r')}, \quad (29)$$

where the summation on the right-hand side is over bound states, which are not valid in the presence of redundant poles, regain their validity in the present formulation.

[11] C. Møller, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. 23, No. 1 (1945).

[12] T. Regge, Nuovo cimento 9, 295 (1958).

[13] R. Jost, Helv. Phys. Acta 20, 256 (1947).

V. UNIQUENESS OF THE SCATTERING POTENTIAL FOR A GIVEN PHASE SHIFT

Levinson has shown that a single phase shift, $\eta_l(k)$, l fixed, determines the scattering potential uniquely, provided there are no bound states for the same l and the potential goes sufficiently rapidly to zero at infinity.[7] If, however, there are n discrete bound states for the given l, then examples can be constructed showing continuous families of potentials involving n parameters corresponding to the same given phase shift, the so-called "phase-equivalent potentials."[2] All these examples involve redundant poles for the S matrix. We will show now that with the elimination of the redundant poles there is also a unique connection between a phase shift $\eta_l(k)$ and the scattering potential. The point is that the solutions considered in the case of phase equivalent potentials do not satisfy the asymptotic condition for the solutions of the complex Schrödinger equation, as in the example of Eqs. (26) and (27). We give now a general proof.

Let $\eta_l(k)$ or $S_l(k)$ be the given phase shift or S matrix. Then from Eqs. (12) and (6) we get

$$\int \mathfrak{v}_l(k,\xi)V(\xi)[(-1)^l S_l(k)h_l(k\xi)+h_l^*(k\xi)]d\xi = \frac{1}{ik}[1-(-1)^l S_l(k)]. \quad (30)$$

If there are two distinct potentials $V(\xi)$ and $V'(\xi)$ and two solutions $\mathfrak{v}_l(k,\xi)$ and $\mathfrak{v}_l'(k,\xi)$ corresponding to the same S matrix, we get, from Eq. (30),

$$\int_0^\infty [\mathfrak{v}_l(k,\xi)V(\xi)-\mathfrak{v}_l'(k,\xi)V'(\xi)] \times[(-1)^l S_l(k)h_l(k\xi)+h_l^*(k\xi)]d\xi=0. \quad (31)$$

Moreover, $\mathfrak{v}_l$ and $\mathfrak{v}_l'$ must be the same asymptotically as $\xi\to\infty$, since the phase shift is given. This means that the integrand of Eq. (31) behaves at $\xi\to\infty$ either as $(V-V')e^{2ik\xi}$ or as $(V-V')e^{-2ik\xi}$ for all complex k, depending whether $\chi>0$, or $\chi<0$. The integrand must, however, go to zero as $\xi\to\infty$, if the integral has to vanish. This is only possible for all complex k, if $(V-V')$ is identically zero (not just approaching zero!). Then by continuity $V-V'$ must be zero for all ξ.

VI. RELATIVISTIC CASE

The previous discussion can be extended readily to the relativistic case. The radial wave function for the Klein-Gordon equation satisfies[14]

$$\mathfrak{v}_l''(k,r)+\left[\lambda^2-V(r,k)-\frac{l(l+1)}{r^2}\right]\mathfrak{v}_l(k,r)=0, \quad (32)$$

where

$$V(r,k)=[2(k^2+1)^{\frac{1}{2}}-V(r)]V(r). \quad (33)$$

Equations (2) and (6) hold exactly here too with $V(\xi)$ replaced by $V(\xi,k)$ as given by Eq. (33). The condition on the potential are here

$$\int_0^\infty \frac{\xi|V(\xi,k)|d\xi}{1+(k\xi)}<\infty \qquad \int_0^\infty \frac{|V(\xi,k)|}{|k|}d\xi<\infty. \quad (34)$$

By the same methods as in the nonrelativistic case, one finds that $F_l(k)$ has the same analytical properties as before except that it does not approach unity as $|k|\to\infty$ in the upper half-plane. Indeed for $k=\lambda\to\infty$ we have the asymptotic form

$$\lim_{\lambda\to\infty} F_l(\lambda)=1+i\int_0^\infty j_l^2(\lambda\xi)\frac{V(\xi,\lambda)}{\lambda}d\xi,$$

which is clearly different from 1 since $V(\xi,\lambda)/\lambda\to 2V(\xi)$ from Eq. (33). Thus, $S_l(\lambda)$ will not approach 1 as $\lambda\to\infty$. Hence the phase shifts approach a constant value at high energies rather than the value $n\pi$ as in the nonrelativistic case.

For Dirac particles, the radial wave function corresponding to our $\mathfrak{v}_l(k,\xi)$ consists of two functions which we denote by $M_K(r)$ and $N_K(r)$ satisfying the coupled first-order equations[15]

$$[E+1-V(r)]M_K(r)-N_K(r)-(K/r)N_K(r)=0$$
$$[E-1-V(r)]N_K(r)+M_K(r)-(K/r)M_K(r)=0,$$

in the state with the quantum number K, where $K^2=(j+\frac{1}{2})^2$. M_K and N_K correspond to $j=l-\frac{1}{2}$ and $j=l+\frac{1}{2}$, respectively, and satisfy the same second order equation.

$$M_K''(r)+\{\lambda^2-[K(K-1)/r^2]-V(\lambda,r)\}M_K(r)=0$$
$$N_K''(r)+\{\lambda^2-[K(K+1)/r^2]-V(\lambda,r)\}N_K(r)=0,$$

where $V(\lambda,r)$ is the same as in Eq. (33). Thus we reach the same conclusions as in the case of Klein-Gordon equation. In particular, the phase shifts will approach a constant value at high energies.

CONCLUSIONS

We believe that we have clarified some of the ambiguous and unphysical features concerning the analytical properties of the S matrix for potential scattering. The results we have proved are the following:

1. In order for the Schrödinger equation with complex k to have asymptotic solutions for all k, the potential must have a cutoff at arbitrary large distances, or it must be such that the integral in Eq. (6) exists.

[14] L. Schiff, footnote 9, p. 321.

[15] L. Schiff, footnote 9, p. 335.

2. This leads to an S matrix given by

$$S_l(k)=(-1)^l[F_l^*(k)/F_l(k)],$$

which differs for real k from the conventional S matrix as little as one pleases, but in general has completely different analytical properties outside the real axis.

3. In the foregoing expression for $S_l(k)$, the functions $F_l^*(k)$ and $F_l(k)$ are analytic in the whole complex k plane with no singularities except an essential singularity at infinity in the upper half-plane, for $F_l^*(k)$ and in the lower half-plane for $F_l(k)$.

4. The S matrix so defined has no redundant poles; all the poles of the S matrix on the positive imaginary axis correspond to true bound states. The solutions in the conventional S matrix theory which give rise to redundant poles do not satisfy our boundary conditions for the solubility of the complex Schrödinger equation.

5. The number of poles of the S matrix in the lower half-plane whlch correspond to decaying and capture state is infinite. Therefore, dispersion relations for the S matrix in the lower-half k plane do not exist.

6. With the elimination of the redundant zeros there is a unique relationship between a phase shift and the scattering potential. The so-called phase-equivalent potentials do not survive.

7. The theory can be extended to the relativistic equations with a central potential and essentially the same results hold.

APPENDIX I. ASYMPTOTIC EXPRESSIONS

Equation (2) in the limit $r\to\infty$ becomes

$$\lim_{r\to\infty}\mathcal{U}_l(r,k)=\frac{j_l(kr)}{k^{l+1}}-\frac{1}{k}\int_0^\infty g_l(r,\xi k)V(\xi)\mathcal{U}_l(\xi,k)d\xi$$

or using Eq. (8),

$$\lim_{r\to\infty}\mathcal{U}_l(r,k)=\frac{j_l(kr)}{k^{l+1}}-\frac{1}{k}j_l(kr)\int_0^\infty n_l(k\xi)V(\xi)\mathcal{U}_l(\xi,k)d\xi$$
$$+\frac{1}{k}n_l(kr)\int_0^\infty j_l(k\xi)V(\xi)\mathcal{U}_l(\xi,k)d\xi$$
$$=\frac{1}{k^{l+1}}[j_l(kr)F_l^{(1)}(k)+n_l(kr)F_l^{(2)}(k)],\quad\text{(I.1)}$$

where we have used Eqs. (7). Asymptotically

$$j_l(kr)\to\sin(kr-\tfrac{1}{2}\pi l),\quad n_l(kr)\to-\cos(kr-\tfrac{1}{2}\pi l),$$

and defining

$$\cos\eta_l(k)=\frac{F_l^{(1)}(k)}{A_l(k)},\quad \sin\eta_l(k)=-\frac{F_l^{(2)}(k)}{A_l(k)},\quad\text{(I.2)}$$

where

$$A_l(k)=[F_l|^{(1)^2}(k)+F_l^{(2)^2}(k)]^{\frac{1}{2}}=[F_l(k)F_l^*(k)]^{\frac{1}{2}},$$

we get

$$\lim_{r\to\infty}\mathcal{U}_l(r,k)=\frac{A_l(k)}{k^{l+1}}[\sin(kr-\tfrac{1}{2}\pi l)\cos\eta_l(k)$$
$$+\cos(kr-\tfrac{1}{2}\pi l)\sin\eta_l(k)],$$

which is just the Eq. (5). To show Eq. (10), we insert in (I.1)

$$j_l(kr)\to\frac{1}{2i}(e^{i(kr-\frac{1}{2}\pi l)}-e^{-i(kr-\frac{1}{2}\pi l)})$$

$$n_l(kr)\to-\tfrac{1}{2}(e^{i(kr-\frac{1}{2}\pi l)}+e^{-i(kr-\frac{1}{2}\pi l)}),$$

and then use $e^{i\frac{1}{2}\pi l}=i^l$ and Eq. (6).

Equations (11) and (12) follow from the phase shifts $\eta_l(k)$:

$$S_l(k)=(-1)^l e^{2i\eta_l(k)},$$

using Eq. (I.2).

APPENDIX II. ANALYTICAL PROPERTIES

(a) Analyticity of $\mathcal{U}_l(k,r)$

$\mathcal{U}_l(k,r)$ depends on $j_l(k,r)$ and $g_l(r,\xi,k)$ [Eq. (2)]. These functions are bounded for $\chi\geqslant 0$ by the following expressions[7]:

$$|j_l(kr)|\leqslant Ke^{\chi r}\frac{|kr|^{l+1}}{(1+|kr|)^{l+1}},\quad r\geqslant 0$$
$$|g_l(r,\xi,k)|\leqslant Ke^{\chi(r-\xi)}\frac{(1+|k\xi|)^l}{|k\xi|^l}\frac{|kr|^{l+1}}{(1+|kr|)^{l+1}},\quad\text{(II.1)}$$
$$0\leqslant\xi\leqslant r,$$

where K is some finite constant. Let us now consider a sequence of functions $\mathcal{U}_l^{(n)}(k,r)$ defined by

$$\mathcal{U}_l(k,r)=\frac{j_l(kr)}{k^{l+1}}-\frac{1}{k}\int_0^r g_l(r,\xi,k)V(\xi)\mathcal{U}_l^{(n-1)}(k,\xi)d\xi$$
$$\mathcal{U}_l^{(0)}(k,r)=0.$$

By iteration and using the inequalities (II.1) we get

$$|\mathcal{U}_l^{(n)}(r,k)-\mathcal{U}_l^{(n-1)}(r,k)|\leqslant K^n\frac{|r|^{l+1}}{(1+|kr|)^{l+1}}e^{\chi r}\frac{[L(r)]^{n-1}}{(n-1)!},$$

where

$$L(r)=\int_0^r\frac{\xi|V(\xi)|d\xi}{1+|k\xi|}.$$

Thus the sequence $\mathcal{U}_l^{(n)}(r,k)$ approaches uniformly to the limit $\mathcal{U}_l(r,k)$, provided $L(r)$ is finite. This gives us the condition (15) in the text. Each $\mathcal{U}_l^{(n)}(k,r)$ is analytic, since it involves analytic functions $j_l(kr)$ and $g_l(r,\xi,k)$ only; hence the limit $\mathcal{U}_l(r,k)$ is analytic in the upper half-plane. Furthermore, since each $\mathcal{U}_l^{(n)}(r,k)$ is an even function of k, the limit $\mathcal{U}_l(r,k)$ is also an even function of k.

Next we show that $\mathcal{U}_l(r,k)$ itself is bounded by

$$|\mathcal{U}_l(r,k)| \leqslant Q[e^{\chi r}|r|^{l+1}/(1+|kr|)^{l+1}], \quad \text{(II.2)}$$

where Q is a constant. To see this we write

$$\mathcal{U}_l(r,k)=M(r,k)\frac{Ke^{\chi r}|kr|^{l+1}}{(1+|kr|)^{l+1}|k|^{l+1}};$$

upon inserting this into Eq. (2) and using (II.1), we obtain

$$M(r,k)\leqslant 1+\int_0^r \frac{\xi}{1+|k\xi|}|V(\xi)|M(\xi,k)d\xi,$$

or

$$M(r,k)\leqslant K\exp\left(\int_0^r \xi|V(\xi)|d\xi\right)=Q.$$

Hence we obtain Eq. (II.2).

(b) Analytic Properties of $F_l(k)$

Since $\mathcal{U}_l(r,k)$ is analytic, we immediately see from Eq. (6) that $F_l(k)$ is also analytic in the upper half-plane. To show that

$$\lim_{|k|\to\infty} F_l(k)=1 \quad \text{for} \quad \chi\geqslant 0, \quad \text{(II.3)}$$

we write from Eq. (6)

$$|F_l(k)-1| \leqslant |k|^l\int_0^\infty |\mathcal{U}_l(\xi,k)||V(\xi)||h_l(k\xi)|d\xi$$
$$=\int_0^\infty A(k,\xi)|V(\xi)|d\xi,$$

where we have set

$$A(k,\xi)=|k|^l|\mathcal{U}_l(\xi,k)||h_l(k\xi)|.$$

Now $h_l(k\xi)$ is singular at $\xi=0$ for $l\neq 0$, and in the neighborhood of zero

$$\lim_{\xi\to 0}|h_l(k\xi)|=\frac{1\cdot 3\cdots(2l-1)}{|k\xi|^l},$$

since

$$\lim_{z\to 0} j_l(z)=\frac{z^{l+1}}{1\cdot 3\cdots(2l+1)}, \quad \lim_{z\to 0} n_l(z)=\frac{1\cdot 3\cdots(2l-1)}{z^l}.$$

One might think, therefore, that $A(\xi,k)$ is singular for $\xi\to 0$. However, the other term in $A(k,\xi)$, $|\mathcal{U}_l(k,\xi)|$ is bounded as shown by Eq. (II.2). Hence

$$\lim_{\xi\to 0} A(\xi,k)\leqslant\frac{1\cdot 3\cdots(2l-1)}{\xi^l}\cdot\frac{Qe^{\chi\xi}|\xi|^{l+1}}{(1+|k\xi|)^{l+1}}$$

or $A(0,k)=0$ for all k.

If ξ is finite, we can use the bound of $h_l(k\xi)$, i.e.,

$$\lim_{|k|\to\infty}|h_l(k\xi)|=e^{-\chi\xi}; \quad \text{(II.3)}$$

then

$$\lim_{|k|\to\infty} A(k,\xi)=\lim_{|k|\to\infty}\frac{1}{|k|}\frac{Q|k\xi|^{l+1}}{(1+|k\xi|)^{l+1}}$$
$$=\lim_{|k|\to\infty}\frac{Q}{|k|}=0.$$

$A(k,\xi)$ approaches, therefore, zero as $|k|\to\infty$ everywhere $\xi\geqslant 0$ and is continuous. Then

$$\lim_{|k|\to\infty}|F_l(k)-1|=\lim_{|k|\to\infty}\int_0^\infty A(\xi,k)|V(\xi)|d\xi=0,$$

provided the second condition in the text, Eq. (17), is satisfied. This completes the proof of Eq. (II.3).

The asymptotic formulas (18) and (18′) follow also from here by noting that the exponential factor in Eq. (II.2) in the lower half-plane is $e^{-\chi r}$. This together with the $e^{-\chi\xi}$ in Eq. (II.3) gives $e^{-2\chi\xi}$.

APPENDIX III. PROOF OF EQUATION (22)

If $S_l(k)$ did have a finite number of poles, it could always be written in the form

$$S_l(k)=P_l(k)Q_l(k), \quad \text{(III.1)}$$

where $P_l(k)$ contains *all* the zeros and poles other than the essential singularity at infinity which is contained in $Q_l(k)$. Then upon using the unitarity of $S_l(k)$, we can write $P_l(k)$ as

$$P_l(k)=\prod_\alpha\frac{k-k_\alpha^*}{k-k_\alpha}\prod_\beta\frac{(k-k_\beta^*)(k+k_\beta)}{(k-k_\beta)(k+k_\beta^*)},$$

where k_α lie on the positive imaginary axis (bound states), k_β in the lower half-plane off the imaginary axis (decay and capture states).

$Q_l(k)$ having no zeros and poles except the essential singularity must be of the form[16]

$$Q_l(k)=c_l e^{iG_l(k)},$$

where $G_l(k)$ can be either an integral rational or an integral transcendental function. Since $Q_l(k)$ has to satisfy unitarity, $Q_l(k)Q_l^*(k)=1$, and symmetry conditions $Q_l(k)=Q_l^*(-k)$, we get

$$c_l=\pm 1,$$

and

$$G_l(k) \text{ real and odd function of } k.$$

From Eqs. (18) and (18′) we know that $F_l^*(k)$ in the upper half-plane [or $F_l(k)$ in the lower half-plane] has the asymptotic factor e^{ick} where c is a negative number. A rational odd function $G_l(k)$ which behaves at infinity as ck must be itself ck. Therefore,

$$Q_l(k)=\pm e^{ick},$$

and we get Eq. (22) of the text. Thus, this formula is meaningful only for finite number of poles.

[16] N. Harkness and N. Morley, *Introduction to the Theory of Analytic Functions* (G. E. Stechert and Company, New York, 1924), p. 193.

Structure of the Many-Channel S Matrix*†

ROGER G. NEWTON
Department of Physics, Indiana University, Bloomington, Indiana
(Received September 23, 1960)

We are considering the nonrelativistic elastic and inelastic scattering of two particles with internal degrees of freedom, or reactions giving rise to two particles. It is shown under very general conditions that all elements of the S matrix can be simply obtained from a single analytic function of all channel momenta, the Fredholm determinant of the scattering and reaction integral equations. Its properties are investigated and the restrictions are established which are necessary and sufficient in order to assure that the unitarity condition is fulfilled. The square well and a superposition of Yukawa potentials are considered as examples.

1. INTRODUCTION

FOR elastic nonrelativistic scattering of particles interacting via local potentials of various degrees of generality, the analytic properties of the S matrix as a function of the energy, its regularity or lack of it in the complex plane, are by now rather well understood.[1] The same cannot be said for S matrices describing inelastic processes, i.e., reactions as well. Quite apart from the characteristic branch points introduced by conservation of energy between channels, there are difficulties in continuing the known functions in terms of which the S matrix can be simply expressed, even to real energies below the highest threshold.

A recent paper[2] by LeCouteur made an important contribution which, it turns out, much facilitates the understanding of such analytic properties. In the very special case of an S-matrix meromorphic in all the channel momenta, he showed that there exists a single function from which all of its elements can be obtained in a simple way. The assumption that, except for poles, the S matrix is regular everywhere in the complex plane of all channel momenta plays an important role in his demonstration; it can, therefore, not be assumed without proof to hold in a more general case. Nor is there any indication in his work of the general properties of the function or of its connection with the interparticle forces.

In the present paper the function of LeCouteur is exhibited under extremely general conditions. It is shown to be the determinant of the generalized Jost matrix function introduced earlier by the author.[3] Unfortunately, that matrix function does not have simple regularity properties unless the potential is of a very restricted class. It is, therefore, significant that it is also shown that its *determinant* is equal to the Fredholm determinant of the set of coupled integral equations for the scattering wave function; under very general conditions it is, therefore, a regular analytic function in the whole upper half of the complex plane of each channel momentum. Furthermore, its zeros give directly the bound states, and in somewhat more restricted circumstances, the resonances.

The existence of a single analytic function of all channel momenta underlying the whole scattering and reaction matrix, diagonal elements as well as off-diagonal, is interesting for a number of reasons. Not only does it allow an insight into the structure of the S matrix under much weaker assumptions than heretofore necessary, but it may also be a useful tool where the unitarity condition makes a direct use of the S matrix cumbersome. It is true that for more than two channels the equation for the Fredholm determinant equivalent to unitarity appears as a rather complicated functional restriction; but perhaps it will be possible in the future to understand the implications of this restriction a little better and thus to come to a better understanding of the functional nature of the unitarity condition.

In Sec. 2 we briefly review the handling of the many channel problem introduced in footnote 3. We restrict ourselves to discrete channels, i.e., containing no more than two particles. The generalized Jost matrix function is introduced and all elements of the S matrix are expressed simply in terms of its determinant. Section 3 deals with the Fredholm method. A recursion procedure is introduced which allows a straightforward generalization to coupled equations. It is then shown that the determinant used in the previous section is identical with the Fredholm determinant of the set of coupled scattering and reaction integral equations. In Sec. 4 the properties of this Fredholm determinant are exhibited. The necessary and sufficient restrictions are derived which take the place of the unitarity of the open channel part of the S matrix. Finally the zeros in the complex plane are related to bound states and resonances. Section 5 contains two examples: The square-well potential, and the continuous superposition of Yukawa potentials.

2. SCHRÖDINGER EQUATION AND S MATRIX

We briefly outline the procedure of footnote 3 in somewhat simplified notation.[4] The starting point is a

* Supported in part by the National Science Foundation.

† A brief summary of this paper was presented at the Tenth Annual International Conference on High Energy Physics, Rochester, New York, August 25–September 1, 1960.

[1] See, for example, the recent review article by R. G. Newton, J. Math. Phys. 1, 319 (1960).

[2] K. J. LeCouteur, Proc. Roy. Soc. (London) **A256**, 115 (1960).

[3] R. G. Newton, Ann. Phys. 4, 29 (1958).

[4] In order to conform to the more customary notation adopted here, all matrix equations of footnote 3 must be read from right to left.

set of coupled radial Schrödinger equations

$$-\tfrac{1}{2}\hbar^2 m_\alpha^{-1}\Psi_\alpha''+\sum_{\beta=1}^{n} U_{\alpha\beta}\Psi_\beta=\mathcal{E}_\alpha\Psi_\alpha$$

for the channel components Ψ_α of the time independent wave function. The "potential" matrix[5] $U_{\alpha\beta}=U_{\beta\alpha}$ contains the (diagonal) centrifugal terms; $\mathcal{E}_\alpha$ are the channel energies which differ from one another by fixed given amounts; and m_α are the channel (reduced) masses. If we introduce new wave function components

$$\psi_\alpha\equiv m_\alpha^{-\frac{1}{2}}\Psi_\alpha$$

and write

$$V_{\alpha\beta}\equiv 2\hbar^{-2}m_\alpha^{\frac{1}{2}}U_{\alpha\beta}m_\beta^{\frac{1}{2}},$$

then the Schrödinger equations read

$$-\psi_\alpha''+\textstyle\sum_\beta V_{\alpha\beta}\psi_\beta=k_\alpha^2\psi_\alpha, \tag{2.1}$$

where

$$k_\alpha=(2m_\alpha\mathcal{E}_\alpha)^{\frac{1}{2}}/\hbar$$

are the channel wave numbers.

We combine n different column wave functions $\{\psi_\alpha\}$, $\alpha=1,\cdots,n$, into a square matrix $\{\psi_{\alpha\beta}\}$, $\alpha,\beta=1,\cdots,n$, and then write (2.1) in matrix notation

$$-\psi''+V\psi=K^2\psi, \tag{2.1'}$$

K being the diagonal matrix of the channel wave numbers. Each column of ψ then solves (2.1) and the columns differ from one another by their boundary condition.

For the sake of simplicity, we shall restrict ourselves to s waves only. The case of higher l values and couplings between them does not present any difficulties in principle but it introduces sometimes bothersome complications. Furthermore, we assume that all elements of V are local, i.e., functions of r, and energy independent.

The solution[6] $F(K,r)=F(k_1,\cdots;r)$ is defined by the boundary condition

$$\lim_{r\to\infty} e^{iKr}F(K,r)=1, \tag{2.2}$$

i.e., more explicitly and less precisely,

$$F_{\alpha\beta}(K,r)\underset{r\to\infty}{\sim}\delta_{\alpha\beta}e^{-ik_\beta r}. \tag{2.2'}$$

Thus the β column has only an incoming wave in the β channel and no particles at infinity in any other channel. Such a solution is in general irregular at the origin. A regular function $\phi(K,r)$ is defined by the boundary condition

$$\phi(K,0)=0,\quad \phi'(K,0)=1. \tag{2.3}$$

It can be expressed in terms of the two linearly independent solutions $F(K,r)$ and

$$F(-K,r)=F(-k_1,-k_2,\cdots;r)^{7}:$$

$$\phi(K,r)=(i/2)[F(K,r)K^{-1}F^T(-K)-F(-K,r)K^{-1}F^T(K)]. \tag{2.4}$$

The matrix function $F(K)$ is obtained by taking the Wronskian matrix of $F(K,r)$ and $\phi(K,r)$:

$$F^T(K)\equiv F^T(K,r)\phi'(K,r)-F^{T\prime}(K,r)\phi(K,r)=F^T(K,0), \tag{2.5}$$

which is independent of r by virtue of the differential equation (2.1') and the symmetry of V.

The modified scattering matrix is obtained from (2.4) together with the boundary condition (2.2):

$$S'(K)=K^{-1}F^T(K)F^{T-1}(-K)K=F^{-1}(-K)F(K). \tag{2.6}$$

The last line follows from the boundary condition (2.3) inserted in (2.4). The symmetric S matrix whose open channel submatrix is unitary, is related to (2.6) by[8]

$$S(K)=K^{\frac{1}{2}}S'K^{\frac{1}{2}}. \tag{2.7}$$

In order to discuss the properties of the relevant functions, it is convenient for the time being to disregard the restrictions on the channel momenta imposed by the fixed differences between the various channel energies. In other words, we regard all the k's as independent variables.

It is clear from the differential equation (2.1') and the boundary condition (2.3) that $\phi(k_1,\cdots;r)$ is an even function of all the k's. It is the unique solution of the matrix integral equation

$$\phi(K,r)=K^{-1}\sin Kr+\int_0^r dr'K^{-1}\sin K(r-r')\times V(r')\phi(K,r'), \tag{2.8}$$

which is always solvable by successive approximations provided only that the first absolute moments of all elements of $V(r)$ are finite.[9] It is then straightforward to show that $\phi(k_1,\cdots;r)$ is an analytic function of all the k's regular in the entire complex plane.[9]

The function $F(K,r)$, on the other hand, has in general much less regularity. It follows from the boundary condition (2.2') that $F_{\alpha\beta}(k_1,\cdots;r)$ *is an even*

[5] The symmetry of the potential matrix V entails the symmetry of the S matrix, i.e., the reciprocity theorem. Both follow for a suitable choice of phases of the angular momentum functions from an assumed time reversal invariance of the interaction hamiltonian; see, e.g., footnote 1.

[6] The letter K stands both for the diagonal matrix of channel wave numbers and for the set of k's. Sometimes we make things more explicit by writing instead $(k_1,\cdots)$.

[7] The superscript "T" indicates the transposed, "$*$" the complex conjugate, and "$\dagger$" the Hermitian conjugate.

[8] The scattering (or reaction) amplitude is obtained from S by Eq. (2.15) of footnote 1, where now, of course, S and Θ have additional channel subscripts and $k\neq k'$. The cross section is still directly the square modulus of Θ.

[9] The proof is no different from that in footnote 1 for the one channel case.

function of all k's, except of k_β; by (2.5) the same then is true of $F_{\alpha\beta}(k_1,\cdots)$. $F(K,r)$ is the solution of the integral equation

$$F(K,r)=e^{-iKr}-\int_r^\infty dr'K^{-1}\sin K(r-r')\times V(r')F(K,r'), \quad (2.9)$$

which is solvable by successive approximations provided only that the first absolute moments of V are finite.[9] If the second absolute moments of all elements of V also exist, then it is straight forward to show[9] that $F_{\alpha\beta}(k_1,\cdots;r)$ is an analytic function of k_β regular in the lower half of the complex plane, but as a function of the other k's it has no general regularity properties unless much stronger assumptions are made concerning the potential. In other words, we cannot even continue all elements of $F(K,r)$ to values of $k_1, \cdots, k_n$ which correspond to a real energy below the highest threshold, where one or more of the k's are imaginary. The same then holds for $F(K)$, which by (2.5) and (2.9) can be written

$$F(K)=1+\int_0^\infty drK^{-1}\sin KrV(r)F(K,r). \quad (2.10)$$

It is a remarkable fact that in spite of this lack of regularity of the elements of $F(K)$ the determinant

$$f(K)\equiv f(k_1,\cdots)\equiv \det F(-K) \quad (2.11)$$

is an analytic function of all k's regular in the whole upper half of the complex plane. Thus, although as soon as one of the k's leaves the real axis, singularities of some elements of $F(K)$ can in general not be ruled out, such singularities conspire to make the combination of elements that make up the determinant, regular in the lower half-plane. The proof of this is given in Sec. 3.

Let us now rewrite (2.6) in the way appropriate to the formation of the inverse:

$$F^{-1}(-K)=X(-K)/f(K), \quad (2.12)$$

$$S'(K)=X(-K)F(K)/f(K). \quad (2.13)$$

Thus $X(K)$ is the transposed of the matrix made up of the co-factors of $F(K)$. It follows that $X_{\alpha\beta}(k_1,\cdots)$ *is an even function of* k_α.

The diagonal elements of S can by (2.13) be written explicitly by developing the determinant

$$S_{\alpha\alpha}=S_{\alpha\alpha}'=\frac{\sum_\gamma X_{\alpha\gamma}(-K)F_{\gamma\alpha}(K)}{\sum_\gamma X_{\alpha\gamma}(-K)F_{\gamma\alpha}(-K)}$$

$$=\frac{\sum_\gamma X_{\alpha\gamma}(-k_1,\cdots,k_\alpha,\cdots)F_{\gamma\alpha}(-k_1,\cdots,k_\alpha,\cdots)}{\sum_\gamma X_{\alpha\gamma}(-k_1,\cdots)F_{\gamma\alpha}(-k_1,\cdots)}$$

or

$$S_{\alpha\alpha}=f(k_1,\cdots,-k_\alpha,\cdots)/f(k_1,\cdots)$$
$$=1/S_{\alpha\alpha}(k_1,\cdots,-k_\alpha,\cdots). \quad (2.14)$$

Thus all diagonal elements of S are obtainable from the function $f(k_1,\cdots)$.

Before proving a similar result for the off-diagonal elements, we derive a number of general symmetry properties of the S matrix.

Suppose we define, for $\alpha\neq\beta$,

$$f_{\alpha\beta}(K)\equiv\sum_\gamma X_{\alpha\gamma}(-K)F_{\gamma\beta}(K)k_\beta^{-1}, \quad (2.15)$$

so that

$$S_{\alpha\beta}'=k_\beta f_{\alpha\beta}/f \quad (2.16)$$

and

$$f_{\alpha\beta}=f_{\beta\alpha} \quad (2.17)$$

expresses the symmetry of the S matrix. $X_{\alpha\beta}$ being an even function of k_α, it follows that $f_{\alpha\beta}$ *is an even function of* k_α *and* k_β. Consequently we have by (2.14) for $\alpha\neq\beta$,

$$S_{\alpha\alpha}'S_{\alpha\beta}'(k_1,\cdots,-k_\alpha,\cdots)=S_{\alpha\beta}', \quad (2.18a)$$

$$S_{\beta\alpha}'S_{\alpha\alpha}'(k_1,\cdots,-k_\alpha,\cdots)=-S_{\beta\alpha}'(k_1,\cdots,-k_\alpha,\cdots). \quad (2.18b)$$

Furthermore, we evaluate, for $\alpha\neq\beta$, $\alpha\neq\gamma$,

$$[X(-K)F(K)f(k_1,\cdots,-k_\alpha,\cdots)-X(-k_1,\cdots,k_\alpha,\cdots)\times F(k_1,\cdots,-k_\alpha,\cdots)f(k_1,\cdots)]_{\beta\gamma}$$

once by inserting

$$\mathbf{1}f(K)=F(-K)X(-K)=X(-K)F(-K)$$

in the middle and once on the left:

$$[\ ]_{\beta\gamma}=[X(-K)F(-k_1,\cdots,k_\alpha,\cdots)\times X(-k_1,\cdots,k_\alpha,\cdots)F(K)]_{\beta\gamma}$$
$$-[X(-K)F(-K)X(-k_1,\cdots,k_\alpha,\cdots)\times F(k_1,\cdots,-k_\alpha,\cdots)]_{\beta\gamma}$$
$$=[X(-K)F(K)]_{\beta\alpha}[X(-K)F(K)]_{\alpha\gamma}=k_\alpha k_\gamma f_{\beta\alpha}f_{\alpha\gamma},$$

which implies that

$$S_{\beta\alpha}'S_{\alpha\gamma}'(k_1,\cdots,-k_\alpha,\cdots)=S_{\beta\gamma}'-S_{\beta\gamma}'(k_1,\cdots,-k_\alpha,\cdots). \quad (2.19)$$

Equations (2.14), (2.18), and (2.19) can be written compactly

$$S'P^{(\alpha)}S'(k_1,\cdots,-k_\alpha,\cdots)=P^{(\alpha)}+S'Q^{(\alpha)}-Q^{(\alpha)}S'(k_1,\cdots,-k_\alpha,\cdots), \quad (2.20)$$

where $P^{(\alpha)}$ is the projection on the α channel, and $Q^{(\alpha)}=1-P^{(\alpha)}$. This equation contains its own generalization. If we write $P^{(\alpha\beta\cdots)}$ for the projection on the $(\alpha,\beta,\cdots)$, channels, then it is a straightforward algebraic exercise to show by repeated use of (2.20) and its version for $k_\alpha\to-k_\alpha$ that

$$S'P^{(\alpha\beta\cdots)}S'(k_1,\cdots,-k_\alpha,-k_\beta,-\cdots)=P^{(\alpha\beta\cdots)}+S'Q^{(\alpha\beta\cdots)}-Q^{(\alpha\beta\cdots)}S'(k_1,\cdots,-k_\alpha,-k_\beta,-\cdots). \quad (2.21)$$

The most important special case is obtained by multiplying it on both sides by $P^{(\alpha\beta\cdots)}$:

$$P^{(\alpha\beta\cdots)}S'P^{(\alpha\beta\cdots)}S'(k_1,\cdots,-k_\alpha,-k_\beta,-\cdots)\times P^{(\alpha\beta\cdots)}=P^{(\alpha\beta\cdots)}, \quad (2.22)$$

which says that the $(\alpha,\beta,\cdots)$ submatrix of

$$S'(k_1,\cdots,-k_\alpha,-k_\beta,-\cdots)$$

is the inverse of the $(\alpha,\beta,\cdots)$ submatrix of S'.[10]

We now look back at (2.20) or its detailed statements (2.14), (2.18), and (2.19). On inserting (2.16) and (2.14) in (2.19) for $\beta=\gamma$ and using the evenness of $f_{\alpha\beta}$ as a function of k_α and k_β, we obtain the equation

$$[f_{\alpha\beta}(k_1,\cdots)]^2 =[f(k_1,\cdots,-k_\alpha,\cdots)f(k_1,\cdots,-k_\beta,\cdots) -f(k_1,\cdots)f(k_1,\cdots,-k_\alpha,-k_\beta,\cdots)]/k_\alpha k_\beta \quad (2.23)$$

or, equivalently,

$$S_{\alpha\beta}{}^2=S_{\alpha\alpha}S_{\beta\beta}-f(k_1,\cdots,-k_\alpha,-k_\beta,\cdots)/f(k_1,\cdots). \quad (2.23')$$

Equations (2.14) and (2.23) completely determine the S matrix from the function $f(k_1,\cdots)$. [The sign ambiguity inherent in (2.23) is of no consequence.]

3. FREDHOLM METHOD

Before we can prove that the function $f(k_1,\cdots)$ of (2.11) is equal to the Fredholm determinant of the scattering integral equations, it is necessary to generalize the Fredholm method to coupled equations, i.e., matrix integral equations. In order to facilitate matters, we shall use a general matrix notation also for the "continuous indices," i.e., for the arguments of the integral kernels. The formal development will hold for finite matrices as well as for infinite, discrete or continuous matrices provided certain convergence conditions are fulfilled.

We want to find the inverse of the matrix

$$M=\mathbf{1}-\alpha R, \quad (3.1)$$

where α is a parameter introduced for convenience. We form the inverse in the familiar way by constructing the matrix N which is the transposed of the (signed) co-factors of M, and divide by the determinant:

$$NM=MN=\mathbf{1}\det M\equiv\mathbf{1}\Delta. \quad (3.2)$$

Both N and Δ are expanded in a power series in α:

$$N=\sum_0^\infty \alpha^n N^{(n)}, \quad N^{(0)}=\mathbf{1}; \quad (3.3)$$

$$\Delta=\sum_0^\infty \alpha^n \Delta^{(n)}, \quad \Delta^{(0)}=\mathbf{1}. \quad (3.4)$$

We then use the well-known differentiation rule for the determinant[11]:

$$(d/d\alpha)\Delta=\mathrm{Tr}\,N(d/d\alpha)M=-\mathrm{Tr}\,NR \quad (3.5)$$

and evaluate the left-hand side by (3.2):

$$-\mathbf{1}\,\mathrm{Tr}\,NR=-RN+(\mathbf{1}-\alpha R)(d/d\alpha)N =-NR+[(d/d\alpha)N](\mathbf{1}-\alpha R).$$

On inserting the expansion (3.3), we obtain the recursion relation

$$N^{(n)}=RN^{(n-1)}-(1/n)\mathbf{1}\,\mathrm{Tr}\,RN^{(n-1)} =N^{(n-1)}R-(1/n)\mathbf{1}\,\mathrm{Tr}\,RN^{(n-1)}, \quad (3.6)$$

while (3.4) substituted in (3.5) leads to[12]

$$\Delta=1-\sum_1^\infty \frac{\alpha^n}{n}\,\mathrm{Tr}\,RN^{(n-1)}. \quad (3.7)$$

For infinite dimensional matrices, the expansions (3.3) and (3.7) together with the recursion (3.6) constitute the Fredholm method for the formation of the inverse

$$M^{-1}=N/\Delta. \quad (3.8)$$

Usually it is more convenient to write this in the form

$$M^{-1}=\mathbf{1}+(\alpha/\Delta)NR=\mathbf{1}+(\alpha/\Delta)RN\equiv 1+(\alpha/\Delta)Y, \quad (3.9)$$

which is readily demonstrated by the use of (3.3), (3.6), and (3.7). In that case

$$Y=\sum_0^\infty \alpha^n Y^{(n)} \quad (3.10)$$

and the $Y^{(n)}$ are determined by the recursion

$$Y^{(n)}=RY^{(n-1)}-(1/n)R\,\mathrm{Tr}\,Y^{(n-1)} =Y^{(n-1)}R-(1/n)R\,\mathrm{Tr}\,Y^{(n-1)}, \quad Y^{(0)}=R, \quad (3.11)$$

while

$$\Delta=1-\sum_1^\infty \frac{\alpha^n}{n}\,\mathrm{Tr}\,Y^{(n-1)}. \quad (3.12)$$

The foregoing procedure is used to solve the equations

$$\Psi=\Psi_0+\alpha R\Psi$$

or

$$M\Psi=\Psi_0.$$

[10] This is Peierls' version of the unitarity condition; see R. E. Peierls, Proc. Roy. Soc. (London) **A253**, 16 (1959).

[11] "Tr" stands for the trace of the matrix, including the continuous indices.

[12] If M has the finite dimensionality D then, of course, the power series (3.3) and (3.4) must break up and reduce to polynomials of degree $D-1$ and D, respectively. It is readily checked by (3.6) that a necessary condition for $N^{(m+1)}$ to vanish is indeed that $D=m+1$. Since this condition must also be sufficient we find that

$$RN^{(D-1)}=\mathbf{1}[\mathrm{Tr}\,RN^{(D-1)}]/D$$

or

$$R^{-1}=N^{(D-1)}[D/\mathrm{Tr}\,RN^{(D-1)}],$$

which is a simple recursion method of constructing the inverse of a D-dimensional matrix.

Equation (3.9) then states that the solution is

$$\Psi=\Psi_0+(\alpha/\Delta)Y\Psi_0.$$

At this point we may revert to a more explicit notation. Let R be a finite matrix of integral kernels. Then (suppressing the finite matrix indices)

$$\Psi(r)=\Psi_0(r)+\alpha\int_0^\infty dr'R(r,r')\Psi(r')$$

is solved by

$$\Psi(r)=\Psi_0(r)+\frac{\alpha}{\Delta}\int_0^\infty dr'Y(r,r')\Psi_0(r'),$$

where

$$Y(r,r')=\sum_0^\infty \alpha^n Y^{(n)}(r,r'),$$

$$\Delta=1-\sum_1^\infty \frac{\alpha^n}{n}\,\mathrm{tr}\int_0^\infty dr Y^{(n-1)}(r,r),$$

$$\begin{aligned}Y^{(n)}(r,r')&=\int_0^\infty dr''R(r,r'')Y^{(n-1)}(r'',r')\\&\quad-\frac{1}{n}R(r,r')\,\mathrm{tr}\int_0^\infty dr''Y^{(n-1)}(r'',r'')\\&=\int_0^\infty dr''Y^{(n-1)}(r,r'')R(r'',r')\\&\quad-\frac{1}{n}R(r,r')\,\mathrm{tr}\int_0^\infty dr''Y^{(n-1)}(r'',r''),\end{aligned}$$

$$Y^{(0)}(r,r')=R(r,r').$$

and "tr" now refers to the trace only over the finite indices.

For the type of integral equation arising in scattering theory we have

$$R(r,r')=\mathcal{G}(r,r')V(r'),$$

where $\mathcal{G}(r,r')$ is a "zero order" outgoing-wave Green's function. It is then customary to write the solution of the integral equation

$$\Psi(r)=\Psi_0(r)+\alpha\int_0^\infty dr'\mathcal{G}(r,r')V(r')\Psi(r')$$

in terms of a resolvent or "complete Green's function":

$$\Psi(r)=\Psi_0(r)+\alpha\int_0^\infty dr'\mathfrak{G}(r,r')V(r')\Psi_0(r').$$

Our result then is that

$$\mathfrak{G}(r,r')=\mathfrak{z}(r,r')/\Delta, \tag{3.13}$$

where both $\mathfrak{z}$ and Δ are obtained as power series expansions in α (the "potential strength"):

$$\mathfrak{z}(r,r')=\sum_0^\infty \alpha^n\mathfrak{z}^{(n)}(r,r') \tag{3.14}$$

$$\Delta=1-\sum_1^\infty \frac{\alpha^n}{n}\,\mathrm{tr}\int_0^\infty dr\,\mathfrak{z}^{(n-1)}(r,r)V(r), \tag{3.15}$$

and the $\mathfrak{z}^{(n)}$ are determined by the recursion

$$\begin{aligned}\mathfrak{z}^{(n)}(r,r')&=\int_0^\infty dr''\mathcal{G}(r,r'')V(r'')\mathfrak{z}^{(n-1)}(r'',r')\\&\quad-\frac{1}{n}\mathcal{G}(r,r')\,\mathrm{tr}\int_0^\infty dr''\mathfrak{z}^{(n-1)}(r'',r'')V(r'')\\&=\int_0^\infty dr''\mathfrak{z}^{(n-1)}(r,r'')V(r'')\mathcal{G}(r'',r')\\&\quad-\frac{1}{n}\mathcal{G}(r,r')\,\mathrm{tr}\int_0^\infty dr''\mathfrak{z}^{(n-1)}(r'',r'')V(r''),\end{aligned} \tag{3.16}$$

$$\mathfrak{z}^{(0)}(r,r')=\mathcal{G}(r,r').$$

There remains the problem of proving the convergence of the series in α. The recursion method for obtaining the Fredholm resolvent and determinant outlined above is more convenient and more manageable in the matrix case than is the customary way of writing them. However, when it comes to the convergence proof then the usual explicit solution of the recursion is much the more powerful because it allows the use of Hadamard's lemma.[13]

We shall write the solution of the recursion (3.11) for the case of discrete indices. The continuum case differs then only in an obvious notational way. The claim is that

$$Y_{\alpha\beta}{}^{(n)}=\frac{(-)^n}{n!}\sum_{\alpha_1,\cdots,\alpha_n}\begin{vmatrix}R_{\alpha\beta} & R_{\alpha\alpha_1} & \cdots & R_{\alpha\alpha_n}\\ R_{\alpha_1\beta} & R_{\alpha_1\alpha_1} & \cdots & R_{\alpha_1\alpha_n}\\ \vdots & & & \vdots\\ R_{\alpha_n\beta} & R_{\alpha_n\alpha_1} & \cdots & R_{\alpha_n\alpha_n}\end{vmatrix} \tag{3.17}$$

solves the recursion (3.11). The proof is immediate since the development of the determinant in (3.17) according to the first row yields

$$\begin{aligned}Y_{\alpha\beta}{}^{(n)}=\frac{(-)^n}{n!}\sum_{\alpha_1,\cdots,\alpha_n}\Bigg\{&R_{\alpha\beta}\begin{vmatrix}R_{\alpha_1\alpha_1} & \cdots & R_{\alpha_1\alpha_n}\\ \vdots & & \vdots\\ R_{\alpha_n\alpha_1} & \cdots & R_{\alpha_n\alpha_n}\end{vmatrix}\\&-R_{\alpha\alpha_1}\begin{vmatrix}R_{\alpha_1\beta} & R_{\alpha_1\alpha_2} & \cdots\\ \vdots & & \\ R_{\alpha_n\beta} & \cdots & \end{vmatrix}\\&+R_{\alpha\alpha_2}\begin{vmatrix}R_{\alpha_1\beta} & R_{\alpha_1\alpha_1} & \cdots\\ \vdots & & \\ R_{\alpha_n\beta} & \cdots & \end{vmatrix}+-\cdots\Bigg\}\\=&-(1/n)R_{\alpha\beta}\,\mathrm{Tr}\,Y^{(n-1)}+[RY^{(n-1)}]_{\alpha\beta}.\end{aligned}$$

In the case of matrix integral equations, each index in (3.17) comprizes two (or more) numbers, one of which runs over a discrete and the other over a continuous

[13] See, for example, E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (The MacMillan Company, New York, 1948), Chap. XI.

range. That is why that way of writing the Fredholm solution becomes cumbersome. The convergence proofs now start from (3.17) in the standard manner.[13,14]

The matrix integral equation for the physical s-wave scattering wave function is

$$\psi(K,r)=K^{-1}\sin Kr + \alpha\int_0^\infty dr'\mathcal{G}(K;r,r')V(r')\psi(K,r'), \quad (3.18)$$

where

$$\mathcal{G}(K;r,r')=-K^{-1}\sin Kr_< e^{iKr_>}. \quad (3.19)$$

If we take any k into the upper half of the complex plane then $\mathcal{G}$ never gives rise to exponential increase as $r'\to\infty$. As a result the usual convergence proofs of the Fredholm theory[13,14] apply there, and it is straightforward to see that both $\mathfrak{d}(K;r,r')$ and $\Delta(K)$ of (3.13) are analytic functions of all k's regular in the whole upper half of each complex plane.[15] It follows from a comparison of the incoming wave part of the asymptotic form of (3.18) and (2.4) that

$$\psi(K,r)=\phi(K,r)[F^T(-K)]^{-1}. \quad (3.20)$$

A further function needed is $G(K,r)$ which solves

$$G(K,r)=e^{iKr}+\alpha\int_0^\infty dr'\mathcal{G}(K;r,r')V(r')G(K,r'). \quad (3.21)$$

Since $G(K,0)=1$ and asymptotically $G(K,r)$ has no incoming waves, we have

$$G(K,r)=F(-K,r)F^{-1}(-K). \quad (3.22)$$

Multiplication of (2.10) by $F^{-1}(K)$ on the right, therefore, yields

$$F^{-1}(-K)=1-\alpha\int_0^\infty drK^{-1}\sin KrV(r)G(K,r). \quad (3.23)$$

We now differentiate G with respect to α. By (3.21)

$$(d/d\alpha)G=\mathcal{G}VG+\alpha\mathcal{G}V(d/d\alpha)G,$$

the solution of which is

$$(d/d\alpha)G=\mathfrak{G}VG, \quad (3.24)$$

since

$$\mathfrak{G}=\mathcal{G}+\alpha\mathfrak{G}V\mathcal{G}=\mathcal{G}+\alpha\mathcal{G}V\mathfrak{G} \quad (3.25)$$

is the integral equation for the complete Green's function $\mathfrak{G}$. Next we differentiate (3.23) with respect to α and insert (3.24):

$$(d/d\alpha)F^{-1}(-K)=-[\psi_0+\alpha\psi_0V\mathfrak{G}]VG,$$

where

$$\psi_0=K^{-1}\sin Kr.$$

A glance at (3.18) shows, therefore, that

$$(d/d\alpha)F^{-1}(-K)=-\int_0^\infty dr\psi^T(K,r)V(r)G(K,r)$$

and consequently, by (3.20),

$$\operatorname{tr}F(-K)(d/d\alpha)F^{-1}(-K) = -\operatorname{tr}\int_0^\infty drG(K,r)\phi^T(K,r)V(r). \quad (3.26)$$

Now the complete Green's function $\mathfrak{G}$ can be represented as[16]

$$\mathfrak{G}(K;r,r')=\begin{cases}-\phi(K,r)G^T(K,r'), & r<r',\\ -G(K,r)\phi^T(K,r'), & r>r'.\end{cases} \quad (3.27)$$

Equation (3.26), therefore, says that

$$\operatorname{tr}F^{-1}(-K)(d/d\alpha)F(-K)=-\operatorname{tr}\int_0^\infty dr\,\mathfrak{G}(K;r,r)V(r) = -\operatorname{Tr}\mathfrak{G}V.$$

As a result we get for the derivative of the determinant of $F(-K)$

$$(d/d\alpha)f(K)/f(K)=\operatorname{tr}F^{-1}(K)(d/d\alpha)F(-K) = -\operatorname{Tr}\mathfrak{G}V,$$

or if we define

$$Z(K;r,r')\equiv f(K)\mathfrak{G}(K;r,r') \quad (3.28)$$

then

$$(d/d\alpha)f(K)=-\operatorname{Tr}ZV. \quad (3.29)$$

We now want to show that the function Z of (3.28) is equal to $\mathfrak{d}$ of (3.13). We know that both $f(K)$ and $Z(K;r,r')$ can be expanded in convergent power series in α:

$$f(K)=\sum_0^\infty\alpha^n f^{(n)}(K),\quad f^{(0)}=1;$$

$$Z(K;r,r')=\sum_0^\infty\alpha^nZ^{(n)}(K;r,r');$$

$$Z^{(0)}(K;r,r')=\mathcal{G}(K;r,r').$$

Equation (3.29), therefore, states that

$$f^{(n)}=-(1/n)\operatorname{Tr}Z^{(n-1)}V.$$

On the other hand, multiplication of (3.25) by f and expansion in α yields

$$Z^{(n)}=\mathcal{G}VZ^{(n-1)}+\mathcal{G}f^{(n)} =\mathcal{G}VZ^{(n-1)}-\mathcal{G}(1/n)\operatorname{Tr}Z^{(n-1)}V.$$

Since that agrees with (3.16), we have $Z=\mathfrak{d}$ and hence

$$f(K)=\Delta(K). \quad (3.30)$$

[14] R. Jost and A. Pais, Phys. Rev. 82, 840 (1951).

[15] We always assume that all elements of V have finite first and second absolute moments.

[16] The proof of this is the same as for (9.22) of footnote 1.

We have, therefore, the important result that the function $f(K)$ in terms of which all elements of the S matrix can be expressed by (2.14) and (2.21), is identical with the Fredholm determinant of the set of coupled scattering integral equations (3.18).

4. PROPERTIES OF $f(K)$

It follows from the boundary condition (2.2) that when all k's are real,

$$F(-k_1, -k_2, \cdots; r)=F^*(k_1,k_2,\cdots; r)$$

and, $\phi(K,r)$ being even in all k's and real for real k's,

$$F(-k_1, -k_2, \cdots)=F^*(k_1,k_2,\cdots); \tag{4.1}$$

consequently

$$f(-k_1, -k_2, \cdots)=f^*(k_1,k_2,\cdots).$$

This implies that in any region of analyticity connected with the real axis

$$f^*(-k_1^*, -k_2^*, \cdots)=f(k_1,k_2,\cdots). \tag{4.2}$$

Since the Fredholm determinant Δ is an analytic function of all k's regular in the whole upper half of each complex plane, so is f by (3.30), and (4.2) must hold in the upper half-plane.

A rather more complicated property of f follows from the off-diagonal elements of (2.19). On taking $\beta\neq\gamma$, we obtain

$$ff_{\beta\gamma}(k_1, \cdots, -k_\alpha, \cdots)=f(k_1, \cdots, -k_\alpha, \cdots)f_{\beta\gamma} - k_\alpha f_{\alpha\beta}f_{\alpha\gamma}. \tag{4.3}$$

On inserting (2.23) in this we obtain, after squaring, an explicit expression for $f(k_1, \cdots, -k_\alpha, -k_\beta, -k_\gamma, \cdots)$ in terms of f with only one or two k's changed in sign:

$$f^2g_{\alpha\beta\gamma}=f(g_\alpha g_{\beta\gamma}+g_\beta g_{\gamma\alpha}+g_\gamma g_{\alpha\beta})-2g_\alpha g_\beta g_\gamma + 2[(g_\alpha g_\beta-g_{\alpha\beta}f)(g_\beta g_\gamma-g_{\beta\gamma}f)(g_\gamma g_\alpha-g_{\gamma\alpha}f)]^{\frac{1}{2}}, \tag{4.4}$$

where

$$g_\alpha\equiv f(k_1, \cdots, -k_\alpha, \cdots),$$
$$g_{\alpha\beta}\equiv f(k_1, \cdots, -k_\alpha, -k_\beta, \cdots), \quad \text{etc.}$$

This equation, which comes into play only for three or more channels, together with (2.14) and (2.23), is equivalent to (2.21) and hence implies equations (2.22) for all choices of submatrices.

There remains the questions of the unitarity of the S matrix. Suppose that k_α, k_β, $\cdots k_\lambda$ are positive real and all other k's are positive imaginary. Then we can use (2.14) and (2.23) and (2.16) to form $S_{\delta\gamma}'$, where δ and γ are among the α, β, $\cdots\lambda$. Equation (2.22) then implies that

$$P^{(\alpha\beta\cdots)}S'P^{(\alpha\beta\cdots)}S'^*P^{(\alpha\beta\cdots)}=P^{(\alpha\beta\cdots)}, \tag{4.5}$$

and consequently by (4.7) and its symmetry, that the open-channel submatrix of S is unitary:

$$P^{(\alpha\beta\cdots)}SP^{(\alpha\beta\cdots)}S^\dagger P^{(\alpha\beta\cdots)}=P^{(\alpha\beta\cdots)}, \tag{4.6}$$

provided that

$$S_{\delta\gamma}'(-K^*)=S_{\delta\gamma}'(K). \tag{4.7}$$

Of course, this property follows from (4.1); but if we construct the S matrix via (2.14) and (2.23) we must be sure that it is indeed satisfied. Because of (4.2), Eq. (2.16) shows that (4.7) holds if

$$f_{\delta\gamma}^*(-K^*)=-f_{\delta\gamma}(K). \tag{4.8}$$

It follows from (2.23) and (4.2) that

$$[f_{\delta\gamma}(-K^*)]^{2*}=[f_{\delta\gamma}(K)]^2;$$

but (4.8) is more stringent.

It is clear from the unitarity (4.6) that (2.14) implies the inequality

$$|f(k_1, \cdots, -k_\alpha, \cdots)|\leqslant|f(k_1, \cdots)| \tag{4.9}$$

for real positive k_α and all other k's either positive or positive imaginary. Suppose we take k_α and k_β positive and all other k's positive imaginary; then (4.2) and (2.23) lead to

$$k_\alpha k_\beta f_{\alpha\beta}^2=|f(k_1, \cdots, -k_\alpha, \cdots)|^2-|f(k_1,\cdots)|^2,$$

which by (4.9) is *negative*. As a result (4.8) is indeed fulfilled and the unitarity condition holds. In other words, (4.9) implies (4.8) for k_δ and k_γ real and all other k's positive imaginary. Next we move k_σ off the imaginary axis. Both sides of (4.8) being analytic functions, the equation must continue to hold. It must, therefore, hold also when k_σ is real. This operation may be repeated. As a result the set of inequalities (4.9) is all that is necessary in order to assure (4.8) and hence the unitarity condition.

A final property of $f(k_1,\cdots)$ is readily shown by the Fredholm procedure.[15] If all k's are kept either on the real axis or in the upper half of the complex plane, then

$$\lim_{|k_1|,|k_2|,\cdots\to\infty} f(k_1,\cdots)=1. \tag{4.10}$$

These are all the restrictions on f. Equations (4.2) and (4.4), and the inequalities (4.9) are necessary and sufficient conditions (together with the regularity property of f) in order that (2.14) and (2.23) lead to a symmetric and unitary S matrix. Equation (4.10) is an additional property that leads to $S=1$ at infinite energy.

Equations (2.14) and (2.23) allow in general the construction of the open-channel part of the S matrix only. In other words, $S_{\alpha\beta}$, $S_{\alpha\alpha}$, and $S_{\beta\beta}$ are in general well defined only if k_α and k_β are real and all other k's are either real or positive imaginary. In order to define S in a larger domain, f would have to be continued to a region where some k's are in the lower half plane. There we know nothing about its behavior. However, if the potential matrix V is known to satisfy more stringent conditions than assumed so far, then f may be shown to have a larger region of regularity. The situation is in that respect just as in the simpler case of a single channel.[1] If all elements of V vanish at infinity more

strongly than $\exp(-2\mu r)$, then $f(k_1,\cdots)$ is regular also in a strip of width μ in the lower half-planes of all k's. If all elements of V vanish identically beyond a finite point, then $f(k_1,\cdots)$ is regular for all finite k_1, k_2, $\cdots$.

We now want to discuss the significance of the zeros of f in the upper half plane. For that purpose it is convenient to consider the function

$$H(K,r)\equiv F(-K,r)X(-K)=G(K,r)f(K),$$

which solves (2.1′) and the integral equation

$$H(K,r)=f(K)e^{iKr}+\int_0^\infty dr'\mathcal{G}(K;r,r')V(r')H(K,r'). \quad (4.11)$$

It can therefore be written

$$H(K,r)=f(K)e^{iKr}+\int_0^\infty dr'\mathfrak{z}(K;r,r')V(r')e^{iKr'}. \quad (4.12)$$

For fixed r, H is an analytic function of all k's regular in the upper half-plane[15] of each k. It is also uniformly (in r and all k's) bounded in the upper half plane of all k's.

Suppose then that

$$f(\kappa_1,\kappa_2,\cdots)=0$$

at a point $(\kappa_1,\cdots)$, where all κ's are either positive or in the upper half-plane. We can establish at once that this cannot happen when all κ's are positive, for then there must exist a set of numbers c_β so that

$$\sum_\beta F_{\gamma\beta}(-\mathrm{K})c_\beta=0 \quad \text{for all } \gamma$$

and by (4.1)

$$\sum_\beta c_\beta^*[F^T(\mathrm{K})]_{\beta\gamma}=0.$$

But we can easily evaluate the Wronskian matrix for $F(K,r)$ and $F(-K,r)$ at infinity and at zero:

$$F^T(K)F'(K,0)-F^{T\prime}(K,0)F(-K)=2iK,$$

multiplication of which on the left by $\{c_\beta^*\}$ and on the right by $\{c_\beta\}$ then yields

$$\sum_\beta|c_\beta|^2\kappa_\beta=0.$$

This is impossible if all κ's are positive.

Now, then, for $k_1=\kappa_1$, $k_2=\kappa_2$, $\cdots$, by (4.11) $H(K,r)$ vanishes at $r=0$; it is a regular wave function. As $r\to\infty$ we have the following situation: For[17] $\mathrm{Im}\,\kappa_\alpha>0$, we obtain from (4.11)

$$H_{\alpha\beta}\underset{r\to\infty}{\sim}\frac{1}{2i\kappa_\alpha}\left\{\int_0^r dr'e^{i\kappa_\alpha(r-r')}[V(r')H(\mathrm{K},r')]_{\alpha\beta}+\int_r^\infty dr'e^{i\kappa_\alpha(r'-r)}[V(r')H(\mathrm{K},r')]_{\alpha\beta}\right\}+O(e^{-\mathrm{Im}\,\kappa_\alpha r})$$

[17] "Re" and "Im" denote, respectively, the real and imaginary part of a number.

which is square integrable. For $\mathrm{Im}\,\kappa_\gamma=0$, we write

$$H_{\gamma\beta}=e^{i\kappa_\gamma r}X_{\gamma\beta}(-\mathrm{K})-\kappa_\gamma^{-1}\int_r^\infty dr'\sin\kappa_\gamma(r-r')[V(r')H(\mathrm{K},r')]_{\gamma\beta},$$

where by (2.10)

$$X_{\gamma\beta}(-K)=\delta_{\gamma\beta}f(K)-k_\gamma^{-1}\int_0^\infty dr\sin k_\gamma r[V(r)H(K,r)]_{\gamma\beta}$$

which exists for real k_γ and all other k's real or in the upper half-plane. Moreover, if k_γ is positive and all other k's are either positive or positive imaginary, and $f=0$, then $X_{\gamma\beta}(-K)=0$ for all β. That fact is easily shown by evaluating the Wronskian matrix for H and H^* once at $r=0$ and once at $r=\infty$. When all k^2 are real, we obtain

$$f^*(K)H'(K,0)-f(K)H^{\dagger\prime}(K,0)=-iX^\dagger(-K)RKX(-K),$$

R being the projection on the channels with *real* k's. Hence if $f(\mathrm{K})=0$ then

$$X^\dagger(-\mathrm{K})\mathrm{K}RX(-\mathrm{K})=0,$$

and consequently, if all the real κ's are positive, then

$$X_{\gamma\beta}(-\mathrm{K})=0$$

for all β.

In other words, when $f(\kappa_1,\cdots)=0$ at a point $(\kappa_1,\cdots)$ where each κ is either real or has a positive imaginary part, then there is not, in general, a bound state; but if that happens at a point where each κ is either positive or positive imaginary, then we have a bound state. If all κ's are positive imaginary then the bound state is of the type usually considered: all channels are closed. If some are real, then the bound state is "embedded in the continuum" a state of affairs discussed at length in a recent paper by Fonda and Newton.[18] Finally, if the forces that would lead to a bound state of the latter type are slightly altered, then the zero of f, with all but one, say, of the κ's fixed, will move off the imaginary axis and cause a resonance (in the absence of other disturbing nearby singularities of the S-matrix).[19]

5. EXAMPLES

Square Well

Assume that for $r<r_0$ the potential matrix V is constant and for $r>r_0$, it vanishes. We write (2.1′)

$$\psi''+A\psi=0,$$

[18] L. Fonda and R. G. Newton, Ann. Phys. **10**, 490 (1960).

[19] An analytic function of n variables vanishes on an $(n-1)$-dimensional complex hypersurface. Insertion of energy conservation among all the κ's generally eliminates the zero altogether after the perturbation. But if between two groups of κ's the energy conservation is relaxed, then the zero is retained and moves generally off the real axis. That is the state of affairs treated from a somewhat different point of view in footnote 18.

where

$$A=\begin{cases}K^2-V, & r<r_0,\\ K^2, & r>r_0.\end{cases}$$

We diagonalize the matrix A for $r<r_0$:

$$A=Ta^2T^{-1},$$

where a^2 is the diagonal matrix of the eigenvalues of A. The solution $F(K,r)$ must be of the form

$$F(K,r)=\begin{cases}T(e^{iar}B+e^{-iar}C), & r<r_0,\\ e^{-iKr}, & r>r_0.\end{cases}$$

The constant matrices B and C are determined by the requirement of continuity of $F(K,r)$ and of its first derivative. We find that for $r<r_0$

$$F(K,r)=[T\cos a(r-r_0)T^{-1}-iTa^{-1}\sin a(r-r_0)T^{-1}K]e^{-iKr_0},$$

and therefore,

$$F(K)=[T\cos ar_0T^{-1}+iTa^{-1}\sin ar_0T^{-1}K]e^{-iKr_0}$$
$$=Ta^{-1}\sin ar_0T^{-1}[Ta\cot ar_0T^{-1}+iK]e^{-iKr_0},$$

so that

$$f(K)=[\Pi_\alpha e^{ik_\alpha r_0}\Pi_i a_i^{-1}\sin a_i r_0]\times\det[Ta\cot ar_0T^{-1}-iK].$$

If we take the case of two channels and set

$$e_i\equiv a_ir_0,\quad \rho_\alpha\equiv k_\alpha r_0,$$
$$R_\alpha\equiv A_{\alpha\alpha}{}^{\frac12}r_0=(\rho_\alpha{}^2-V_{\alpha\alpha}r_0{}^2)^{\frac12},$$

then

$$e_{1,2}{}^2=\tfrac12(R_1{}^2+R_2{}^2)\pm\tfrac12[(R_1{}^2-R_2{}^2)^2+4V_{12}{}^2]^{\frac12}$$

and

$$f(k_1,k_2)=\frac{e^{i(\rho_1+\rho_2)}}{\alpha_2-\alpha_1}\cdot\frac{\sin e_1}{e_1}\cdot\frac{\sin e_2}{e_2}\cdot g(\rho_1,\rho_2),$$

where[20]

$$g(\rho_1,\rho_2)=i\rho_1(\mathcal{C}_1\alpha_1-\mathcal{C}_2\alpha_2)+i\rho_2(\mathcal{C}_2\alpha_1-\mathcal{C}_1\alpha_2)+(\alpha_1-\alpha_2)(\rho_1\rho_2-\mathcal{C}_1\mathcal{C}_2)$$

with

$$\alpha_1=e_1{}^2-R_1{}^2=R_2{}^2-e_2{}^2,\quad \alpha_2=e_2{}^2-R_1{}^2=R_2{}^2-e_1{}^2,$$
$$\mathcal{C}_1=e_1\cot e_1,\quad \mathcal{C}_2=e_2\cot e_2.$$

The S matrix is given by

$$S_{11}=\frac{g(-\rho_1,\rho_2)}{g(\rho_1,\rho_2)}e^{-2i\rho_1},$$

$$S_{22}=\frac{g(\rho_1,-\rho_2)}{g(\rho_1,\rho_2)}e^{-2i\rho_2},$$

$$S_{12}=S_{21}=\frac{2(\mathcal{C}_1-\mathcal{C}_2)(\alpha_1\alpha_2\rho_1\rho_2)^{\frac12}}{g(\rho_1,\rho_2)}e^{-i(\rho_1+\rho_2)}.$$

Potentials of the Yukawa Type

Suppose

$$V(r)=\int d\mu\rho(\mu)e^{-\mu r},$$

where ρ is a matrix and $\rho_{\alpha\beta}(\mu)\equiv0$ for $\mu\leqslant u_{0\alpha\beta}$. Then

$$R(K;r,r')=-K^{-1}\sin Kr_<e^{iKr_>}\int d\mu\rho(\mu)e^{-\mu r'}.$$

It will be advantageous to do everything in Fourier transform language:

$$R(K;k,k')\equiv\frac{1}{2\pi}\int_0^\infty dr\int_0^\infty dr'e^{ikr}e^{-ik'r'}R(K;r,r')$$
$$=\frac{1}{2\pi i}\int d\mu(K+k)^{-1}(k+i\mu-k')^{-1}\times(k'-K-i\mu)^{-1}\rho(\mu).$$

Then

$$\operatorname{Tr}R(K)=-\frac{1}{2\pi}\int\frac{d\mu}{\mu}\int_{-\infty}^\infty dk\,\operatorname{tr}(k+K)^{-1}\times(k-K-i\mu)^{-1}\rho(\mu)$$
$$=-i\int\frac{d\mu}{\mu}\operatorname{tr}(2K+i\mu)^{-1}\rho(\mu)$$
$$=-i\int\frac{d\mu}{\mu}\sum_\alpha(2k_\alpha+i\mu)^{-1}\rho_{\alpha\alpha}(\mu).$$

So as a function of k_α there is a branch line along the negative imaginary axis, starting at $k_\alpha=-\frac12i\mu_{0\alpha\alpha}$.

In order to find the singularities of $Y^{(n)}$ we must, according to (3.6), look at $\operatorname{Tr}R^{n+1}$; the term $R\operatorname{Tr}Y^{(n-1)}$ gives no new singularities.

We need, then, R^2:

$$R^2(k,k')=\frac{1}{(2\pi i)^2}\int d\mu\int d\mu'\int_{-\infty}^\infty\frac{dk''}{(k''-k-i\mu)(k'-k''-i\mu')}\cdot(k''-K-i\mu)^{-1}(k+K)^{-1}\times\rho(\mu)\cdot(k'-K-i\mu')^{-1}(k''+K)^{-1}\rho(\mu)'.$$

All integrations in the trace are readily done and we obtain after some algebra:

$$\operatorname{Tr}R^2(K)=-2i\int d\mu\int d\mu'\sum_{\alpha\beta}\rho_{\alpha\beta}(\mu)\rho_{\beta\alpha}(\mu')\frac{k_\alpha+k_\beta+i\mu+i\mu'}{(\mu+\mu')(2k_\alpha+i\mu+i\mu')(2k_\beta+i\mu+i\mu')(k_\alpha+k_\beta+i\mu)(k_\alpha+k_\beta+i\mu')}.$$

[20] This is the function F of (5.8) in footnote 18 which contains a sign error. Otherwise the notation is the same.

The only new branch line comes from

$$k_\alpha + k_\beta + i\mu = 0$$

or the same with $\mu \to i\mu'$. That branch line starts at

$$k_\beta^{(0)} = -i(\mu_{0\alpha\beta}^2 + 2m_\alpha \Delta_{\alpha\beta}) \times \{\mu_{0\alpha\beta} + (m_\alpha/m_\beta)^{\frac{1}{2}}[\mu_{0\alpha\beta}^2 - 2\Delta_{\alpha\beta}(m_\beta - m_\alpha)]^{\frac{1}{2}}\}^{-1},$$

where

$$\Delta_{\alpha\beta} = (k_\alpha^2/2m_\alpha) - (k_\beta^2/2m_\beta).$$

The branch line extends along the negative imaginary axis only, provided that

$$\mu_{0\alpha\beta} > [2\Delta_{\alpha\beta}(m_\beta - m_\alpha)]^{\frac{1}{2}}; \qquad \text{(A)}$$

otherwise it runs partly parallel to the real axis.

Since the branch lines arise simply from the coincidence of singularities in the original integrands, it is clear that the branch lines of Tr R^2 are the only ones that will occur in Tr R^n and hence in the nth term of the expansion of the Fredholm determinant. We may conclude that $\Delta(k_1, \cdots)$, considered as an over-all function of k_β, i.e., with all k's expressed as functions of k_β, has a branch line along the negative imaginary axis from $k_\beta^{(0)}$ on, provided that (A) is satisfied for all $\mu_{0\alpha\beta}$. In addition, there are the usual branch points coming from the energy conservation between channels; but those lie, of course, necessarily on the real and imaginary axes.

We finally form the S-matrix elements from Δ by (2.14) and (2.23). Then one or two of the k's must change sign. As a result we get branch lines also along the positive imaginary axis. The criterion for branch lines on the imaginary axis only remains (A).

Products of Principal Value Singularities Used in the Formal Theory of Scattering*

S. TANI†
Department of Physics, Washington University, St. Louis, Missouri
(Received May 23, 1960)

A mathematically neat derivation is given of the relation between the S matrix and the transformation function for a finite time. It is shown that one can dispense with the adiabatic switching on and off, and yet one reaches the same result as when one employs it. Necessary conditions are discussed for the validity of this statement. Systematic prescriptions are given of handling products of principal value singularities, which is relevant to the scattering theory in momentum space.

I

THE best way to understand the nature of a propagator as used in the quantum theory of scattering, both in field theory and in ordinary quantum mechanics, is to consider it as the boundary value of some analytic function.[1] Once one admits that the use of such a propagator is well founded, one can proceed formally in the scattering theory if a set of consistent rules is introduced which disposes singularities in momentum space in terms of the delta functions and principal values.

It was pointed out recently[2] that Eqs. (1)–(4) as given in the following play a key role in the so-called synthetic or inverted approach in quantum field theory, so far as one agrees to proceed along the line first mentioned. The advantage of such an approach is that one establishes the quantization even when some phenomenological form factor is introduced. The value of such a form factor is somewhat questionable as the final theoretical tool to be used in the explanation of the properties of elementary particles. Nevertheless, one can emphasize the importance of studying the local field theory as the limit of nonlocal form factor theory. Also, one can show that the relations (1)–(4) provide a clear-cut division between off and on the energy shell of a transition amplitude, however complicated a problem may be.

The relations to be established are as follows. Define the unitary transformation function for a finite time x_0 by

$$\bar{U}(x_0)=\exp\left[\frac{i}{2}\int_{-\infty}^{\infty}\epsilon(x_0,y_0)F(y)d^4y\right]\equiv\exp[iG(x_0)], \quad (1)$$

in terms of a step function

$$\epsilon(a,b)=\begin{cases}+1 & a>b\\ \text{for} & \\ -1 & a<b\end{cases}$$

and an operator $F(x)$ defined at a point x in space-time. The translation of the system, say, from the time t_1 through t_2, is conducted by the transformation function

$$U(t_2,t_1)=\bar{U}(t_2)\cdot\bar{U}(t_1)^{-1}. \quad (2)$$

The S matrix is given then by

$$S=\exp\left[i\int_{-\infty}^{\infty}F(x)d^4x\right], \quad (3)$$

with the same operator $F(x)$ as in (1). If we assume the adiabatic switching on and off,[3] we can establish (3) by substituting (1) into (2). That is to say, we can put

$$S=\lim_{t_2\to\infty,\,t_1\to-\infty}U(t_2,t_1). \quad (4)$$

This means that the S matrix is defined as the translation operator from $-\infty$ through $+\infty$ in time. We are interested first in establishing (1)–(4) without recourse to the adiabatic switching on and off.[4] The validity of (1)–(4) imposes some restriction on the property of $F(x)$, as a matter of fact. This restriction is stated in Sec. II.

If one employs the momentum representation, Eqs. (1)–(4) require that a product of principal values must be carefully treated. This is known in the case of a product of two principal values. The following equality is known under the assumption of regularity of the matrix element of F as a function of energy:

$$(a|F|b)(b|F|c)\left\{P\frac{1}{E_a-E_b}P\frac{1}{E_b-E_c}+\pi^2\delta(E_a-E_b)\delta(E_b-E_c)-P\frac{1}{E_a-E_c}\times\left(P\frac{1}{E_a-E_b}+P\frac{1}{E_b-E_c}\right)\right\}=0, \quad (5)$$

* This work is supported by the Air Force.

† On leave of absence from Tokyo University of Education, Tokyo, Japan. Present address: Physics Department, New York University, University Heights, New York 53, New York.

[1] Cf. e.g., G. Källen and A. Wightman, Kgl. Danske Videnskab. Selskab, Mat. fys. Skrifter 1, No. 6 (1958); there is quite a lot of literature on the dispersion relation.

[2] S. Tani, Phys. Rev. 115, 711 (1959); J. Math. Phys. 2, 46 (1961).

[3] As an example one can refer to Eqs. (3.6) and (3.7), p. 1618, in M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, Phys. Rev. 95, 1612 (1954).

[4] This is known in the case of scattering by a finite-range and regular potential. See N. Levinson, Kgl. Danske Videnskab. Selskab, Mat. fys. Medd. 25, No. 9 (1949); J. M. Cook, J. Math. and Phys. 36, 82 (1957); J. M. Jauch, Helv. Phys. Acta 31, 661

where E_α denotes the energy in the state α. Out of four terms appearing in (5), one cannot perform the integration over E_b earlier than over other variables in the first term; however, one can define such an integral as the combination of three other terms. It is assumed that a matrix element $(a|F|b)$ is regular enough as a function of both E_a and E_b so that a multiple integral of a product of F with a principal value is well defined if performed in a suitable order. The situation in which the number of principal values is larger than two is studied systematically in Sec. III.

II

Let us first define the Hamiltonian as the generating operator of an infinitesimal change in time. It is defined by

$$i[d\bar{U}(t)/dt]=H(t)\bar{U}(t). \tag{6}$$

It is given actually by

$$H(t)=i\left\{\frac{d}{dt}-\bar{U}(t)\cdot\frac{d}{dt}\cdot\bar{U}(t)^{-1}\right\}$$

$$=-\int d^3x\{F(x)+(1/2!)[F(x),-iG(x_0)]$$

$$+(1/3!)[[F(x),-iG(x_0)],-iG(x_0)]+\cdots\}. \tag{7}$$

In the following we suppress the integration over space coordinate in order to shorten the notation. Let us consider the operator $I^{(\pm)}(t)$ which is defined by

$$I^{(\pm)}(t)=\bar{U}(t)\cdot\exp\left[\pm\frac{i}{2}\int_{-\infty}^{\infty}F(x)dx\right]. \tag{8}$$

The most essential step in the proof is to show, for an arbitrary but finite t, that

$$I^{(\pm)}(t)=\mathbf{P}\left(\exp\left[-i\int_{\mp\infty}^{t}H(x)dx\right]\right), \tag{9}$$

where the Hamiltonian $H(t)$ is defined in (7) and $\mathbf{P}$ is the Dyson's chronological ordering operator,[5] namely, the right-hand side of (9) is the shorthand notation for

$$\mathbf{P}\left(\exp\left[-i\int_{-\infty}^{t}H(x)dx\right]\right)$$

$$=1-i\int_{-\infty}^{t}H(x)dx+(-i)^2\int_{-\infty}^{t}H(x)dx$$

$$\times\int_{-\infty}^{x}H(y)dy+\cdots. \tag{10}$$

If one admits (9), it is evident that we must put

$$\lim_{t\to-\infty}I^{(+)}(t)=1. \tag{11}$$

Equation (11) establishes the consistency of the definition

$$\lim_{t\to-\infty}\exp\left[\frac{i}{2}\int_{-\infty}^{\infty}\epsilon(t,x)F(x)dx\right]$$

$$=\exp\left[\frac{-i}{2}\int_{-\infty}^{\infty}F(x)dx\right]. \tag{12}$$

In this way we can dispense with the adiabatic switching off, and yet we reach the same result as when we use it. The limit $t\to+\infty$ of $\bar{U}(t)$ can be established in a similar way. When both limits, $t_2\to+\infty$ and $t_1\to-\infty$ are combined in $U(t_2,t_1)$ we establish (4).

Now, Eq. (9) can be established by expanding both sides into powers of $F(x)$ and by comparing them at the same order. We denote an order of the expansion by a suffix. Accordingly, we write

$$I^{(+)}(t)=1+\sum_{n=1}^{\infty}I_n^{(+)}(t), \tag{13}$$

and

$$H(t)=\sum_{n=1}^{\infty}H_n(t). \tag{14}$$

Also we write for the right-hand side of Eq. (10)

$$\mathbf{P}\left(\exp\left[-i\int_{-\infty}^{t}H(x)dx\right]\right)=1+\sum_{n=1}^{\infty}J_n^{(+)}(t). \tag{15}$$

As a result of the definition given by (10), we have

$$J_{n+1}^{(+)}(t)$$

$$=-i\int_{-\infty}^{t}dx\{H_{n+1}(x)+\sum_{m=1}^{n}H_m(x)J_{n-m+1}(x)\}. \tag{16}$$

Suppose Eq. (9) is established up to the order n. We have to see the condition under which Eq. (9) holds for the order $n+1$, namely,

$$J_{n+1}^{(+)}(t)=I_{n+1}^{(+)}(t). \tag{17}$$

From (7) and (1), we have

$$H_m(t)=(-1)^m\left(\frac{i}{2}\right)^{m-1}\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_mF(x_1)\cdots$$

$$\times F(x_m)\sum_{j=1}^{m}\frac{1}{(j-1)!(m-j+1)!}\epsilon(x_1,t)\cdots$$

$$\times\epsilon(x_{j-1},t)\epsilon(t,x_{j+1})\cdots\epsilon(t,x_m)\delta(t-x_j), \tag{18}$$

where there are $(m-1)\epsilon$'s for all j. From (8) and (1),

(1959); and other papers quoted in them. We reformulate the problem here in order to make our results applicable to wider classes of problems than potential scattering. See the next to the last paragraph in Sec. II.

[5] F. J. Dyson, Phys. Rev. 75, 486 (1949); see p. 492.

we have

$$I_m^{(+)}(t)=\left(\frac{i}{2}\right)^m\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_mF(x_1)\cdots \times F(x_m)\sum_{j=1}^{m}\frac{1}{(j-1)\,!(m-j+1)\,!} \times\epsilon(t,x_1)\epsilon(t,x_2)\cdots\epsilon(t,x_{m-j+1}), \quad (19)$$

where the jth term is a product of only $(m-j+1)\epsilon$'s. If we substitute $I_m(t)$ given by (19) and $H_m(t)$ given by (18) for m up to n on the right-hand side of (16), it turns out that the necessary condition for the validity of (17) is that the following condition holds for an m-times product of $F(x)$ for all m up to $n+1$:

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_mF(x_1)\cdots F(x_m)\{\sum_{j=1}^{m}\epsilon(x_j,x_1) \times\epsilon(x_j,x_2)\cdots\epsilon(x_j,x_{j-1})\epsilon(x_j,t)\epsilon(x_j,x_{j+1})\cdots\epsilon(x_j,x_m) +\epsilon(t,x_1)\epsilon(t,x_2)\cdots\epsilon(t,x_m)+D(m)\}=0, \quad (20)$$

where $m\epsilon$'s appear in each term and the constant $D(m)$ is given by

$$D(m)=\begin{cases}0 & \text{if } m \text{ is odd}\\ -1 & \text{if } m \text{ is even,}\end{cases} \quad (21)$$

and t is arbitrary but finite.

It is not difficult to see that the expression inside the curly brackets in Eq. (20) vanishes almost identically, namely, we have

$$\sum_{j=1}^{m}\epsilon(x_j,x_1)\cdots\epsilon(x_j,x_{j-1})\epsilon(x_j,t)\epsilon(x_j,x_{j+1})\cdots\epsilon(x_j,x_m) +\epsilon(t,x_1)\epsilon(t,x_2)\cdots\epsilon(t,x_m)+D(m)=0, \quad (22)$$

except when several points coincide. The exception arises because a step function is not defined when its arguments coincide.

In order to see that validity of Eq. (22), let us first observe that the left-hand side of (22) is invariant against any permutation of the $m+1$ points, t through x_m. This means that we have only to verify the identity for any arrangement of the $m+1$ points which is convenient; we should get the same result for any other arrangement of $m+1$ points. One can easily see Eq. (22) to hold when the arrangement of $m+1$ points is such that

$$t>x_1>x_2>\cdots>x_m.$$

On coming back to the validity of Eq. (20), we see that it is established if the operator $F(x)$ is regular enough such as:

(i) The integration over x with a step function $\epsilon(tx)$ as a weight is well-defined for an arbitrary but finite t. This means that the contribution from $x=t$ does not introduce any difficulty but can be simply neglected.

(ii) The multiple integral converges uniformly, when the integrations are made according to the order specified by the number of times by which each variable appears. In the jth term as appears in (20), the integration over x_j should be performed at the end, the order of integration over other variables should be interchangeable among them.

Given a particular $F(x)$, one may raise the question as to whether the conditions mentioned above are satisfied for such a particular $F(x)$. We do not go into a systematic investigation of a sufficient condition for the validity of our key equations (1)–(4).[4] It consists of two problems: (i) find the sufficient condition for the validity of (20) or (27) and its generalization, and (ii) find the sufficient condition for the convergence of the expansion into powers of $F(x)$. The solutions to these problems are very valuable, especially when they are formulated in momentum space.

Finally we should have shown how to prove (17) when (18)–(20) are used. Since this is a straightforward algebraic calculation which involves only handling a number of binomial coefficients, we only assert our result and skip to give a detail of such a calculation.

III

It is to be remarked that when one introduces the momentum space representation, Eq. (20) formulates a set of rules for handling a product of principal value singularities. Let us write the matrix element of $F(x)$ between two states, say, a and b, in momentum space as

$$(a\,|\,F(t)\,|\,b)=(a\,|\,F\,|\,b)\exp[-i(E_a-E_b)t]. \quad (23)$$

Recall that the following symbolic calculations are valid under the regularity conditions as discussed in Sec. II:

$$\int_{-\infty}^{\infty}\exp[-i(E_a-E_b)x]dx=2\pi\delta(E_a-E_b), \quad (24)$$

$$\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(t,x)\exp[-i(E_a-E_b)x]dx =\exp[-i(E_a-E_b)t]P\frac{i}{E_a-E_b}. \quad (25)$$

Substitute (23) into (20) and perform the integration according to (24) and (25). Then, we have an equation, which is valid as a distribution equation among products of principal values.

Let us quote an example for which $m=2$. Equation (20) in this case specifically reads

$$\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dyF(x)F(y)\{\epsilon(tx)\epsilon(ty)+\epsilon(xt)\epsilon(xy) +\epsilon(yt)\epsilon(yx)-1\}=0. \quad (26)$$

When we calculate the matrix element between states,

say, a and c, we have the following results:

$$\left\langle a\left|\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(tx)dx\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(ty)dyF(x)F(y)\right|c\right\rangle$$
$$=\langle a|F|b\rangle\langle b|F|c\rangle\cdot\exp[-i(E_a-E_c)t]$$
$$\cdot(i)^2\frac{P}{E_a-E_b}\frac{P}{E_b-E_c}, \quad (27a)$$

$$\left\langle a\left|\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(xt)dx\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(xy)dyF(x)F(y)\right|c\right\rangle$$
$$=\langle a|F|b\rangle\langle b|F|c\rangle\cdot\exp[-i(E_a-E_c)t]$$
$$\cdot P\frac{1}{E_a-E_c}P\frac{1}{E_b-E_c}, \quad (27b)$$

$$\left\langle a\left|\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(yt)dy\int_{-\infty}^{\infty}\tfrac{1}{2}\epsilon(yx)dxF(x)F(y)\right|c\right\rangle$$
$$=\langle a|F|b\rangle\langle b|F|c\rangle\cdot\exp[-i(E_a-E_c)t]$$
$$\cdot P\frac{1}{E_a-E_c}P\frac{1}{E_a-E_b}, \quad (27c)$$

$$\left\langle a\left|\int_{-\infty}^{\infty}\tfrac{1}{2}dx\int_{-\infty}^{\infty}\tfrac{1}{2}dyF(x)F(y)\right|c\right\rangle$$
$$=\langle a|F|b\rangle\langle b|F|c\rangle\pi^2\delta(E_a-E_b)\delta(E_b-E_c), \quad (27d)$$

respectively, in the order as they appear in Eq. (26). One should be careful to keep a proper order of integration so that the integral is meaningful. In writing the preceding results, the integration over the density of the intermediate state b is suppressed. When (27a)–(27d) are summed up, we see that the relation (5) mentioned in Sec. I is established.

In this way, if represented in momentum space, Eq. (20) will provide a systematic method of exploring the products of principal values in general situations such as are encountered in the scattering theory. Actually the rules defined in this way reproduce all of the rules used in our earlier publications.[2] They were formulated by a more primitive method. In that method one starts with

$$P\frac{1}{E_a-E_b}\sim\frac{E_a-E_b}{(E_a-E_b)^2+\epsilon^2}$$
$$\pi\delta(E_a-E_b)\sim\frac{\epsilon}{(E_a-E_b)^2+\epsilon^2},$$

and takes the limit $\epsilon\to 0$ at the end.

In conclusion, we have established that if the regularity conditions formulated in Sec. II are satisfied, the adiabatic switching on and off can be justified as the conventional means to derive the right result. We have dealt with only the exponential form of a unitary transformation function. This is made in connection with its usefulness, since the exponential form is unique in that we can ignore processes which are represented by reducible diagrams. As far as justification of the adiabatic switching is concerned, we can treat the other types of transformation function equally well.

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to Professor Eugene Feenberg, Professor Leslie Foldy, and Professor Henry Primakoff for their kind interest and useful comments. He is deeply indebted to Dr. Joseph Dreitlein for a very enlightening discussion which brought a considerable elegance into the proof.

Classification of Four-Dimensional Riemannian Spaces

J. F. SCHELL
Aeronautical Research Laboratories, Wright-Patterson Air Force Base, Ohio
(Received August 23, 1960)

A classification of four-dimensional Riemann spaces with signature +2 is given. The classification depends upon the differential as well as the algebraic properties of the Riemann tensor. The tool employed is the infinitesimal-holonomy group of the space. An introduction to the concept of the holonomy group is given, and the technique of classification is outlined. A comparison with the classification of empty spaces given by A. Z. Petrov and with the recent work of E. Newman is also given.

1. INTRODUCTION

A CLASSIFICATION of Einstein spaces according to the eigenbivectors of the curvature tensor has been given by Petrov.[1] This classification depends only upon the algebraic properties of the curvature tensor and does not take into account its differential properties. The present paper gives a classification of general V_4's with signature $+2$ [$V_4(+++-)$'s] which is based on the theory of the infinitesimal-holonomy group (ihg) for a V_4. This theory involves not only the curvature tensor but also brings in all of its covariant derivatives. Thus we are led to a classification of $V_4(+++-)$'s which brings in the differential as well as the algebraic properties of the curvature tensor.

We begin (Sec. 2) by introducing the concept of an ihg for a general V_n, and indicate its relation to rotation groups. This is followed (Sec. 3) by an elementary discussion of continuous rotation groups in an n-dimensional flat space. We then return (Sec. 4) to a further discussion of ihg's in a general V_n. The classification of $V_4(+++-)$'s is obtained (Sec. 5) from a classification of rotation groups in a four-dimensional flat space. Specialization to Einstein spaces is then discussed (Sec. 6), and, finally (Sec. 7), the relation of this classification to that of Petrov and the work of Newman[2] is treated.

2. INFINITESIMAL-HOLONOMY GROUP FOR AN ANALYTIC REGION $R \subset V_n$

We denote by V_n an n-dimensional Riemann space. Let P be a point in an analytic region $R \subset V_n$, in which the Christoffel symbols are also analytic. Let C be a closed curve in R which can be continuously shrunk to a point in R and which passes through P. The vector transformation at P which is generated by parallel displacement (of the set of vectors at P) around C will be a rotation. The set of all such curves C in R will give rise to a set of rotations at P. It can be shown that this set of rotations actually forms a connected continuous group,[3,4] the group called the infinitesimal-holonomy group of V_n at P. It is also easily shown that the same abstract group obtains at each point of R, so that we may speak of the ihg of $R \subset V_n$.

Since the ihg is a rotation group in the (flat) tangent space to V_n at P, it is profitable at this point to digress a bit and discuss the properties of a general group of rotations in an n-dimensional flat space.

3. ROTATION GROUPS IN AN n-DIMENSIONAL FLAT SPACE

Let ξ^κ ($\kappa = I, \cdots$, roman numeral n) denote a set of coordinates in an n-dimensional flat space. Consider a group of rotations about the origin of this coordinate system. It is known that the group germ (that is, the set of elements of the group which can be continuously transformed into the identity) is also a Lie group and will have its transformations expressible in the form

$$\eta^\kappa = e^{Xt}\xi^\kappa = \left[1 + Xt + \frac{(Xt)^2}{2!} + \cdots\right]\xi^\kappa, \tag{1}$$

where η^κ denotes the coordinates of the point into which ξ^κ is rotated and X is an operator of the form

$$X = L_\lambda{}^\kappa \xi^\lambda (\partial/\partial \xi^\kappa),^5 \tag{2}$$

the

$$L_{\lambda\kappa} \equiv g_{\kappa\mu} L_\lambda{}^\mu = -L_{\kappa\lambda} \tag{3}$$

being components of a bivector in the space. The infinitesimal transformation

$$\eta^\kappa = \xi^\kappa + L_\lambda{}^\kappa \xi^\lambda dt = (1 + X dt)\xi^\kappa, \tag{4}$$

obtained from Eq. (1) by neglecting higher-order terms and, for purely notational reasons, by replacing t by dt, is said to generate the set of transformations in Eq. (1). The operator X is called the generator of the transformations; the tensor $L_{\lambda\kappa}$ is called the generating bivector. In the remainder of this discussion, we confine our attention solely to the group germ.

Because of the bivector character of $L_{\lambda\kappa}$, the rotation

[1] A. Z. Petrov, Sci. Note Kazan State Univ. 114, 55 (1954).
[2] E. Newman, J. Math. Phys. (to be published).
[3] A. Nijenhuis, Koninkl. Ned. Akad. Wetenschap. Proc. Ser. A 56, 233, 241 (1953).
[4] J. A. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin, 1954), p. 361 ff.

[5] Throughout this paper the summation convention is used with respect to all types of indices. Bracketed indices, for example, $[\alpha\beta]$, indicate skew-symmetrization.

group can have at most $n(n-1)/2$ independent generators [see Eq. (2)], or independent generating bivectors [see Eq. (3)]. The number r of independent generators is referred to as the number of parameters of the group.[6] If the only element in the group is the identity, we shall say that the group has zero parameters.

Consider now an r-parameter group G,

$$0<r\leqq n(n-1)/2.$$

Denote by

$$L_{\mathbf{p}\alpha\beta}, \quad \mathbf{p}=1, \cdots, r \tag{5}$$

a set of r independent generating bivectors. Then every bivector

$$L_{\alpha\beta}=C^{\mathbf{p}}L_{\mathbf{p}\alpha\beta} \quad (C^{\mathbf{p}}=\text{const}) \tag{6}$$

will generate a one-parameter subgroup of G. Moreover, every element of G is generated by an $L_{\alpha\beta}$ in the form of Eq. (6) with uniquely determined $C^{\mathbf{p}}$.

The generators corresponding to Eq. (5) are

$$X_{\mathbf{p}}=L_{\mathbf{p}\lambda}{}^{\kappa}\xi^{\lambda}(\partial/\partial\xi^{\kappa}). \tag{7}$$

From this equation and the Lie structural formulas for continuous groups,

$$\begin{gathered}2X_{[\mathbf{p}}X_{\mathbf{q}]}\equiv X_{\mathbf{p}}X_{\mathbf{q}}-X_{\mathbf{q}}X_{\mathbf{p}}=C_{\mathbf{pq}}{}^{\mathbf{r}}X_{\mathbf{r}} \\ (C_{\mathbf{pq}}{}^{\mathbf{r}}=C_{[\mathbf{pq}]}{}^{\mathbf{r}}),\end{gathered} \tag{8}$$

we obtain the following commutation identity for the generating bivectors:

$$L_{\mathbf{p}\alpha\beta}L_{\mathbf{q}}{}^{\beta}{}_{\gamma}-L_{\mathbf{q}\alpha\beta}L_{\mathbf{p}}{}^{\beta}{}_{\gamma}=C_{\mathbf{pq}}{}^{\mathbf{r}}L_{\mathbf{r}\alpha\gamma}. \tag{9}$$

This relation is very important for the classification of n-dimensional rotation groups and hence for the classification of ihg's. Ultimately, the classification rests on the well-known second part of the second theorem of Lie.

A necessary and sufficient condition that a set $L_{\mathbf{p}\alpha\beta}$ $(\mathbf{p}=1, \cdots, r)$ *of* r *bivectors be a complete set*[7] *of generating bivectors for a group (of rotations) is that the relations in Eq. (9) hold.*

4. INFINITESIMAL-HOLONOMY GROUP OF AN ANALYTIC REGION $R\subset V_n$ (CONTINUED)

Since the ihg of a region $R\subset V_n$ is a group of rotations at a point P, it will possess generating bivectors which have all of the properties described in the preceding paragraph. Moreover, these bivectors are intimately related to the curvature tensor.[3,4,8] Indeed, a complete set of generating bivectors $L_{\mathbf{p}\lambda\nu}$ for the ihg spans the same set of bivectors as do

$$\begin{gathered}R_{ab\lambda\nu}, \nabla_{x_1}R_{ab\lambda\nu}, \cdots, \nabla_{x_p}\cdots\nabla_{x_1}R_{ab\lambda\nu}, \cdots \\ (a, b, x_i=\bar{1}, \cdots, \bar{n}).\end{gathered} \tag{10}$$

The bivectors in Eq. (10) are obtained from the curvature tensor and its derivatives by contraction over all but the $\lambda\nu$ indices with the vectors $e_a{}^{\alpha}$ $(a=\bar{1}, \cdots, \bar{n})$ of an arbitrary ennuple. (The indices a, b, x_i are merely labeling indices for the different bivectors.) Thus, we have

$$R_{\alpha\beta\gamma\delta}=M^{\mathbf{q}}{}_{\alpha\beta}L_{\mathbf{q}\gamma\delta},$$

and similar relations for the covariant derivatives of $R_{\alpha\beta\gamma\delta}$. It can be shown[3,4,8] that $M^{\mathbf{p}}{}_{\alpha\beta}=h^{\mathbf{pq}}L_{\mathbf{q}\gamma\delta}$ and, hence,

$$R_{\alpha\beta\gamma\delta}=h^{\mathbf{pq}}L_{\mathbf{p}\alpha\beta}L_{\mathbf{q}\gamma\delta}, \tag{11}$$

where $h^{\mathbf{pq}}$ is symmetric in its indices.

One final property of an ihg which is relevant to the classification of V_4's is that of perfectness.[8] The ihg of a region $R\subset V_n$ is called *perfect* if there is at least one point $P\epsilon R$ such that the set $(R_{ab\lambda\nu})_P$ *alone* spans the entire set of generating bivectors of the group. In other words, an ihg will be called perfect, if the bivectors in Eq. (10) arising from the covariant derivatives are linearly dependent on those arising from the curvature tensor alone. An ihg which is not perfect will be called *imperfect*. There will be various degrees of imperfectness depending upon the number of linearly independent bivectors in Eq. (10) which are given by the curvature tensor alone, how many come from first derivatives, etc. Clearly, from Eq. (11), a necessary and sufficient condition for perfectness is (remember that for an r-parameter ihg $\mathbf{p}, \mathbf{q}=1, \cdots, r$)

$$\det(h^{\mathbf{pq}})\neq 0 \quad \text{for some } P\epsilon R.$$

In the next section we shall give a summary of the results of a classification of groups of rotation in a four-dimensional Minkowskian space[9] and shall use these results to obtain a preliminary classification of ihg's in a $V_4(+++-)$.

5. CLASSIFICATION OF INFINITESIMAL-HOLONOMY GROUPS IN $V_4(+++-)$

In our classification of ihg's of a $V_4(+++-)$, we shall employ the results of another paper[9] which gives a classification of groups of rotation in four-dimensional flat spaces R_4 whose metric tensors have the various possible inequivalent signatures: $(++++)$, $(+++-)$, $(++--)$. We begin first, then, with a brief outline of the method of classification of rotation groups.

A rotation group in an R_4 can be at most six parametric, but it may also have fewer than six parameters. Because of the commutation relations in Eq. (9), not every arbitrarily chosen set of r bivectors [see Eq. (5)] will be a complete set of generators of a group. Indeed, in $R_4(+++-)$ the restriction which Eq. (9) puts on the generating tensors is such as to exclude five-parameter rotation groups.[10] Moreover, the nature of

[6] For $n=4$, a rotation group can have at most $4(4-1)/2=$six independent generators and, hence, at most six parameters.

[7] A set of bivectors [see Eq. (5)] is said to be a complete set of generating bivectors of a group G (of rotations), if every generating bivector of the group can be expressed in the form of Eq. (6).

[8] V. Hlavatý, J. Math. and Mech. 8, 285, 597 (1959); 9, 89, 453 (1960).

[9] J. F. Schell (unpublished).

[10] Similarly, in an $R_4(++++)$ five-parameter rotation groups are also excluded; however, in an $R_4(++--)$ five-parameter rotation groups are possible.

TABLE I. Classification of rotation groups in $R_4(+++-)$.

Number of group parameters	Class symbol	Canonical basis for system of generators
0	R_1	...
1	R_2	$[pq]$
	R_3	$[px]$
	R_4	$[xy]$
	R_5	$[pq]+[xy]$
2	R_6	$[pq]$ $[px]$
	R_7	$[pq]$ $[xy]$
	R_8	$[px]$ $[py]$
3	R_9	$[pq]$ $[px]$ $[py]$
	R_{10}	$[pq]$ $[px]$ $[qx]$
	R_{11}	$[px]$ $[py]$ $[xy]$
	R_{12}	$[px]$ $[py]$ $[xy]+[pq]$
	R_{13}	$[xy]$ $[xz]$ $[yz]$
4	R_{14}	$[pq]$ $[px]$ $[py]$ $[xy]$
6	R_{15}	$[pq]$ $[px]$ $[py]$ $[qx]$ $[qy]$ $[xy]$

two-, three-, and four-parameter groups are restricted. The theory of line geometry in a three-dimensional space and its representation in the five-dimensional homogeneous Klein space K_5 [11,12] are employed to obtain an exhaustive method of selecting various types of r-parameter sets of bivectors.[13] From each set we select a basis consisting of r linearly independent bivectors $L_{p\alpha\beta}$. By checking the validity of Eq. (9), those sets which form a complete set of generators for a group of rotations are singled out.

In K_5 the various classes of rotation groups are characterized by an algorithm which makes use of geometrical representations (in K_5) for the Levi-Civita tensor density $\epsilon_{\alpha\beta\gamma\delta}$ and for the tensor $g_{\alpha[\gamma}g_{\delta]\beta}$ which plays an important role in Petrov's classification. In the present paper, however, an equivalent but seemingly more physical characterization by canonical bases for the set of generating bivectors is given. A rotation group will belong to a given class, if its generating bivectors have a basis expressible in the canonical form for that class. The canonical basis for the 15 different (mutually exclusive) classes of rotation groups in an $R_4(+++-)$ are given in Table I. In this table the following symbolism is used: x^α, y^α are two mutually orthogonal spacelike real vectors; p^α, q^α are the two null vectors orthogonal to both x^α, y^α; z^α is an arbitrary real spacelike vector in the plane of p^α and q^α and is thus orthogonal to x^α, y^α; $[pq]$ denotes the skew-symmetric product $p^{[\alpha}q^{\beta]}$, etc.

It is interesting to note that only in two classes, R_5 and R_{12}, are we *compelled* to employ a general bivector in the canonical basis.

Not all of the classes of rotation groups contain groups which may be infinitesimal holonomy groups for some $V_4(+++-)$. For, if a rotation group with a canonical basis $L_{p\alpha\beta}$ is an ihg for a $V_4(+++-)$, then the curvature tensor is, in some nonholonomic frame (vierbein), expressible in the form of Eq. (11), and must satisfy the cyclic identity

$$R_{[\alpha\beta\gamma]\delta}=0. \tag{12}$$

(The other algebraic symmetries are automatically satisfied.) By using the canonical basis for class R_5, it is easily shown that a tensor given by Eq. (11) will satisfy Eq. (12) only if $h^{11}=0$; this implies that $R_{\alpha\beta\gamma\delta}=0$ everywhere. Thus the space would be flat and the holonomy group would consist of the identity alone and would be of class R_1 contrary to assumption. All other classes of rotation groups lead to tensors [see Eq. (11)] which are compatible with the cyclic identity. Nonetheless, whether each of the other classes contain or do not contain ihg's for some $V_4(+++-)$, is, as of this writing, still an unanswered problem.

6. *IHG* CLASSIFICATION OF $(+++-)$-EINSTEIN SPACES

A $V_4(+++-)$ whose Ricci tensor satisfies the Einstein condition

$$R_{\alpha\beta}=\kappa g_{\alpha\beta} \quad (R_{\alpha\beta}\equiv R_{\mu\alpha\beta}{}^{\mu}, \quad \kappa=a \text{ scalar}), \tag{13}$$

is termed a $(+++-)$-Einstein space. In this section we outline the technique for showing that the ihg's for $(+++-)$-Einstein spaces are contained in at most five classes of rotation groups and that empty spaces ($\kappa=0$) are contained in only four of the classes of rotation groups.

Clearly, no Einstein space can have its ihg in class R_5 since, as shown previously, there is no ihg of any $V_4(+++-)$ in this class. On the other hand, the only $V_4(+++-)$ with an ihg in class R_1 is flat space ($R_{\alpha\beta\gamma\delta}=0$) and is a trivial example of an empty space. That there can be no Einstein space with an ihg of class R_k, $k=2, 3, 4, 6, 10, 13$, follows from the following fact[1,14]:

In a nonholonomic frame (vierbein, tetrad, 4-nuple)

$$u_a{}^\alpha, \quad u^a{}_\alpha, \quad u_a{}^\alpha u^b{}_\alpha=\delta_a{}^b \quad (a, b=\bar{1}, \cdots, \bar{4}), \tag{14}$$

in which the matrix of the nonholonomic (physical)

[11] V. Hlavatý, *Differential Line Geometry* (P. Noordhoff, Groningen, 1953).

[12] V. Hlavatý, *Geometry of Einstein's Unified Field Theory* (P. Noordhoff, Groningen, 1958), Appendix III.

[13] The homogeneous Klein space K_5 may be regarded as a five-dimensional homogeneous bivector space. A point in K_5 has six homogeneous coordinates l^A ($A=1\cdots6$) and is associated with a bivector $L^{\alpha\beta}=L^{[\alpha\beta]}$ in R_4 according to an arbitrarily chosen correspondence $A \leftrightarrow [\alpha\beta]$ between the six indices A and the six skew-symmetric index pairs $[\alpha\beta]=-[\beta\alpha]$. The group of transformations in K_5 (other than the homogeneous transformation $l^{A'}=\rho l^A$, $\rho=$const) is induced by coordinate transformations in R_4, namely, a transformation

$$l^{A'}=a_A{}^{A'}l^A,$$

in K_5 is admissible, if and only if,

$$a_A{}^{A'} \leftrightarrow a_{\alpha\beta}{}^{\alpha'\beta'}\equiv a_{[\alpha}{}^{\alpha'}a_{\beta]}{}^{\beta'}=a_{[\alpha\beta]}{}^{[\alpha'\beta']},$$

where

$$x^{\alpha'}=a_\alpha{}^{\alpha'}x^\alpha+a^{\alpha'}$$

is an admissible transformation in R_4.

[14] H. S. Ruse, Proc. London Math. Soc. **50**, 75 (1944).

components of the metric tensor has the form

$$(g_{ab})\equiv(g_{\alpha\beta}u_a{}^\alpha u_b{}^\beta)=\mathrm{diag}(1\ 1\ 1\ -1), \qquad (15)$$

the (nonholonomic) components R_{abcd} of the curvature tensor of an Einstein space are such that

$$(R_{AB})=\begin{pmatrix} M & N \\ N & -M \end{pmatrix}, \quad (A,B=1,\cdots,6), \qquad (16)$$

where

$$M \text{ and } N \text{ are } 3\times 3 \text{ symmetric matrices,} \qquad (17)$$

$$\text{(a) } \mathrm{tr}M=\kappa \quad \text{(b) } \mathrm{tr}N=0, \qquad (18)$$

and

A and B are composite indices representing skew-symmetric index pairs according to the scheme:

$$\begin{array}{ccccccc} A & 1 & 2 & 3 & 4 & 5 & 6 \\ [ab] & [\overline{14}] & [\overline{24}] & [\overline{34}] & [\overline{23}] & [\overline{31}] & [\overline{12}]. \end{array} \qquad (19)$$

The condition (18b) is the cyclic identity of Eq. (12). The conditions (16), (17), and (18a) result from the Einstein condition of Eq. (13).

Thus, if a given class of rotation groups contains ihg's for $(+++-)$-Einstein spaces then the curvature tensor given by Eq. (11) must satisfy the conditions (16)–(18) in an arbitrary orthonormal nonholonomic frame. If there is an Einstein space with ihg of class R_k, $k=1,\cdots,15$, its curvature tensor is given by Eq. (11), where the $L_p{}^{\alpha\beta}$ may be taken as the bivectors of the canonical basis for class R_k (as given in Table I). For each class, introduce the orthonormal nonholonomic frame

$$u_{\bar{1}}{}^\alpha=\hat{x}^\alpha, \qquad u_{\bar{2}}{}^\alpha=\hat{y}^\alpha$$
$$u_{\bar{3}}{}^\alpha=(\hat{p}^\alpha+\hat{q}^\alpha)/\sqrt{2}, \quad u_{\bar{4}}{}^\alpha=(\hat{p}^\alpha-\hat{q}^\alpha)/\sqrt{2},$$

where $\hat{x}^\alpha$ and $\hat{y}^\alpha$ are unit vectors in the direction of x^α and y^α, and $\hat{p}^\alpha$, $\hat{q}^\alpha$ are vectors along p^α and q^α which satisfy the relation $\hat{p}^\alpha\hat{q}^\alpha=1$. It is easily shown that when $k=2$, 3, 4, 6, 10, 13, condition (17) requires that all $h^{pq}=0$ and, hence, that the space be flat space, but then the ihg would be of class R_1, contrary to assumption. Thus there are no Einstein spaces with ihg's in these classes. On the other hand, when $k=7$, 8, 9, 11, 12, 14, 15, the conditions (15)–(18) will be satisfied only if the h^{pq} satisfy certain conditions. These conditions on h^{pq} are given in Table II. For completeness we also include class R_1.

From Table II it is seen that if Einstein spaces with three-parameter holonomy groups exist (classes R_9, R_{11}, R_{12}), they are empty: $\kappa=0$. The following considerations will show that in fact they do not exist.

It is easy to show that an empty Einstein space with (R_{AB})-rank 2 has its curvature tensor expressible in the form

$$R_{\alpha\beta\gamma\delta}=\rho(p_{[\alpha}x_{\beta]}p_{[\gamma}x_{\delta]}-p_{[\alpha}y_{\beta]}p_{[\gamma}y_{\delta]}), \qquad (20)$$

TABLE II. Conditions on h^{pq} imposed by conditions (16)–(18).

Number of group parameters	Rotation class	Conditions on h^{pq}	
0	R_1	$\cdots$	$\kappa=0$
2	R_7	$h^{11}=-h^{22}\neq 0$, $h^{12}=0$,	$\kappa=h^{11}\neq 0$
	R_8	$h^{11}=-h^{22}$	
3	R_9, R_{11}, R_{12}	$h^{11}=-h^{22}$, $h^{3p}=h^{p3}=0$,	$\kappa=0$
4	R_{14}	$h^{11}=-h^{22}$, $h^{33}=-h^{44}$, $h^{13}=-h^{24}$, $h^{23}=h^{14}=h^{34}=0$	$\kappa=h^{33}$
6	R_{15}	$h^{11}=-h^{44}$, $h^{22}=-h^{55}$, $h^{33}=-h^{66}$, $h^{12}=-h^{45}$, $h^{13}=-h^{46}$, $h^{23}=-h^{56}$, $h^{15}=h^{42}$, $h^{43}=h^{16}$, $h^{53}=h^{26}$, $h^{14}+h^{25}+h^{36}=0$,	$\kappa=h^{11}+h^{22}+h^{33}$

where p_α, x_α, y_α have the same significance as before in connection with Table I. Thus the bivectors $[px]$, $[py]$ are two independent generating bivectors of the ihg. Consideration of the covariant derivatives of Eq. (20) discloses that either no more linearly independent generating bivectors, or four such bivectors, arise from the derivatives. Thus an empty Einstein space of (R_{AB})-rank $r'=2$ has an ihg of either two or six parameters.

It follows from conditions (16) and (17) that the rank r' of (R_{AB}) for an Einstein space must be 0, 2, 4, or 6. This implies that there are r' linearly independent rows in the matrix (R_{AB}). The rows of (R_{AB}) will then determine r' linearly independent bivectors, according to the scheme (19). These bivectors (given by their components in the nonholonomic frame $u_a{}^\alpha$, $u^a{}_\alpha$) are precisely the bivectors of Eq. (10) which come from the curvature tensor alone and thus are linearly independent bivector generators of the holonomy group. Thus if an Einstein space with a three-parameter holonomy group exists, the rank r' of (R_{AB}) must be an even number such that $0<r'<3$. We conclude that $r'=2$, and hence $r=2$ or 6, contrary to the assumption that $r=3$.

7. RELATION BETWEEN THE *IHG* CLASSIFICATION OF EMPTY SPACES AND THE PETROV CLASSIFICATION

Petrov[1] has given a classification of $(+++-)$-Einstein spaces according to the eigenbivectors of the eigenvalue problem

$$R_{\alpha\beta\gamma\delta}L^{\gamma\delta}=KL_{\alpha\beta}(=Kg_{\alpha[\gamma}g_{\delta]\beta}L^{\gamma\delta}).$$

He found three types. Type I is characterized by the existence of six independent eigenbivectors; type II by four independent eigenbivectors; type III by two independent eigenbivectors. Petrov also showed that by a suitable choice for a nonholonomic frame, the Riemann tensor of each type assumes a characteristic

canonical form:

Type I.

$$(R_{AB})=\begin{pmatrix} \alpha_1 & \cdot & \cdot & \beta_1 & \cdot & \cdot \\ \cdot & \alpha_2 & \cdot & \cdot & \beta_2 & \cdot \\ \cdot & \cdot & \alpha_3 & \cdot & \cdot & \beta_3 \\ \beta_1 & \cdot & \cdot & -\alpha_1 & \cdot & \cdot \\ \cdot & \beta_2 & \cdot & \cdot & -\alpha_2 & \cdot \\ \cdot & \cdot & \beta_3 & \cdot & \cdot & -\alpha_3 \end{pmatrix}$$

$$\sum_{i=1}^{3}\alpha_i=\kappa, \quad \sum_{i=1}^{3}\beta_i=0.$$

Type II.

$$(R_{AB})=\begin{pmatrix} \alpha_1 & \cdot & \cdot & \beta_1 & \cdot & \cdot \\ \cdot & \alpha_2-\sigma & \cdot & \cdot & \beta_2 & \sigma \\ \cdot & \cdot & \alpha_2+\sigma & \cdot & \sigma & \beta_2 \\ \beta_1 & \cdot & \cdot & -\alpha_1 & \cdot & \cdot \\ \cdot & \beta_2 & \sigma & \cdot & -\alpha_2+\sigma & \cdot \\ \cdot & \sigma & \beta_2 & \cdot & \cdot & -\alpha_2-\sigma \end{pmatrix}$$

$$\alpha_1+2\alpha_2=\kappa, \quad \beta_1+2\beta_2=0$$
$$\sigma\neq 0.$$

Type III.

$$(R_{AB})=\begin{pmatrix} \alpha & \sigma & \cdot & \cdot & \cdot & \cdot \\ \sigma & \alpha & \cdot & \cdot & \cdot & \sigma \\ \cdot & \cdot & \alpha & \cdot & \sigma & \cdot \\ \cdot & \cdot & \cdot & -\alpha & -\sigma & \cdot \\ \cdot & \cdot & \sigma & -\sigma & \alpha & \cdot \\ \cdot & \sigma & \cdot & \cdot & \cdot & \alpha \end{pmatrix}$$

$$3\alpha=\kappa.$$

Table III shows the relation between the Petrov classification of empty spaces and the ihg classification. The concept of the perfectness of the holonomy group was used in establishing the comparison. As demonstrated in the preceding section, the number of generating bivectors of the ihg coming from the curvature tensor alone is always even and is equal to the rank of (R_{AB}). Thus the degree of imperfectness of the holonomy group is given by the rank of (R_{AB}) in comparison with the number of group parameters. With the exception of the impossibility of an Einstein space with a four-parameter imperfect holonomy group, the relation between the two classifications is readily established by the use of the canonical bases of Table I, Eq. (11), and the Petrov canonical forms of (R_{AB}). That the four-parameter imperfect case does not occur, follows from the fact that, if such a case existed, the (R_{AB})-rank would necessarily be 2, but in that event, as demonstrated before, the number of parameters of the holonomy group must be 2 or 6 and could not be 4.

TABLE III. Relation between ihg classification of empty spaces and the Petrov classification.

Number of group parameters	Rotation group class	(R_{AB}) rank	Petrov type
0	R_1	0	I
2	R_7	2	II
4	R_{14}	4	III
		6	I II
6	R_{15}	4	I III
		2	II

A study of Table II discloses the following:

(a) Petrov type I spaces have either 0- or 6-parameter ihg's.

(b) Petrov type II spaces have either 2- or 6-parameter ihg's.

(c) Petrov type III spaces have 4- or 6-parameter ihg's and the (R_{AB})-rank is necessarily 4.

(d) There are no empty spaces with 1-, 3-, or 5-parameter holonomy groups.

(e) All empty Einstein spaces with an imperfect ihg are six parametric.

Newman[2] has developed the following treatment of empty space Riemann tensors. He considers a non-holonomic frame with base vectors $\hat{x}^\alpha$, $\hat{y}^\alpha$, $\hat{p}^\alpha$, $\hat{q}^\alpha$ (these vectors have the same significance as in Sec. 6). From these vectors are constructed the six bivectors $[\hat{p}\hat{q}]$, $[\hat{p}\hat{x}]$, $[\hat{p}\hat{y}]$, $[\hat{q}\hat{x}]$, $[\hat{q}\hat{y}]$, $[\hat{x}\hat{y}]$, and their quadratic products. Among the linear combinations of such products, 10 independent tensors are selected which possess all the symmetries of the curvature tensor and have vanishing "Ricci tensor." An arbitrary curvature tensor for empty space is then a linear combination of these 10 tensors. The covariant derivatives of the 10 "basic curvature tensors" are computed and expressed as linear combinations of the "basic curvature tensors" with vector coefficients. Newman also gives the form of the curvature tensors of the three Petrov types as linear combinations of the 10 "basic curvature tensors"; the coefficients are just the α's, β's, and σ's of Petrov's canonical form.

It follows from this that the formalism used by Newman and that employed in the ihg classification are closely related. Indeed, using the forms of the curvature tensor which Newman gives and the differential relations which he also derives, it is an easy matter to check the validity of Table III.

ACKNOWLEDGMENTS

We are indebted to Professor V. Hlavaty for suggesting this research and for many helpful discussions. We also wish to acknowledge discussions with Dr. J. N. Goldberg and Professor E. Newman, and to thank the former for many suggestions for improvement in presentation.

Geometry of Light Paths between Two Material Bodies

E. P. WIGNER
Palmer Physical Laboratory, Princeton, New Jersey
(Received September 23, 1960)

The pattern of light signals, which was proposed before for the measurement of the curvature, is investigated in a two-dimensional manifold of constant curvature (deSitter space). The pattern consists of light signals between two freely moving bodies, each signal being emitted when the signal from the other body arrives. It is shown that the coordinates of the arrivals (or emissions) of the light signals can be obtained from the coordinates of the emission of the first signal by means of projective transformations [see Eq. (18)] which are iterates of a single such transformation. The same applies to the proper times at which these signals are received.

PRELIMINARY REMARKS ABOUT THE deSITTER SPACE

IN connection with the measurement of the curvature in a two-dimensional universe, the paths of light signals between two freely moving bodies were considered in some detail.[1] Since the curvature is, in general, different for different points of space time, the earlier investigation was restricted to the case in which the spatial distance of the two material bodies is small as compared with the radius of curvature. This restriction appears necessary, if one wants to assume that the curvature is constant throughout the region in which the measurement takes place. It may be worthwhile to point out, nevertheless, that in a (two-dimensional) space of constant curvature, i.e., in a two-dimensional deSitter space, the light signals between freely moving bodies form a simple geometric pattern, no matter what is the state of relative motion of the two material bodies.

The two-dimensional deSitter space can be visualized most easily[2] as a hyperboloid which is embedded into a three-dimensional space x, y, τ; the equation of the hyperboloid is

$$x^2+y^2-\tau^2=a^2, \tag{1}$$

where a is the "radius of the universe." It is connected with the Riemann tensor by the equation

$$R_{0101}=2/a^2. \tag{1a}$$

Both the hyperboloid and the metric are invariant with respect to linear transformations which leave the form

$$F=x^2+y^2-\tau^2 \tag{2}$$

invariant. This enables one to obtain all geodesics by transformations of a single geodesic. For reasons of symmetry

$$x=a\cosh s, \quad y=0, \quad \tau=a\sinh s, \tag{3}$$

is a geodesic; others will be obtained therefrom by transformations which leave F invariant. Furthermore, the geodesic (3), as a whole, is invariant under the Lorentz transformation in x and τ,

$$\begin{aligned} x'&=\cosh\chi x+\sinh\chi\tau \\ y'&=y \\ \tau'&=\sinh\chi x+\cosh\chi\tau. \end{aligned} \tag{4}$$

In the parametric representation (3) of this geodesic, the transformation (4) replaces s by $s+\chi$. It then follows that the distance of two points of (3), characterized by two values s_1 and s_2 of the parameter s, depends only on the difference s_2-s_1 of the parameters. One infers from this that the distance of the two points of the geodesic, measured along the geodesic, is proportional to the difference of the values of the parameter s which characterize these points. The constant of proportionality is easily calculated to be a, so that the distance becomes, simply, $a(s_2-s_1)$.

A Lorentz transformation in y and τ produces from (3) the new geodesic

$$\begin{aligned} x'&=x=a\cosh s \\ y'&=y\cosh\varphi+\tau\sinh\varphi=a\sinh\varphi\sinh s \\ \tau'&=y\sinh\varphi+\tau\cosh\varphi=a\cosh\varphi\sinh s, \end{aligned} \tag{5}$$

and all timelike geodesics which go through the point $x=a$, $y=\tau=0$ have a parametric representation of the form (5) with a suitable φ. This φ is simply the hyperbolic angle between the velocities which correspond to the geodesics (3) and (5). The distance of two points of (5), characterized by the values s_1 and s_2 of the parameter s, as measured along the geodesic, is still $a(s_2-s_1)$, since this distance is invariant with respect to the transformation which led from (3) to (5).

In order to obtain a two-dimensional picture of our deSitter space, we again suppress the variable τ. Then, the points of the xy plane which are outside the circle

$$x^2+y^2=a^2 \tag{6}$$

each correspond to two points of the deSitter space. We shall say that the xy plane outside the circle (6) has two sheets: the x, y point of the lower sheet corresponds to the point x, y, $\tau=-(x^2+y^2-a^2)^{\frac{1}{2}}$ of the hyperboloid, the x, y point of the upper sheet to the point x, y, $\tau=+(x^2+y^2-a^2)^{\frac{1}{2}}$ of the hyperboloid. The inside of the circle (6) does not exist for our purposes. The

[1] E. P. Wigner, Revs. Modern Phys. **29**, 255 (1957); Phys. Rev. **120**, 643 (1960).
[2] See, e.g., H. P. Robertson, Revs. Modern Phys. 5, 62 (1933). See, however, footnote 4.

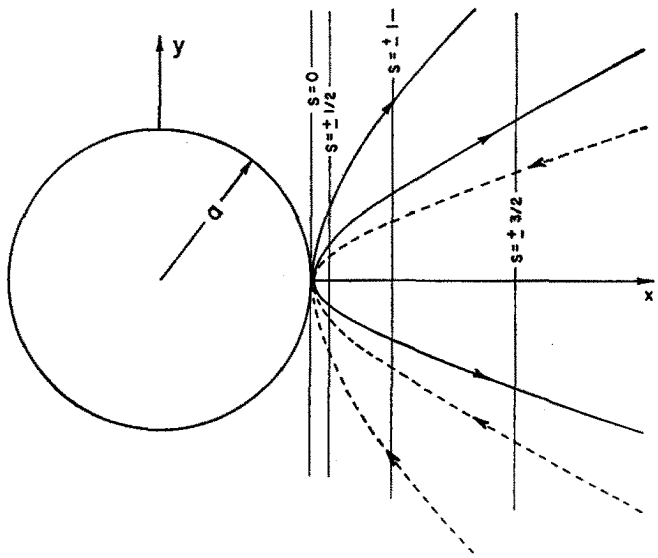


FIG. 1. Geodesics obtained by the Lorentz transformation (4) from the geodesic $y=0$. The geodesics are represented by their projections into the xy plane. Broken line indicates the lower sheet of plane ($\tau<0$); full line indicates the upper sheet ($\tau>0$). The proper time elapsed after the passage through the a,0 point is given by as; points of constant s lie on lines parallel to the y axis. Positive s values refer to the upper sheet; negative values refer to the lower sheet.

geodesic (3) is represented on the xy plane by the line segment

$$y=0, \quad x>a. \tag{7}$$

The image of a material body travels on this segment first on the lower sheet in the direction of decreasing x, then on the upper sheet in the direction of increasing x. The picture of the geodesic (5) in the xy plane is the hyperbola

$$x'^2-(y'^2/\sinh^2\varphi)=a^2, \tag{8}$$

which is traversed in the direction indicated by the arrow in Fig. 1, starting again on the lower sheet but passing into the upper one at the point a, 0. The distance of a point x, y from the point a, 0, as measured along the geodesic which connects the two points, is as, where s is the parameter of the point x, y on the geodesic (5). Since as can be expressed by x alone,

$$as=a \operatorname{arc\,cosh}(x/a)=a \ln[x+(x^2-a^2)^{\frac{1}{2}}]/a. \tag{9}$$

The points which are at equal distance from the a, 0 point lie on straight lines parallel to the y axis. Figure 1 gives the lines which represent points at distances 0, $\pm\frac{1}{2}a$, $\pm a$, $\pm\frac{3}{2}a$ from the point a, 0. The square root in (9) must be taken with the negative sign on the lower sheet; the corresponding s are then also negative.

It follows from the development of the last paragraph that the points at distance 0 from a, 0 lie on the line which is tangent to the circle (6) at $x=a$, $y=0$. The two light signals (our space has only one spatial dimension) which pass through the point a, 0, travel in opposite directions along this line, both passing at a, 0 from the lower into the upper sheet. Their paths are straight lines on the hyperboloid, called rulings. All other light signals are represented on our diagram by tangents of the circle (6); they can be obtained from the two light signals just described by rotations in the xy plane. All the preceding results have been established before and are well known; they are repeated here for the reader's convenience.

LIGHT SIGNALS BETWEEN TWO MATERIAL BODIES

The world lines of two material bodies in the same space-time plane can either intersect or not. The latter alternative would mean, in a flat space, that they are at rest with respect to each other—which is an exceptional situation. The same is not true in hyperbolic space and two timelike geodesics can approach each other and again separate without ever intersecting. The geodesic (3) and any other geodesic obtained from it by a rotation in the xy plane are in this relation. Hence, the case of nonintersection is not an exceptional one in the hyperbolic case. Nevertheless, only the case of two intersecting timelike geodesics will be considered in the present note. This will be done, partly for reasons of space, and partly because the earlier publications[1] are principally concerned with nonintersecting world lines.

The intersection of two world lines, as every point on the hyperboloid, can be brought into the point $x=a$, $y=\tau=0$ by a transformation which leaves the F of Eq. (2) invariant. The two world lines will then have parametric representations of the form given in (5), with different hyperbolic angles $\varphi=\varphi_1$ and $\varphi=\varphi_2$. A further transformation of the type used in (5), but with the angle $-\frac{1}{2}(\varphi_1+\varphi_2)$ will then change φ_1 into $\frac{1}{2}(\varphi_1-\varphi_2)$ and φ_2 into $\frac{1}{2}(\varphi_2-\varphi_1)$. Denoting the former angle by $\frac{1}{2}\varphi$, the parametric representations of the two geodesics become

$$\begin{aligned} x_1&=x_2=a\cosh s \\ y_1&=-y_2=a\sinh\tfrac{1}{2}\varphi\sinh s \\ \tau_1&=\tau_2=a\cosh\tfrac{1}{2}\varphi\sinh s. \end{aligned} \tag{10}$$

The transformations just carried out shift the point of coincidence of the two bodies to the point $x=a$, $y=\tau=0$, and reduce the velocity of their center of mass to zero. These transformations give the two world lines a convenient form; they do not change any of their invariant properties. In particular, they do not affect relations between traveling times of light quanta passing back and forth between them, if these traveling times are proper times measured along the world lines themselves.

The images of the two world lines in the xy plane now coincide and form the hyperbola

$$x^2-(y^2/\sinh^2\tfrac{1}{2}\varphi)=a^2, \tag{11}$$

which is traversed, by the two bodies, in opposite directions. Before they reach the $x=a$, $y=0$ point, both are on the lower sheet of the xy plane; they pass at that point into the upper sheet. The hyperbolic angle of their relative velocity at coincidence is φ.

The image of a light signal between the two bodies is that segment of a tangent to the circle (6) which lies inside the hyperbola. It is on the lower sheet of the xy plane, if the emission takes place before the coincidence of the two particles, i.e., if the images of the particles are themselves on the lower sheet. In this case, the light signal travels toward the circle. If the emission of the light signal takes place after the coincidence of the two particles, their image is on the upper sheet and so is the image of the light signal which travels, in this case, away from the circle.

Figure 2 illustrates the construction of a series of light signals, each emitted from one of the bodies when the previous signal, emitted by the other body, arrives. Naturally, there must be a first light signal which initiates the series. The images of these light signals form a polygon which is circumscribed around the circle (6) and is inscribed into the hyperbola (11). If the emission of the first signal takes place before the two bodies meet, the series has no end and infinitely many light signals can be exchanged before the two bodies come to coincidence. If the first light signal is emitted after this coincidence, only a finite number of light signals can be exchanged before the two bodies "recede under each other's horizon," and thus lose contact. In order to obtain the successive corners of the polygon, one has to draw, from the corner last obtained, the tangent to the circle which corresponds to the emission of a signal, and bring this tangent into intersection with the other branch of the hyperbola. The following section will describe the properties of this polygon; in particular, it will give an expression for the proper times at which the successive light quanta are received.

ARRIVAL TIMES OF THE LIGHT SIGNALS

The circle and hyperbola of Fig. 2 have two pairs of coincident points in common. They are, therefore, in

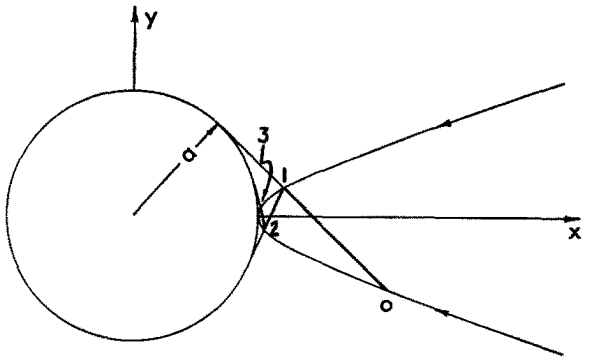


FIG. 2. Construction of light signals between two bodies approaching each other with equal velocities. All points of the hyperbola are on the lower sheet, i.e., all events take place before the bodies reach the a,0 point. If the convention of Fig. 1 had been followed, all lines would be broken. The images of light signals are tangents to the circle; the signal emitted at 0 reaches the second body at 1; the signal emitted at 1 reaches the first body at 2, and so on.

the terminology of projective geometry,[3] members of a pencil of conics, as well as members of a range of conics. It is well known,[4] that there exists a one parametric manifold of projective transformations which leave two such conics invariant, and that the points of coincidence are fixed points of these transformations. In our case, the projective transformations in question are

$$x'=a\frac{x\cosh\chi-a\sinh\chi}{-x\sinh\chi+a\cosh\chi}, \qquad y'=a\frac{y}{-x\sinh\chi+a\cosh\chi}, \tag{12}$$

with an arbitrary χ. Incidentally, these transformations also leave all other conics of the pencil or range invariant, i.e., all conics

$$x^2\pm y^2/c^2=a^2, \tag{13}$$

but this is of no significance for our discussion. Note, however, that for positive x (since $x>a$ on the hyperbola and $\cosh\chi>\sinh\chi$ for all real χ), the numerator of the expression for x' will be positive. Hence, if we restrict

[3] See, e.g., C. W. O'Hara and D. R. Ward, *An Introduction to Projective Geometry* (Clarendon Press, Oxford, England, 1937), p. 126 ff.

[4] It is not easy to find an explicit statement of this theorem in the literature. See, however, O. Veblen and J. W. Young, *Projective Geometry* (Ginn and Company, Boston, Massachusetts, 1910), Chap. X, or Figure 6.8C of H. S. M. Coxeter's *The Real Projective Plane* (Cambridge University Press, Cambridge, England, 1955). The conics of this figure are all tangent to the lines AP and AQ at the points P and Q respectively; these points, as well as A, are fixed points of the transformations in question. They correspond to our points $(a,0)$, $(-a,0)$, $(0,\infty)$. The transformations are the products of the harmonic homologies with axis PQ and center A, and with axis AB and center C. The position of the point B is the free parameter; C is the harmonic conjugate of B with respect to P and Q. I am indebted to Dr. Coxeter for this reference. However, no general proof of the theorem will be given since the transformations in question are exhibited explicitly in our Eq. (12).

ourselves to transformations which do not interchange the two branches of the hyperbola, we must restrict χ to such values that the denominators in Eqs. (12) are also positive.

In addition to Eqs. (12), the conics are evidently invariant also under the reflections

$$x'=x \quad y'=-y \tag{12a}$$

$$x'=-x \quad y'=y \tag{12b}$$

$$x'=-x \quad y'=-y. \tag{12c}$$

Only Eq. (12a), or rather, the combination of Eqs. (12) and (12a), will be used in what follows.

If we subject a tangent to the circle to a transformation (12) or a combination of Eqs. (12) and (12a), the resulting line will still be a tangent to the circle. Furthermore, the points of intersection with the hyperbola will be transformed into points of intersection with the hyperbola. Hence Eq. (12) transforms a light signal from the first body to the second body into a similar light signal. If we want to transform a light signal from the first body to the second into a light signal from the second body into the first, we must use a combination of Eqs. (12) and (12a), i.e., write

$$x'=a\frac{x\cosh\chi-a\sinh\chi}{-x\sinh\chi+a\cosh\chi}, \qquad y'=\frac{-ay}{-x\sinh\chi+a\cosh\chi}. \tag{14}$$

By choosing χ appropriately, the point of intersection of the light signal with the upper branch of the hyperbola can be transformed into the point of intersection of the light signal with the lower branch. Since this point is the starting point of the next (responding) light signal, Eqs. (14) transform the equation of each light signal into the equation of the next light signal. They also transform the point of emission of each light signal into the point of emission of the next light signal. The same applies for the points of arrival. In fact, the whole polygon is transformed onto itself by the transformation (14) with the proper χ.

The proper value of χ remains to be determined. This can be obtained from the condition that the intersection of a tangent to the circle

$$\xi x+\eta y=a^2 \quad (\text{where } \xi^2+\eta^2=a^2), \tag{15}$$

with the transform of this tangent

$$a\xi\frac{x\cosh\chi-a\sinh\chi}{-x\sinh\chi+a\cosh\chi}-a\eta\frac{y}{-x\sinh\chi+a\cosh\chi}=a^2 \tag{15a}$$

lie on the hyperbola (11). The point of intersection of the lines (15) and (15a) is

$$x=a\frac{a(1+\cosh\chi)+\xi\sinh\chi}{a\sinh\chi+\xi(1+\cosh\chi)}, \qquad y=\frac{a\eta\sinh\chi}{a\sin\chi+\xi(1+\cosh\chi)}. \tag{16}$$

With $\xi^2+\eta^2=a^2$, the condition that the point (16) lie on the hyperbola (11), reduces to

$$\sinh^2\tfrac{1}{2}\varphi=\tfrac{1}{2}(\cosh\chi-1) \tag{17}$$

or

$$\chi=\pm\varphi. \tag{17a}$$

The upper sign holds on the upper sheet where the image of the light signal moves away from the circle so that $|y'|>|y|$; the lower sign holds on the lower sheet where the opposite is true. It would seem though that it should be possible to obtain Eq. (17a) with less computation than Eqs. (15)–(17) imply.

Let us denote the coordinates of the point at which the first light signal is emitted by x_0, y_0, τ_0. The image of this point on the xy plane is x_0, y_0, and the image of the arrival point of this signal, which is also the departure point of the second signal, can be obtained by the transformation (14) with $\chi=\pm\varphi$. The coordinates of the image of the arrival point of the signal $n-1$, which is also the departure point of the signal n, will be denoted by x_n, y_n. This can be obtained from x_0, y_0 by the transformation which is the n-fold iterate of the transformation (14), with $\chi=\pm\varphi$. Hence,

$$x_n=a\frac{x_0\cosh n\varphi\mp a\sinh n\varphi}{\mp x_0\sinh n\varphi+a\cosh n\varphi} \tag{18}$$

$$y_n=\frac{(-)^n ay_0}{\mp x_0\sinh n\varphi+a\cosh n\varphi}. \tag{18a}$$

The corresponding τ_n can be calculated from expression (1):

$$\tau_n=\frac{a\tau_0}{\mp x_0\sinh n\varphi+a\cosh n\varphi}. \tag{18a}$$

As was mentioned before, n can grow indefinitely if the first signal is emitted before the bodies come to a coincidence, i.e., if one is on the lower sheet of the xy plane and uses the lower sign in Eqs. (18). The number n has an upper limit, if the signals are emitted after the coincidence; x_n becomes negative for larger n. The points with even n represent arrival and departure points at the first body; the points with odd n refer to events at the second body.

Finally, we calculate the proper time t_{2n} which a clock on the first body would attribute to the nth arrival of a signal. If the clock measures the time from the time at which the two bodies are in coincidence, its time will be the same as of Eq. (9). Hence,

$$t_{2n}=a\ln q_{2n}, \tag{19}$$

where

$$q_{2n}=a^{-1}[x_{2n}+(x_{2n}^2-a^2)^{\frac{1}{2}}]. \tag{19a}$$

It follows from Eq. (19a) that

$$x_{2n}=\tfrac{1}{2}a(q_{2n}+1/q_{2n}). \tag{20}$$

Clearly, x_{2n} and q_{2n} mutually determine each other. Hence, the equation

$$q_{2n}=\frac{q_0\cosh n\varphi\mp\sinh n\varphi}{\mp q_0\sinh n\varphi+\cosh n\varphi} \tag{21}$$

will be established, if the x_{2n} calculated from it by Eq. (20) becomes equal to the expression (18) obtained earlier. Hence, we calculate

$$\begin{aligned}&\tfrac{1}{2}a(q_{2n}+1/q_{2n})\\&=\tfrac{1}{2}a\frac{(q_0\cosh n\varphi\mp\sinh n\varphi)^2+(\mp q_0\sinh n\varphi+\cosh n\varphi)^2}{(\mp q_0\sinh n\varphi+\cosh n\varphi)(q_0\cosh n\varphi\mp\sinh n\varphi)}\\&=\tfrac{1}{2}a\frac{(1+q_0^2)\cosh 2n\varphi\mp 2q_0\sinh 2n\varphi}{\mp\tfrac{1}{2}(1+q_0^2)\sinh 2n\varphi+q_0\cosh 2n\varphi}.\end{aligned} \tag{22}$$

Multiplying numerator and denominator by (a/q_0) gives, since q_0 is so defined that Eq. (20) holds there for,

$$\tfrac{1}{2}a(q_{2n}+1/q_{2n})=a\frac{x_0\cosh 2n\varphi\mp a\sinh 2n\varphi}{\mp x_0\sinh 2n\varphi+a\cosh 2n\varphi}=x_{2n}. \tag{22a}$$

This then verifies Eq. (21), and gives the rather simple expression

$$t_{2n}=a\ln\frac{e^{t_0/a}\cosh n\varphi\mp\sinh n\varphi}{\mp e^{t_0/a}\sinh n\varphi+\cosh n\varphi} \tag{23}$$

for the time of the nth event at the first body. The lower sign applies if these events precede the coincidence of the two bodies; the upper sign applies in the opposite case.

It should be mentioned perhaps that the earlier[1] publications, t_1, t_2, and t_3 are, in the present notation, t_2-t_0, t_4-t_2, and t_6-t_4.

Lorentz Invariance and the Gravitational Field

T. W. B. KIBBLE*
Department of Mathematics, Imperial College, London, England
(Received August 19, 1960)

An argument leading from the Lorentz invariance of the Lagrangian to the introduction of the gravitational field is presented. Utiyama's discussion is extended by considering the 10-parameter group of inhomogeneous Lorentz transformations, involving variation of the coordinates as well as the field variables. It is then unnecessary to introduce *a priori* curvilinear coordinates or a Riemannian metric, and the new field variables introduced as a consequence of the argument include the vierbein components $h_k{}^\mu$ as well as the "local affine connection" $A^i{}_{j\mu}$. The extended transformations for which the 10 parameters become arbitrary functions of position may be interpreted as general coordinate transformations and rotations of the vierbein system. The free Lagrangian for the new fields is shown to be a function of two covariant quantities analogous to $F_{\mu\nu}$ for the electromagnetic field, and the simplest possible form is just the usual curvature scalar density expressed in terms of $h_k{}^\mu$ and $A^i{}_{j\mu}$. This Lagrangian is of first order in the derivatives, and is the analog for the vierbein formalism of Palatini's Lagrangian. In the absence of matter, it yields the familiar equations $R_{\mu\nu}=0$ for empty space, but when matter is present there is a difference from the usual theory (first pointed out by Weyl) which arises from the fact that $A^i{}_{j\mu}$ appears in the matter field Lagrangian, so that the equation of motion relating $A^i{}_{j\mu}$ to $h_k{}^\mu$ is changed. In particular, this means that, although the covariant derivative of the metric vanishes, the affine connection $\Gamma^\lambda{}_{\mu\nu}$ is nonsymmetric. The theory may be reexpressed in terms of the Christoffel connection, and in that case additional terms quadratic in the "spin density" $S^k{}_{ij}$ appear in the Lagrangian. These terms are almost certainly too small to make any experimentally detectable difference to the predictions of the usual metric theory.

1. INTRODUCTION

IT has long been realized that the existence of certain fields, notably the electromagnetic field, can be related to invariance properties of the Lagrangian.[1] Thus, if the Lagrangian is invariant under phase transformations $\psi \rightarrow e^{ie\lambda}\psi$, and if we wish to make it invariant under the general gauge transformations for which λ is a function of x, then it is necessary to introduce a new field A_μ which transforms according to $A_\mu \rightarrow A_\mu - \partial_\mu\lambda$, and to replace $\partial_\mu\psi$ in the Lagrangian by a "covariant derivative" $(\partial_\mu + ieA_\mu)\psi$. A similar argument has been applied by Yang and Mills[2] to isotopic spin rotations, and in that case yields a triplet of vector fields. It is thus an attractive idea to relate the existence of the gravitational field to the Lorentz invariance of the Lagrangian. Utiyama[3] has proposed a method which leads to the introduction of 24 new field variables $A^{ij}{}_\mu$ by considering the homogeneous Lorentz transformations specified by six parameters ϵ^{ij}. However, in order to do this it was necessary to introduce *a priori* curvilinear coordinates and a set of 16 parameters $h_k{}^\mu$. Initially, the $h_k{}^\mu$ were treated as given functions of x, but at a later stage they were regarded as field variables and interpreted as the components of a vierbein system in a Riemannian space. This is a rather unsatisfactory procedure since it is the purpose of the discussion to supply an argument for introducing the gravitational field variables, which include the metric as well as the affine connection. The new field variables $A^{ij}{}_\mu$ were subsequently related to the Christoffel connection $\Gamma^\lambda{}_{\mu\nu}$ in the Riemannian space, but this could only be done uniquely by making the *ad hoc* assumption that the quantity $\Gamma'^\lambda{}_{\mu\nu}$ calculated from $A^{ij}{}_\mu$ was symmetric.

It is the purpose of this paper to show that the vierbein components $h_k{}^\mu$, as well as the "local affine connection" $A^{ij}{}_\mu$, can be introduced as new field variables analogous to A_μ if one considers the full 10-parameter group of inhomogeneous Lorentz transformations in place of the restricted six-parameter group. This implies that one must consider transformations of the coordinates as well as the field variables, which will necessitate some changes in the argument, but it also means that only one system of coordinates is required, and that a Riemannian metric need not be introduced *a priori*. The interpretation of the theory in terms of a Riemannian space may be made later if desired. The starting point of the discussion is the ordinary formulation of Lorentz invariance (including translational invariance) in terms of rectangular coordinates in flat space. We shall follow the analogy with gauge transformations as far as possible, and for purposes of comparison we give in Sec. 2 a brief discussion of linear transformations of the field variables. This is essentially a summary of Utiyama's argument, though the emphasis is rather different, particularly with regard to the covariant and noncovariant conservation laws.

In Sec. 3 we discuss the invariance under Lorentz transformations, and in Sec. 4 we extend the discussion to the corresponding group in which the ten parameters become arbitrary functions of position. We show that to maintain invariance of the Lagrangian, it is necessary to introduce 40 new variables so that a suitable covariant derivative may be constructed. To make the action integral invariant, one actually requires the Lagrangian to be an invariant density rather than an invariant, and one must, therefore, multiply the invariant by a suitable (and uniquely determined) function of the new fields. In Sec. 5 we consider the possible forms of the free Lagrangian for the new fields. As in the case of the

* NATO Research Fellow.

[1] See, for example, H. Weyl, *Gruppentheorie und Quantenmechanik* (S. Hirzel, Leipzig, 1931), 2nd ed., Chap. 2, p. 89; and earlier references cited there.

[2] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[3] Ryoyu Utiyama, Phys. Rev. **101**, 1597 (1956).

electromagnetic field, we choose the Lagrangian of lowest degree which satisfies the invariance requirements.

The geometrical interpretation in terms of a Riemannian space is discussed in Sec. 6, where we show that the free Lagrangian we have obtained is just the usual curvature scalar density, though expressed in terms of an affine connection $\Gamma^{\lambda}{}_{\mu\nu}$ which is not necessarily symmetric. In fact, when no matter is present it is symmetric as a consequence of the equations of motion, but otherwise it has an antisymmetric part expressible in terms of the "spin density" $\mathfrak{S}^{\mu}{}_{ij}$. Thus there is a difference between this theory and the usual metric theory of gravitation. This difference was first pointed out by Weyl,[4] and has more recently been discussed by Sciama.[5] It arises from the fact that our free Lagrangian is of first order in the derivatives, with the $h_k{}^{\mu}$ and $A^{ij}{}_{\mu}$ as independent variables. It is possible to re-express the theory in terms of the Christoffel connection ${}^0\Gamma^{\lambda}{}_{\mu\nu}$ or its local analog ${}^0A^{ij}{}_{\mu}$, and this is done in Sec. 7. In that case, additional terms quadratic in $\mathfrak{S}^{\mu}{}_{ij}$, and multiplied by the gravitational constant, appear in the Lagrangian.

2. LINEAR TRANSFORMATIONS

We consider a set of field variables $\chi_A(x)$, which we regard as the elements of a column matrix $\chi(x)$, with the Lagrangian

$$L(x)\equiv L\{\chi(x),\chi_{,\mu}(x)\},$$

where $\chi_{,\mu}=\partial_\mu\chi$. We also consider linear transformations of the form

$$\delta\chi=\epsilon^a T_a\chi, \tag{2.1}$$

where the ϵ^a are n constant infinitesimal parameters, and the T_a are n given matrices satisfying commutation rules appropriate to the generators of a Lie group,

$$[T_a,T_b]=f_a{}^c{}_b T_c.$$

The Lagrangian is invariant under these transformations if the n identities

$$(\partial L/\partial\chi)T_a\chi+(\partial L/\partial\chi_{,\mu})T_a\chi_{,\mu}\equiv 0, \tag{2.2}$$

are satisfied, and we shall assume that this is so. Note that $\partial/\partial\chi$ must be regarded as a row matrix. The equations of motion imply n conservation laws

$$J^{\mu}{}_{a,\mu}=0,$$

where the "currents" are defined by[6]

$$J^{\mu}{}_a\equiv-(\partial L/\partial\chi_{,\mu})T_a\chi. \tag{2.3}$$

Now, under the more general transformations of the form (2.1), but in which the parameters ϵ^a become arbitrary functions of position, the Lagrangian is no longer invariant, because the derivatives transform according to

$$\delta\chi_{,\mu}=\epsilon^a T_a\chi_{,\mu}+\epsilon^a{}_{,\mu}T_a\chi, \tag{2.4}$$

and the terms in $\epsilon^a{}_{,\mu}$ do not cancel. In fact, one finds

$$\delta L\equiv-\epsilon^a{}_{,\mu}J^{\mu}{}_a.$$

However, one can obtain a modified Lagrangian which is invariant by replacing $\chi_{,\mu}$ in L by a quantity $\chi_{;\mu}$ which transforms according to

$$\delta\chi_{;\mu}=\epsilon^a T_a\chi_{;\mu}. \tag{2.5}$$

To do this[7] it is necessary to introduce $4n$ new field variables $A^a{}_{\mu}$ whose transformation properties involve $\epsilon^a{}_{,\mu}$. In fact, if one takes

$$\chi_{;\mu}\equiv\chi_{,\mu}+A^a{}_{\mu}T_a\chi, \tag{2.6}$$

then the condition (2.5) determines the transformation properties of the new fields uniquely. They are

$$\delta A^a{}_{\mu}=\epsilon^b f_b{}^a{}_c A^c{}_{\mu}-\epsilon^a{}_{,\mu}. \tag{2.7}$$

In this way one obtains the invariant Lagrangian

$$L'\{\chi,\chi_{,\mu},A^a{}_{\mu}\}\equiv L\{\chi,\chi_{;\mu}\}.$$

The expression $\chi_{;\mu}$ may be called the covariant derivative of χ with respect to the transformations (2.1). One may define covariant currents by

$$J'^{\mu}{}_a\equiv-(\partial L'/\partial A^a{}_{\mu})\equiv-(\partial L/\partial\chi_{;\mu})T_a\chi, \tag{2.8}$$

where L is regarded as a function of χ and $\chi_{;\mu}$. They transform linearly according to

$$\delta J'^{\mu}{}_a=-\epsilon^b f_b{}^c{}_a J'^{\mu}{}_c,$$

and their covariant divergences vanish in virtue of the equations of motion and the identities (2.2):

$$J'^{\mu}{}_{a;\mu}\equiv J'^{\mu}{}_{a,\mu}-A^b{}_{\mu}f_b{}^c{}_a J'^{\mu}{}_c$$
$$=0.$$

Two covariant differentiations do not in general commute. From (2.6) one finds

$$\chi_{;\mu\nu}-\chi_{;\nu\mu}=F^a{}_{\mu\nu}T_a\chi,$$

where

$$F^a{}_{\mu\nu}\equiv A^a{}_{\mu,\nu}-A^a{}_{\nu,\mu}-f_b{}^a{}_c A^b{}_{\mu}A^c{}_{\nu}. \tag{2.9}$$

Unlike $A^a{}_{\mu}$, the expression $F^a{}_{\mu\nu}$ is a covariant quantity transforming according to

$$\delta F^a{}_{\mu\nu}=\epsilon^b f_b{}^a{}_c F^c{}_{\mu\nu},$$

and one may, therefore, define its covariant derivative in an obvious manner. It satisfies the cyclic identity

$$F^a{}_{\mu\nu;\rho}+F^a{}_{\nu\rho;\mu}+F^a{}_{\rho\mu;\nu}\equiv 0.$$

[4] H. Weyl, Phys. Rev. 77, 699 (1950).

[5] D. W. Sciama, *Festschrift for Infeld* (Pergamon Press, New York), to be published.

[6] We have defined $J^{\mu}{}_a$ with the opposite sign to that used by Utiyama.[3] This is because with this choice of sign the analogous quantity for translations is $T^{\mu}{}_{\nu}$ rather than $-T^{\mu}{}_{\nu}$. The change may be considered as a change of sign of ϵ^a and T_a, and there is a corresponding change of sign in (2.6). This convention has the additional advantage that the "local affine connection" $A^i{}_{j\mu}$ defined in Sec. 4 specifies covariant derivatives according to the same rule as $\Gamma^{\lambda}{}_{\mu\nu}$.

[7] For a full discussion, see footnote 3.

It remains to find a free Lagrangian L_0 for the new fields. Clearly L_0 must be separately invariant, and it is easy to see[3] that this implies that it must contain $A^a{}_\mu$ only through the covariant combination $F^a{}_{\mu\nu}$. The simplest such Lagrangian is[8]

$$L_0=-\tfrac{1}{4}F^a{}_{\mu\nu}F_a{}^{\mu\nu}, \tag{2.10}$$

where the tensor indices are raised with the flat-space metric $\eta^{\mu\nu}$ with diagonal elements $(1,-1,-1,-1)$, and the index a is lowered with the metric[8a]

$$g_{ab}\equiv f_a{}^c{}_d f_c{}^d{}_b$$

associated with the Lie group (except of course for a one-parameter group). It is clear that this Lagrangian is not unique. All that is required is that it should be a scalar both in coordinate space and in the Lie-group space, and one could add to it terms of higher degree in $F^a{}_{\mu\nu}$. However, it seems reasonable to choose the Lagrangian of lowest degree which satisfies the invariance requirements.

With the choice (2.10) of L_0, the equations of motion for the new fields are

$$F_a{}^{\mu\nu}{}_{;\nu}=J'^\mu{}_a.$$

Because of the antisymmetry of $F_a{}^{\mu\nu}$ one can define another current which is conserved in the strict sense:

$$(J'^\mu{}_a+j^\mu{}_a)_{,\mu}=0, \tag{2.11}$$

where

$$j^\mu{}_a\equiv A^b{}_\nu f_b{}^c{}_a F_c{}^{\mu\nu}.$$

This extra current $j^\mu{}_a$ may be regarded as the current of the new field $A^a{}_\mu$ itself, since it is expressible in the form

$$j^\mu{}_a\equiv-(\partial L_0/\partial A^a{}_\mu)\equiv-(\partial L_0/\partial A^b{}_{\nu,\mu})f_a{}^b{}_c A^c{}_\nu, \tag{2.12}$$

which should be compared with (2.8). Note, however, that it is not a covariant quantity. To obtain a strict conservation law one must sacrifice the covariance of the current.

3. LORENTZ TRANSFORMATIONS

We now wish to consider infinitesimal variations of both the coordinates and the field variables,

$$\begin{aligned} x^\mu &\longrightarrow x'^\mu=x^\mu+\delta x^\mu,\\ \chi(x) &\longrightarrow \chi'(x')=\chi(x)+\delta\chi(x). \end{aligned} \tag{3.1}$$

It will be convenient to allow for the possibility that the Lagrangian may depend on x explicitly. Then, under a variation (3.1), the change in L is

$$\delta L\equiv(\partial L/\partial\chi)\delta\chi+(\partial L/\partial\chi_{,\mu})\delta\chi_{,\mu}+(\partial L/\partial x^\mu)\delta x^\mu,$$

where $\partial L/\partial x^\mu$ denotes the partial derivative with fixed χ. It is sometimes useful to consider also the variation at a fixed value of x,

$$\delta_0\chi=\chi'(x)-\chi(x)=\delta\chi-\delta x^\mu\chi_{,\mu}. \tag{3.2}$$

In particular, it is obvious that δ_0 commutes with ∂_μ, whence

$$\delta\chi_{,\mu}=(\delta\chi)_{,\mu}-(\delta x^\nu)_{,\mu}\chi_{,\nu}. \tag{3.3}$$

The action integral

$$I(\Omega)\equiv\int_\Omega L(x)d_4x$$

over a space-time region Ω is transformed under (3.1) into

$$I'(\Omega)\equiv\int_\Omega L'(x')\|\partial_\nu x'^\mu\|d_4x.$$

Thus the action integral over an arbitrary region is invariant if[9]

$$\delta L+L(\delta x^\mu)_{,\mu}\equiv\delta_0 L+(L\delta x^\mu)_{,\mu}\equiv 0. \tag{3.4}$$

This is of course the typical transformation law of an invariant density.

We now consider the specific case of Lorentz transformations,

$$\delta x^\mu=\epsilon^\mu{}_\nu x^\nu+\epsilon^\mu,\quad \delta\chi=\tfrac{1}{2}\epsilon^{\mu\nu}S_{\mu\nu}\chi, \tag{3.5}$$

where ϵ^μ and $\epsilon^{\mu\nu}=-\epsilon^{\nu\mu}$ are 10 real infinitesimal parameters, and the $S_{\mu\nu}$ are matrices satisfying

$$\begin{aligned} &S_{\mu\nu}+S_{\nu\mu}=0,\\ &[S_{\mu\nu},S_{\rho\sigma}]=\eta_{\nu\rho}S_{\mu\sigma}+\eta_{\mu\sigma}S_{\nu\rho}-\eta_{\nu\sigma}S_{\mu\rho}-\eta_{\mu\rho}S_{\nu\sigma}\equiv\tfrac{1}{2}f_{\mu\nu}{}^{\kappa\lambda}{}_{\rho\sigma}S_{\kappa\lambda}. \end{aligned}$$

From (3.3) one has

$$\delta\chi_{,\mu}=\tfrac{1}{2}\epsilon^{\rho\sigma}S_{\rho\sigma}\chi_{,\mu}-\epsilon^\rho{}_\mu\chi_{,\rho}. \tag{3.6}$$

Moreover, since $(\delta x^\mu)_{,\mu}=\epsilon^\mu{}_\mu=0$, the condition (3.4) for invariance of the action integral again reduces to $\delta L\equiv 0$, and yields the 10 identities[10]

$$\partial L/\partial x^\rho\equiv L_{,\rho}-(\partial L/\partial\chi)\chi_{,\rho}-(\partial L/\partial\chi_{,\mu})\chi_{,\mu\rho}\equiv 0, \tag{3.7}$$

$$\begin{aligned}(\partial L/\partial\chi)S_{\rho\sigma}\chi+(\partial L/\partial\chi_{,\mu})(S_{\rho\sigma}\chi_{,\mu}\\ +\eta_{\mu\rho}\chi_{,\sigma}-\eta_{\mu\sigma}\chi_{,\rho})\equiv 0.\end{aligned} \tag{3.8}$$

These are evidently the analogs of the identities (2.2), and we shall assume that they are satisfied. Note that (3.7), which express the conditions for translational invariance, are equivalent to the requirement that L be explicitly independent of x, as might be expected.

As before, the equations of motion may be used to obtain 10 conservation laws which follow from these identities, namely,

$$T^\mu{}_{\rho,\mu}=0,\quad (S^\mu{}_{\rho\sigma}-x_\rho T^\mu{}_\sigma+x_\sigma T^\mu{}_\rho)_{,\mu}=0,$$

[8] There could of course be a constant factor multiplying (2.10), but this can be absorbed by a trivial change of definition of $A^a{}_\mu$ and T_a.

[8a] The discussion here applies only to semisimple groups since otherwise g_{ab} is singular. (I am indebted to the referee for this remark.)

[9] See L. Rosenfeld, Ann. Physik 5, 113 (1930).

[10] Compare L. Rosenfeld, Ann. inst. Henri Poincaré 2, 25 (1931).

where

$$T^{\mu}{}_{\rho}\equiv(\partial L/\partial\chi_{,\mu})\chi_{,\rho}-\delta^{\mu}{}_{\rho}L,\quad S^{\mu}{}_{\rho\sigma}\equiv-(\partial L/\partial\chi_{,\mu})S_{\rho\sigma}\chi.$$

These are the conservation laws of energy, momentum, and angular momentum.

It is instructive to examine these transformations in terms of the variation $\delta_0\chi$ also, which in this case is

$$\delta_0\chi=-\epsilon^{\rho}\partial_{\rho}\chi+\tfrac{1}{2}\epsilon^{\rho\sigma}(S_{\rho\sigma}+x_{\rho}\partial_{\sigma}-x_{\sigma}\partial_{\rho})\chi.$$

On comparing this with (2.1), one sees that the role of the matrices T_a is played by the differential operators $-\partial_{\mu}$ and $S_{\rho\sigma}+x_{\rho}\partial_{\sigma}-x_{\sigma}\partial_{\rho}$. Thus, by analogy with the definition (2.3) of the currents $J^{\mu}{}_{a}$, one might expect the currents in this case to be

$$J^{\mu}{}_{\rho}\equiv(\partial L/\partial\chi_{,\mu})\chi_{,\rho},\quad J^{\mu}{}_{\rho\sigma}\equiv S^{\mu}{}_{\rho\sigma}-x_{\rho}J^{\mu}{}_{\sigma}+x_{\sigma}J^{\mu}{}_{\rho},$$

corresponding to the parameters ϵ^{ρ}, $\epsilon^{\rho\sigma}$, respectively. However, in terms of δ_0, the condition for invariance (3.4) is not simply $\delta_0 L\equiv 0$, and the additional term $\delta x^{\rho}L_{,\rho}$ is responsible for the appearance of the term $L_{,\rho}$ in the identities (3.7), and hence for the term $\delta^{\mu}{}_{\rho}L$ in $T^{\mu}{}_{\rho}$.

4. GENERALIZED LORENTZ TRANSFORMATIONS

We now turn to a consideration of the generalized transformations (3.5) in which the parameters ϵ^{μ} and $\epsilon^{\mu\nu}$ become arbitrary functions of position. It is more convenient, and clearly equivalent, to regard as independent functions $\epsilon^{\mu\nu}$ and

$$\xi^{\mu}\equiv\epsilon^{\mu}{}_{\nu}x^{\nu}+\epsilon^{\nu},$$

since this avoids the explicit appearance of x. Moreover, one could consider generalized transformations with $\xi^{\mu}=0$ but nonzero $\epsilon^{\mu\nu}$, so that the coordinate and field transformations can be completely separated. In view of this fact, it is convenient to use Latin indices for ϵ^{ij} (and for the matrices S_{ij}), retaining the Greek ones for ξ^{μ} and x^{μ}. Thus the transformations under consideration are

$$\delta x^{\mu}=\xi^{\mu},\quad \delta\chi=\tfrac{1}{2}\epsilon^{ij}S_{ij}\chi \tag{4.1}$$

or

$$\delta_0\chi=-\xi^{\mu}\chi_{,\mu}+\tfrac{1}{2}\epsilon^{ij}S_{ij}\chi. \tag{4.2}$$

This notation emphasizes the similarity of the ϵ^{ij} transformations to the linear transformations discussed in Sec. 2. These transformations alone were considered by Utiyama.[3] Evidently, the four functions ξ^{μ} specify a general coordinate transformation. The geometrical significance of the ϵ^{ij} will be discussed in Sec. 6.

According to our convention, the differential operator ∂_{μ} must have a Greek index. However, in the Lagrangian function L it would be inconvenient to have two kinds of indices, and we shall, therefore, regard L as a given function of χ and χ_k (no comma),[11] satisfying the identities (3.7) and (3.8). The original Lagrangian is then obtained by setting

$$\chi_k=\delta_k{}^{\mu}\chi_{,\mu}.$$

It is of course not invariant under the generalized transformations (4.1), but we shall later obtain an invariant expression by replacing χ_k by a suitable quantity $\chi_{;k}$.

The transformation of $\chi_{,\mu}$ is given by

$$\delta\chi_{,\mu}=\tfrac{1}{2}\epsilon^{ij}S_{ij}\chi_{,\mu}+\tfrac{1}{2}\epsilon^{ij}{}_{,\mu}S_{ij}\chi-\xi^{\nu}{}_{,\mu}\chi_{,\nu}, \tag{4.3}$$

and so the original Lagrangian transforms according to

$$\delta L\equiv-\xi^{\rho}{}_{,\mu}J^{\mu}{}_{\rho}-\tfrac{1}{2}\epsilon^{ij}{}_{,\mu}S^{\mu}{}_{ij}.$$

Note that it is $J^{\mu}{}_{\rho}$ rather than $T^{\mu}{}_{\rho}$ which appears here. The reason for this is that we have not included the extra term $L(\delta x^{\mu})_{,\mu}$ in (3.4). The left-hand side of (3.4) actually has the value

$$\delta L+L(\delta x^{\mu})_{,\mu}\equiv-\xi^{\rho}{}_{,\mu}T^{\mu}{}_{\rho}-\tfrac{1}{2}\epsilon^{ij}{}_{,\mu}S^{\mu}{}_{ij}.$$

We now look for a modified Lagrangian which makes the action integral invariant. The additional term just mentioned is of a different kind to those previously encountered, in that it involves L and not $\partial L/\partial\chi_k$. In particular, it includes contributions from terms in L which do not contain derivatives. Thus it is clear that we cannot remove it by replacing the derivative by a suitable covariant derivative. For this reason, we shall consider the problem in two stages. We first eliminate the noninvariance arising from the fact that $\chi_{,\mu}$ is not a covariant quantity, and thus obtain an expression L' satisfying

$$\delta L'\equiv 0. \tag{4.4}$$

Then, because the condition (3.4) for invariance of the action integral requires the Lagrangian to be an invariant density rather than an invariant, we make a further modification, replacing L' by $\mathfrak{L}'$, which satisfies

$$\delta\mathfrak{L}'+\xi^{\mu}{}_{,\mu}\mathfrak{L}'\equiv 0. \tag{4.5}$$

The first part of this program can be accomplished by replacing χ_k in L by a "covariant derivative" $\chi_{;k}$ which transforms according to

$$\delta\chi_{;k}=\tfrac{1}{2}\epsilon^{ij}S_{ij}\chi_{;k}-\epsilon^{i}{}_{k}\chi_{;i}. \tag{4.6}$$

The condition (4.4) then follows from the identities (3.8). To do this it is necessary to introduce forty new field variables. We consider first the ϵ^{ij} transformations, and eliminate the $\epsilon^{ij}{}_{,\mu}$ term in (4.3) by setting[12]

$$\chi_{|\mu}\equiv\chi_{,\mu}+\tfrac{1}{2}A^{ij}{}_{\mu}S_{ij}\chi, \tag{4.7}$$

where the $A^{ij}{}_{\mu}=-A^{ji}{}_{\mu}$ are 24 new field variables. We can then impose the condition

$$\delta\chi_{|\mu}=\tfrac{1}{2}\epsilon^{ij}S_{ij}\chi_{|\mu}-\xi^{\nu}{}_{,\mu}\chi_{|\nu} \tag{4.8}$$

which determines the transformation properties of $A^{ij}{}_{\mu}$

[11] Note that since we are using Latin indices for S_{ij} the various tensor components of χ must also have Latin indices, and for spinor components the Dirac matrices must be γ^k.

[12] Our $A^{ij}{}_{\mu}$ differs in sign from that of Utiyama.[3] Compare footnote 6.

uniquely. They are

$$\delta A^{ij}{}_{\mu}=\epsilon^{i}{}_{k}A^{kj}{}_{\mu}+\epsilon^{j}{}_{k}A^{ik}{}_{\mu}-\xi^{\nu}{}_{,\mu}A^{ij}{}_{\nu}-\epsilon^{ij}{}_{,\mu}. \quad (4.9)$$

The position with regard to the last term in (4.3) is rather different. The term involving $\epsilon^{ij}{}_{,\mu}$ is inhomogeneous in the sense that it contains χ rather than $\chi_{,\mu}$, just like the second term of (2.4), but this is not true of the last term.[13] Correspondingly, the transformation law (4.8) of $\chi_{|\mu}$ is already homogeneous. This means that to obtain an expression $\chi_{;k}$ transforming according to (4.6) we should add to $\chi_{|\mu}$ not a term in χ but rather a term in $\chi_{|\mu}$ itself. In other words, we can merely multiply by a new field:

$$\chi_{;k}\equiv h_{k}{}^{\mu}\chi_{|\mu}. \quad (4.10)$$

Here the $h_{k}{}^{\mu}$ are 16 new field variables with transformation properties determined by (4.6) to be

$$\delta h_{k}{}^{\mu}=\xi^{\mu}{}_{,\nu}h_{k}{}^{\nu}-\epsilon^{i}{}_{k}h_{i}{}^{\mu}. \quad (4.11)$$

It should be noted that the fields $h_{k}{}^{\mu}$ and $A^{ij}{}_{\mu}$ are quite independent and unrelated at this stage, though of course they will be related by equations of motion.

We have now found an invariant L'. We can easily obtain an invariant density $\mathfrak{L}'$ by multiplying by a suitable function of the fields already introduced:

$$\mathfrak{L}'\equiv\mathfrak{H}L'.$$

Then (4.5) is satisfied provided that $\mathfrak{H}$ is itself an invariant density,

$$\delta\mathfrak{H}+\xi^{\mu}{}_{,\mu}\mathfrak{H}\equiv 0.$$

It is easy to see that the only function of the new fields which obeys this transformation law, and does not involve derivatives, is

$$\mathfrak{H}=[\det(h_{k}{}^{\mu})]^{-1},$$

where the arbitrary constant factor has been chosen so that $\mathfrak{H}$ reduces to 1 when $h_{k}{}^{\mu}$ is set equal to $\delta_{k}{}^{\mu}$.[14]

The final form of our modified Lagrangian is

$$\mathfrak{L}\{\chi,\chi_{,\mu},h_{k}{}^{\mu},A^{ij}{}_{\mu}\}\equiv\mathfrak{H}L\{\chi,\chi_{;k}\}.$$

(We can drop the prime without risk of confusion.) It may be asked whether this Lagrangian is unique in the same sense as the modified Lagrangian L' of Sec. 2, and in fact it is easy to see that it is not. The reason for this is that if one starts with two Lagrangians L_1 and L_2 which differ by an explicit divergence, and are therefore equivalent, then the modified Lagrangians $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are not necessarily equivalent. Consider for example the Lagrangian for a real scalar field written in its first-order form

$$L_1=\pi^{k}\varphi_{,k}-\tfrac{1}{2}\pi^{k}\pi_{k}-\tfrac{1}{2}m^{2}\varphi^{2}. \quad (4.12)$$

This is equivalent to

$$L_2=-\pi^{k}{}_{,k}\varphi-\tfrac{1}{2}\pi^{k}\pi_{k}-\tfrac{1}{2}m^{2}\varphi^{2}, \quad (4.13)$$

but the corresponding modified Lagrangians differ by

$$\begin{aligned}\mathfrak{L}_1-\mathfrak{L}_2&\equiv\mathfrak{H}(\pi^{k}\varphi)_{;k}\\&\equiv\mathfrak{H}h_{k}{}^{\mu}[(\pi^{k}\varphi)_{,\mu}+A^{k}{}_{i\mu}\pi^{i}\varphi]\end{aligned} \quad (4.14)$$

which is not an explicit divergence. Thus in order to define the modified Lagrangian $\mathfrak{L}$ completely it would be necessary to specify which of the possible equivalent forms of the original Lagrangian is to be chosen. The reasons for this situation and the problem of choosing the correct form are discussed in the Appendix.

As in Sec. 2, one may define modified "currents" in terms of $L\equiv L\{\chi,\chi_{;k}\}$ by

$$\mathfrak{T}^{k}{}_{\mu}\equiv\partial\mathfrak{L}/\partial h_{k}{}^{\mu}\equiv\mathfrak{H}b^{i}{}_{\mu}\{(\partial L/\partial\chi_{;k})\chi_{;i}-\delta^{k}{}_{i}L\}, \quad (4.15)$$

$$\mathfrak{S}^{\mu}{}_{ij}\equiv-2(\partial\mathfrak{L}/\partial A^{ij}{}_{\mu})\equiv-\mathfrak{H}h_{k}{}^{\mu}(\partial L/\partial\chi_{;k})S_{ij}\chi, \quad (4.16)$$

where $b^{i}{}_{\mu}$ is the inverse of $h_{i}{}^{\mu}$, satisfying

$$b^{i}{}_{\mu}h_{i}{}^{\nu}=\delta_{\mu}{}^{\nu}, \quad b^{i}{}_{\mu}h_{j}{}^{\mu}=\delta^{i}{}_{j}.$$

To express the "conservation laws" which these currents satisfy in a simple form, it is convenient to extend the definition of the covariant derivative $\chi_{|\mu}$ (not $\chi_{;k}$). Originally, it is defined for χ and, therefore, by a trivial extension for any other quantity which is invariant under ξ^{μ} transformations, and transforms linearly under ϵ^{ij} transformations. We wish to extend it to any quantity which transforms linearly under ϵ^{ij} transformations, by simply ignoring the ξ^{μ} transformation properties altogether. Thus, for example, we would have

$$h_{i}{}^{\mu}{}_{|\nu}\equiv h_{i}{}^{\mu}{}_{,\nu}-A^{k}{}_{i\nu}h_{k}{}^{\mu}, \quad (4.17)$$

according to the ϵ^{ij} transformation law of $h_{i}{}^{\mu}$. We shall call this the ϵ covariant derivative. Later we shall define another covariant derivative which takes account of ξ^{μ} transformations also.

One can easily calculate the commutator of two ϵ covariant differentiations.[15] This gives

$$\chi_{|\mu\nu}-\chi_{|\nu\mu}=\tfrac{1}{2}R^{ij}{}_{\mu\nu}S_{ij}\chi, \quad (4.18)$$

where

$$R^{i}{}_{j\mu\nu}\equiv A^{i}{}_{j\mu,\nu}-A^{i}{}_{j\nu,\mu}-A^{i}{}_{k\mu}A^{k}{}_{j\nu}+A^{i}{}_{k\nu}A^{k}{}_{j\mu}.$$

This quantity is covariant under ϵ^{ij} transformations, and satisfies the cyclic identity

$$R^{i}{}_{j\mu\nu|\rho}+R^{i}{}_{j\nu\rho|\mu}+R^{i}{}_{j\rho\mu|\nu}\equiv 0.$$

[13] The reason for this may be seen in terms of the variation $\delta_0\chi$ given by (4.2). The analogs of the matrices T_a are clearly $-\partial_\mu$ and S_{ij}, so that the presence of the derivative $\chi_{,\mu}$ in the last term of (4.3) is to be expected. By analogy with (2.6) we should expect the covariant derivative to have the form

$$\chi_{;k}=\delta_{k}{}^{\mu}\chi_{,\mu}+\tfrac{1}{2}A^{ij}{}_{k}S_{ij}\chi-A^{\mu}{}_{k}\partial_{\mu}\chi.$$

Because of the appearance of derivatives, the first and last terms can be combined in the form $h_{k}{}^{\mu}\chi_{,\mu}$, where $h_{k}{}^{\mu}=\delta_{k}{}^{\mu}-A^{\mu}{}_{k}$. If we then set $A^{ij}{}_{k}=h_{k}{}^{\mu}A^{ij}{}_{\mu}$, we arrive at the same form for $\chi_{;k}$ as that obtained in the text.

[14] Multiplication of the entire Lagrangian by a constant factor is of course unimportant.

[15] Note that this could not be done without extending the definition, since one must know how to treat the index on $\chi_{1\mu}$. Here, as in Sec. 2, we simply ignore it.

It is thus closely analogous to $F^a{}_{\mu\nu}$. Note that $R^{ij}{}_{\mu\nu}$ is antisymmetric in both pairs of indices.

In terms of the ϵ covariant derivative, the "conservation laws" can be expressed in the form[5]

$$(\mathfrak{T}^k{}_\nu h_k{}^\mu)_{|\mu}+\mathfrak{T}^k{}_\mu h_k{}^\mu{}_{|\nu}=\tfrac{1}{2}\mathfrak{S}^\mu{}_{ij}R^{ij}{}_{\mu\nu}, \tag{4.19}$$

$$\mathfrak{S}^\mu{}_{ij|\mu}=\mathfrak{T}_{i\mu}h_j{}^\mu-\mathfrak{T}_{j\mu}h_i{}^\mu. \tag{4.20}$$

5. FREE GRAVITATIONAL LAGRANGIAN

We now wish to examine the quantity $\chi_{;k}$, rather than $\chi_{|\mu}$. As before, the covariant derivative of any quantity which transforms in a similar way to χ may be defined analogously. Now in particular $\chi_{;k}$ itself (unlike $\chi_{|\mu}$) is such a quantity, and therefore without extending the definition of covariant derivative one can evaluate the commutator $\chi_{;kl}-\chi_{;lk}$. However, this quantity is not simply obtained by multiplying $\chi_{|\mu\nu}-\chi_{|\nu\mu}$ by $h_k{}^\mu h_l{}^\nu$, as one might expect. The reason for this is that in evaluating $\chi_{;kl}$ one differentiates the $h_k{}^\mu$ in $\chi_{;k}$, and moreover adds an extra $A^i{}_{k\mu}$ term on account of the index k. Thus one finds

$$\chi_{;kl}-\chi_{;lk}=\tfrac{1}{2}R^{ij}{}_{kl}S_{ij}\chi-C^i{}_{kl}\chi_{;i}, \tag{5.1}$$

where

$$R^{ij}{}_{kl}\equiv h_k{}^\mu h_l{}^\nu R^{ij}{}_{\mu\nu}, \tag{5.2}$$

$$C^i{}_{kl}\equiv(h_k{}^\mu h_l{}^\nu-h_l{}^\mu h_k{}^\nu)b^i{}_{\mu|\nu}. \tag{5.3}$$

Note that (5.1) is not simply proportional to χ, but involves $\chi_{;i}$ also.[16]

We now look for a free Lagrangian $\mathfrak{L}_0$ for the new fields. Clearly $\mathfrak{L}_0$ must be an invariant density, and if we set

$$\mathfrak{L}_0\equiv\mathfrak{H}L_0,$$

then it is easy to see, as in the case of linear transformations, that the invariant L_0 must be a function only of the covariant quantities $R^{ij}{}_{kl}$ and $C^i{}_{kl}$. As before, there are many possible forms for $\mathfrak{L}_0$, but there is a difference between this case and the previous one in that all the indices on these expressions are of the same type (unlike $F^a{}_{\mu\nu}$), and one can, therefore, contract the upper indices with the lower. In fact, the condition that L_0 be a scalar in two separate spaces is now reduced to the condition that it be a scalar in one space. In particular, this means that there exists a linear invariant which has no analog in the previous case, namely,

$$R\equiv R^{ij}{}_{ij}.$$

There are in addition several quadratic invariants. However, if we again choose for L_0 the form of lowest possible degree, then we are led to the free Lagrangian[17]

$$\mathfrak{L}_0=\tfrac{1}{2}\mathfrak{H}R \tag{5.4}$$

which differs from (2.10) in being only linear in the derivatives.

[16] This is another example of the fact that for ξ^μ transformations derivatives play the role of the matrices T_a. Compare footnote 13.

[17] We choose units in which $\kappa=1$ (as well as $c=\hbar=1$).

With this choice of Lagrangian, the equations of motion for the new fields are

$$\mathfrak{H}(R^{ik}{}_{jk}-\tfrac{1}{2}\delta^i{}_jR)=-\mathfrak{T}^i{}_\mu h_j{}^\mu, \tag{5.5}$$

$$-[\mathfrak{H}(h_i{}^\mu h_j{}^\nu-h_j{}^\mu h_i{}^\nu)]_{|\nu}$$
$$\equiv\mathfrak{H}(h_k{}^\mu C^k{}_{ij}-h_j{}^\mu C^k{}_{ik}-h_i{}^\mu C^k{}_{kj})=\mathfrak{S}^\mu{}_{ij}. \tag{5.6}$$

From Eq. (5.6) one can immediately obtain a strict conservation law

$$(\mathfrak{S}^\mu{}_{ij}+\mathfrak{s}^\mu{}_{ij})_{,\mu}=0, \tag{5.7}$$

where

$$\mathfrak{s}^\mu{}_{ij}\equiv\mathfrak{H}A^k{}_{iv}(h_j{}^\mu h_k{}^\nu-h_k{}^\mu h_j{}^\nu)-\mathfrak{H}A^k{}_{jv}(h_i{}^\mu h_k{}^\nu-h_k{}^\mu h_i{}^\nu).$$

This quantity is expressible in the form

$$\mathfrak{s}^\mu{}_{ij}\equiv-2(\partial\mathfrak{L}_0/\partial A^{ij}{}_\mu)\equiv-\tfrac{1}{2}(\partial\mathfrak{L}_0/\partial A^{mn}{}_{\nu,\mu})f_{ij}{}^{mn}{}_{kl}A^{kl}{}_\nu,$$

which is closely analogous to (2.12), and should be compared with (4.16). Equation (5.7) is a rather surprising result, since $\mathfrak{S}^\mu{}_{ij}$ may very reasonably be interpreted as the spin density of the matter field,[18] so that it appears to be a law of conservation of spin with no reference to the orbital angular momentum. In fact, however, the orbital angular momentum appears in the corresponding "covariant conservation law" (4.20), and therefore part of the "spin" of the gravitational field, $\mathfrak{s}^\mu{}_{ij}$, may be regarded as arising from this source. Nevertheless, Eq. (5.7) differs from other statements of angular momentum conservation in that the coordinates do not appear explicitly.

It would also be possible to deduce from Eq. (5.5) a strict conservation law

$$[h_k{}^\nu(\mathfrak{T}^k{}_\mu+\mathfrak{t}^k{}_\mu)]_{,\nu}=0, \tag{5.8}$$

but there is a considerable amount of freedom in choosing $\mathfrak{t}^k{}_\mu$. The most natural definition, by analogy with (4.15) would be

$$\mathfrak{t}^k{}_\mu\equiv\partial\mathfrak{L}_0/\partial h_k{}^\mu,$$

and this quantity does indeed satisfy (5.8). However, in this case the expression within the parentheses itself vanishes, so that (5.8) is rather trivial. We shall not discuss the question of the correct choice of $\mathfrak{t}^k{}_\mu$ further, as this lies beyond the scope of the present paper.[19]

It should be noted that Eq. (5.6) can be solved, at least in principle, for $A^{ij}{}_\mu$. In the simple case when $\mathfrak{S}^\mu{}_{ij}$ vanishes, one finds[20]

$$A_{ij\mu}={}^0A_{ij\mu}\equiv\tfrac{1}{2}b^k{}_\mu(c_{kij}-c_{ijk}-c_{jki}),$$
$$c^k{}_{ij}\equiv(h_i{}^\mu h_j{}^\nu-h_j{}^\mu h_i{}^\nu)b^k{}_{\mu,\nu}. \tag{5.9}$$

[18] See H. J. Belinfante, Physica **6**, 887 (1939), and footnote 5.

[19] It is well known in the case of the ordinary metric theory of gravitation that many definitions of the energy pseudotensor are possible. See, for example, P. G. Bergmann, Phys. Rev. **112**, 287 (1958).

[20] The ${}^0A^{ij}{}_\mu$ are Ricci's coefficients of rotation. See for instance V. Fock, Z. Physik **57**, 261 (1929).

In general, if we write

$$\mathfrak{S}^{\mu}{}_{ij} \equiv \mathfrak{H} h_k{}^{\mu} S^k{}_{ij},$$

then

$$A_{ij\mu} = {}^0A_{ij\mu} - \tfrac{1}{2} b^k{}_{\mu}(S_{kij} - S_{ijk} - S_{jki} - \eta_{ki} S^l{}_{lj} - \eta_{kj} S^l{}_{il}). \quad (5.10)$$

If the original Lagrangian L is of first order in the derivatives, then $S^k{}_{ij}$ is independent of $A^{ij}{}_{\mu}$ so that (5.10) is an explicit solution. Otherwise, however, $A^{ij}{}_{\mu}$ also appears on the right-hand side of this equation.

We conclude this section with a discussion of the Lagrangian for the fields $A^a{}_{\mu}$ introduced in Sec. 2 when the "gravitational" fields $h_k{}^{\mu}$ and $A^{ij}{}_{\mu}$ are also introduced. The fields $A^a{}_{\mu}$ should not be regarded merely as components of χ when dealing with Lorentz transformations, since one must preserve the invariance under the linear transformations. To find the correct form of the Lagrangian, one should consider simultaneously Lorentz transformations and these linear transformations. This can be done provided that the matrices T^a commute with the S_{ij}, a condition which is always fulfilled in practice. Then one finds that χ_k in L should be replaced by a derivative which is covariant under both (2.1) and (4.1), namely,

$$\chi_{;k} = h_k{}^{\mu}(\chi_{,\mu} + \tfrac{1}{2} A^{ij}{}_{\mu} S_{ij} \chi + A^a{}_{\mu} T_a \chi).$$

The commutator $\chi_{;kl} - \chi_{;lk}$ then contains the extra term

$$F^a{}_{kl} T^a \chi,$$

where

$$F^a{}_{kl} \equiv h_k{}^{\mu} h_l{}^{\nu} F^a{}_{\mu\nu},$$

with $F^a{}_{\mu\nu}$ given by (2.9). It is important to notice that the derivatives of $A^a{}_{\mu}$ in $F^a{}_{\mu\nu}$ are ordinary derivatives, not covariant ones. (We shall see in the next section that the ordinary and covariant curls are not equal, because the affine connection is in general nonsymmetric.) As before, one can see that any invariant function of $A^a{}_{\mu}$ must be a function of $F^a{}_{kl}$ only, and the simplest free Lagrangian for $A^a{}_{\mu}$ is, therefore,

$$-\tfrac{1}{4} \mathfrak{H} F^a{}_{kl} F_a{}^{kl}. \quad (5.11)$$

6. GEOMETRICAL INTERPRETATION

Up to this point, we have not given any geometrical significance to the transformations (4.1), or to the new fields $h_k{}^{\mu}$ and $A^{ij}{}_{\mu}$, but it is useful to do so in order to be able to compare the theory with the more familiar metric theory of gravitation.

Now the ξ^{μ} transformations are general coordinate transformations, and according to (4.11) $h_k{}^{\mu}$ transforms like a contravariant vector under these transformations, while $b^k{}_{\mu}$ and $A^{ij}{}_{\mu}$ transform like covariant vectors. Thus the quantity

$$g_{\mu\nu} \equiv b^k{}_{\mu} b_{k\nu} \quad (6.1)$$

is a symmetric covariant tensor, and may therefore be interpreted as the metric tensor of a Riemannian space. It is moreover invariant under the ϵ^{ij} transformations. Evidently, the Greek indices may be regarded as world tensor indices, and we must of course abandon for them the convention that all indices are to be raised or lowered with the flat-space metric $\eta_{\mu\nu}$, and use $g_{\mu\nu}$ instead. It is easy to see that the scalar density $\mathfrak{H}$ is equal to $(-g)^{\frac{1}{2}}$, where $g = \det(g_{\mu\nu})$.

Now, in view of the relation (6.1), $h_k{}^{\mu}$ and $b^k{}_{\mu}$ are the contravariant and covariant components, respectively, of a vierbein system in the Riemannian space.[21] Thus the ϵ^{ij} transformations should be interpreted as vierbein rotations, and the Latin indices as local tensor indices with respect to this vierbein system. The original field χ may be decomposed into local tensors and spinors,[22] and from the tensors one can form corresponding world tensors by multiplying by $h_k{}^{\mu}$ or $b^k{}_{\mu}$. For example, from a local vector v^i one can form

$$v^{\mu} = h_i{}^{\mu} v^i, \quad v_{\mu} = b^i{}_{\mu} v_i. \quad (6.2)$$

No confusion can be caused by using the same symbol v for the local and world vectors, since they are distinguished by the type of index, and indeed we have already used this convention in (5.2). Note that $v_{\mu} = g_{\mu\nu} v^{\nu}$, so that (6.2) is consistent with the choice of metric (6.1). We shall frequently use this convention of associating world tensors with given local tensors without explicit mention on each occasion.

The field $A^i{}_{j\mu}$ may reasonably be called a "local affine connection" with respect to the vierbein system, since it specifies the covariant derivatives of local tensors or spinors.[23] For a local vector, this takes the form

$$v^i{}_{|\nu} = v^i{}_{,\nu} + A^i{}_{j\nu} v^j, \quad v_{j|\nu} = v_{j,\nu} - A^i{}_{j\nu} v_i. \quad (6.3)$$

It may be noticed that the relation (4.10) between $\chi_{|\mu}$ and $\chi_{;k}$ is of the same type as (6.2) and could be written simply as

$$\chi_{;\mu} = \chi_{|\mu} \quad (6.4)$$

according to our convention. However, we shall retain the use of two separate symbols because we wish to extend the definition of covariant derivative in a different way to that of Sec. 4. It seems natural to define the covariant derivative of a world tensor in terms of the covariant derivative of the associated local tensor. Thus, for instance, to define the covariant derivatives of the world vectors (6.2) one would form the world tensors corresponding to (6.3). This gives

$$v^{\lambda}{}_{;\nu} \equiv h_i{}^{\lambda} v^i{}_{|\nu} = v^{\lambda}{}_{,\nu} + \Gamma^{\lambda}{}_{\mu\nu} v^{\mu},$$

$$v_{\mu;\nu} \equiv b^i{}_{\mu} v_{i|\nu} = v_{\mu,\nu} - \Gamma^{\lambda}{}_{\mu\nu} v^{\lambda},$$

where

$$\Gamma^{\lambda}{}_{\mu\nu} \equiv h_i{}^{\lambda} b^i{}_{\mu|\nu} \equiv -b^i{}_{\mu} h_i{}^{\lambda}{}_{|\nu}. \quad (6.5)$$

Note that this definition of $\Gamma^{\lambda}{}_{\mu\nu}$ is equivalent to the

[21] See for instance H. Weyl, Z. Physik **56**, 330 (1929).
[22] H. J. Belinfante, Physica **7**, 305 (1940).
[23] Compare J. A. Schouten, J. Math. and Phys. **10**, 239 (1931).

requirement that the covariant derivatives of the vierbein components should vanish,

$$h_i{}^\lambda{}_{;\nu}\equiv 0, \quad b^i{}_{\mu;\nu}\equiv 0. \tag{6.6}$$

For a generic quantity α transforming according to

$$\delta\alpha=\tfrac{1}{2}\epsilon^{ij}S_{ij}\alpha+\xi^\lambda{}_{,\mu}\Sigma_\lambda{}^\mu\alpha, \tag{6.7}$$

the covariant derivative is defined by[21]

$$\alpha_{;\nu}\equiv\alpha_{,\nu}+\tfrac{1}{2}A^{ij}{}_\nu S_{ij}\alpha+\Gamma^\lambda{}_{\mu\nu}\Sigma_\lambda{}^\mu\alpha, \tag{6.8}$$

whereas the ϵ covariant derivative defined in Sec. 4 is obtained by simply omitting the last term of (6.8). Note that the two derivatives are equal for purely local tensors or spinors, but not otherwise. One easily finds that the commutator of two covariant differentiations is given by

$$\alpha_{;\mu\nu}-\alpha_{;\nu\mu}=\tfrac{1}{2}R^{ij}{}_{\mu\nu}S_{ij}\alpha+R^\rho{}_{\sigma\mu\nu}\Sigma_\rho{}^\sigma\alpha-C^\lambda{}_{\mu\nu}\alpha_{;\lambda},$$

where $R^\rho{}_{\sigma\mu\nu}$ and $C^\lambda{}_{\mu\nu}$ are defined in the usual way in terms of $R^i{}_{j\mu\nu}$ and $C^i{}_{kl}$. They are both world tensors, and can easily be expressed in terms of $\Gamma^\lambda{}_{\mu\nu}$, in the form[24]

$$R^\rho{}_{\sigma\mu\nu}=\Gamma^\rho{}_{\sigma\mu,\nu}-\Gamma^\rho{}_{\sigma\nu,\mu}-\Gamma^\rho{}_{\lambda\mu}\Gamma^\lambda{}_{\sigma\nu}+\Gamma^\rho{}_{\lambda\nu}\Gamma^\lambda{}_{\sigma\nu}, \tag{6.9}$$

$$C^\lambda{}_{\mu\nu}=\Gamma^\lambda{}_{\mu\nu}-\Gamma^\lambda{}_{\nu\mu}. \tag{6.10}$$

Thus one sees that $R^\rho{}_{\sigma\mu\nu}$ is just the Riemann tensor formed from the affine connection $\Gamma^\lambda{}_{\mu\nu}$.

From (6.6) it follows that

$$g_{\mu\nu;\rho}\equiv 0, \tag{6.11}$$

so that it is consistent to interpret $\Gamma^\lambda{}_{\mu\nu}$ as an affine connection in the Riemannian space. However, the definition (6.5) evidently does not guarantee that it is symmetric, so that in general it is not the Christoffel connection. The curvature scalar R has the usual form

$$R\equiv R^\mu{}_\mu, \quad R_{\mu\nu}\equiv R^\lambda{}_{\mu\lambda\nu},$$

so that the free gravitational Lagrangian is just the usual one except for the nonsymmetry of $\Gamma^\lambda{}_{\mu\nu}$. It should be remarked that it would be incorrect to treat the 64 components of $\Gamma^\lambda{}_{\mu\nu}$ as independent variables, since there are only 24 components of $A^{ij}{}_\mu$. In fact the $\Gamma^\lambda{}_{\mu\nu}$ are restricted by the 40 identities (6.11). Thus there is no contradiction with the well-known fact that the first-order Palatini Lagrangian with nonsymmetric $\Gamma^\lambda{}_{\mu\nu}$ does not yield (6.11) as equations of motion.[25]

The equations of motion (5.5) and (5.6) can be rewritten in the form

$$\mathfrak{H}(R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R)=-\mathfrak{T}_{\mu\nu}, \tag{6.12}$$

$$\mathfrak{H}C^\lambda{}_{\mu\nu}=\mathfrak{S}^\lambda{}_{\mu\nu}-\tfrac{1}{2}\delta^\lambda{}_\mu\mathfrak{S}^\rho{}_{\rho\nu}-\tfrac{1}{2}\delta^\lambda{}_\nu\mathfrak{S}^\rho{}_{\mu\rho}. \tag{6.13}$$

From Eqs. (6.10) and (6.13) one sees that in the absence of matter the affine connection $\Gamma^\lambda{}_{\mu\nu}$ is symmetric, and therefore equal to the Christoffel connection ${}^0\Gamma^\lambda{}_{\mu\nu}$. (This is the analog for world tensors of ${}^0A^i{}_{j\mu}$.) Then $R_{\mu\nu}$ is symmetric, and Eq. (6.12) yields Einstein's familiar equations for empty space,

$$R_{\mu\nu}=0.$$

However, when matter is present, $\Gamma^\lambda{}_{\mu\nu}$ is no longer symmetric, and its antisymmetric part is given by (6.13). Then the tensor $R_{\mu\nu}$ is also nonsymmetric, and correspondingly the energy tensor density $\mathfrak{T}_{\mu\nu}$ is in general nonsymmetric, because $h_k{}^\mu$ does not appear in $\mathfrak{L}$ only through the symmetric combination $g^{\mu\nu}$. Thus the theory differs slightly from the usual one, in a way first noted by Weyl.[4] In the following section, we shall investigate this difference in more detail.[5]

Finally, we can rewrite the covariant conservation laws in terms of world tensors. It is convenient to define the contraction

$$C_\mu\equiv C^\lambda{}_{\mu\lambda},$$

since the covariant divergence of a vector density $\mathfrak{f}^\mu$ is then

$$\mathfrak{f}^\mu{}_{;\mu}=\mathfrak{f}^\mu{}_{,\mu}+C_\mu\mathfrak{f}^\mu. \tag{6.14}$$

The conservation laws become

$$\mathfrak{T}^\nu{}_{\mu;\nu}-C_\nu\mathfrak{T}^\nu{}_\mu+C^\lambda{}_{\mu\nu}\mathfrak{T}^\nu{}_\lambda=\tfrac{1}{2}R^{\rho\sigma}{}_{\mu\nu}\mathfrak{S}^\nu{}_{\rho\sigma},$$

$$\mathfrak{S}^\mu{}_{\rho\sigma;\mu}-C_\mu\mathfrak{S}^\mu{}_{\rho\sigma}=\mathfrak{T}_{\rho\sigma}-\mathfrak{T}_{\sigma\rho}.$$

It may be noticed that these are slightly more complicated than the expressions in terms of the ϵ covariant derivative.

7. COMPARISON WITH METRIC THEORY

For simplicity, we shall assume in this section that L is only of first order in the derivatives, so that (5.10) is an explicit solution for $A^{ij}{}_\mu$. The difference between the theory presented here and the usual one arises because we are using a Lagrangian $\mathfrak{L}_0$ of first order, in which $h_k{}^\mu$ and $A^{ij}{}_\mu$ are independent variables. The situation is entirely analogous to that which obtains for any theory with "derivative" interaction. In first-order form, the "momenta" $A^{ij}{}_\mu$ are not just equal to derivatives of the "coordinates" $h_k{}^\mu$, or in other words to ${}^0A^{ij}{}_\mu$. Thus an interaction which appears simple in first-order form will be more complicated if a second-order Lagrangian is used, and vice versa.

The second-order form of the Lagrangian may be obtained by substituting for $A^{ij}{}_\mu$ the expression (5.10). This gives

$$\mathfrak{L}'={}^0\mathfrak{L}+{}^0\mathfrak{L}_0+{}^1\mathfrak{L},$$

where ${}^0\mathfrak{L}$ and ${}^0\mathfrak{L}_0$ are obtained from $\mathfrak{L}$ and $\mathfrak{L}_0$ by replacing $A^{ij}{}_\mu$ by ${}^0A^{ij}{}_\mu$ (or equivalently $\Gamma^\lambda{}_{\mu\nu}$ by ${}^0\Gamma^\lambda{}_{\mu\nu}$), and ${}^1\mathfrak{L}$ is an additional term quadratic in $S^k{}_{ij}$, namely,

$${}^1\mathfrak{L}=\tfrac{1}{8}\mathfrak{H}(2S_{ijk}S^{jki}-S_{ijk}S^{ijk}+2S^i{}_{ik}S_j{}^{jk}). \tag{7.1}$$

In this Lagrangian, only $h_k{}^\mu$ and χ are treated as inde-

[24] This is a generalization to nonsymmetric affinities of the result proved in the appendix to footnote 3. See also footnotes 4 and 5.

[25] See for instance E. Schrödinger, *Space-time Structure* (Cambridge University Press, New York, 1950).

pendent variables. The equations of motion are equivalent to those previously obtained if the variables $A^{ij}{}_{\mu}$ are eliminated from the latter by using (5.10).

The usual metric theory, on the other hand, is given by the Lagrangian

$$\mathfrak{L}''={}^{0}\mathfrak{L}+{}^{0}\mathfrak{L}_{0},$$

without the extra terms (7.1). If this Lagrangian were written in a first-order form by introducing additional independent variables $A^{ij}{}_{\mu}$, then one would arrive at a form identical to the one given here except for the appearance of extra terms equal to (7.1) with a negative sign.

Thus we see that the only difference between the two theories is the presence or absence of these "direct-interaction" terms. Now if we had not set $\kappa=1$, then $\mathfrak{L}_0$ would have a factor κ^{-1}, whereas the terms (7.1) would appear with the factor κ. They are, therefore, extremely small in comparison to other interaction terms. In particular, for a Dirac field, they would be proportional to (see Appendix)

$$\kappa\bar{\psi}\gamma_k\gamma_5\psi\bar{\psi}\gamma^k\gamma_5\psi.$$

Thus they are similar in form to the Fermi interaction terms, but much smaller in magnitude, so that it seems impossible that they would lead to any observable difference between the predictions of the two theories. Hence we must conclude that for all practical purposes the theory presented here is equivalent to the usual one.

ACKNOWLEDGMENTS

The author is indebted to Drs. J. L. Anderson, P. W. Higgs, and D. W. Sciama for helpful discussions and comments.

APPENDIX

In this appendix we shall discuss the remaining ambiguity in the modified Lagrangian. It was pointed out in Sec. 4 that the generally covariant Lagrangians obtained from two equivalent Lagrangians L_1 and L_2 are in general inequivalent. One can now see that in fact they differ by a covariant divergence. Thus (4.14) can be written in the form

$$\mathfrak{L}_1-\mathfrak{L}_2=(\mathfrak{H}h_k{}^{\mu}\pi^k\varphi)_{;\mu},$$

but in view of (6.14) this is not equal to the ordinary divergence. It is clear that quite generally changing L by a divergence must change $\mathfrak{L}$ by the covariant divergence of a quantity which is a vector density under coordinate transformations, and invariant under all other transformations. This is the reason for the difference between this case and that of the linear transformations of Sec. 2.

We now wish to investigate the possibility of choosing a criterion which will select a particular form of L, and thus specify $\mathfrak{L}$ completely. There does not seem to be any really compelling reason for one choice rather than another, but there are plausible arguments for a particular choice.

The most obvious criterion would be to require that the Lagrangian should be written in the symmetrized first-order form suggested by Schwinger,[26] which in the case of the scalar field discussed in Sec. 4 is

$$L=\tfrac{1}{2}(L_1+L_2).$$

This corresponds to treating φ and π^k on a symmetrical footing. However, this may not in fact be the correct choice, because for some purposes φ and π^k should not be treated in this way. In fact, the two Lagrangians differ in one important respect: $\mathfrak{L}_1$ is independent of $A^{ij}{}_{\mu}$, whereas $\mathfrak{L}_2$ is not. Correspondingly, for L_1 the quantity $S^k{}_{ij}$ vanishes, whereas for L_2 one finds

$$S^k{}_{ij}=(\delta^k{}_i\pi_j-\delta^k{}_j\pi_i)\varphi.$$

The conservation laws in the two cases are of course the same, because the quantities $T^k{}_i$ also differ. Now the tensor $S^k{}_{ij}$ has often been interpreted as the spin density,[18] so that the two cases differ with regard to the separation of the total angular momentum into orbital and spin terms. The scalar field is normally regarded as a field of spinless particles, so that one would naturally expect $S^k{}_{ij}$ to vanish. This, therefore, furnishes a possible criterion, which would select L_1 rather than L_2. With this choice, a preferred position is assigned to the "wave function" φ rather than the "momenta" π^k, and the derivatives are written on φ only. In this way one achieves a vanishing spin tensor, because the matrices S_{ij} are zero for the scalar field φ, but not for the vector π^k. It may be noticed that L_1 is automatically selected if one writes the Lagrangian in its second-order form in terms of φ only:

$$L_1'=\tfrac{1}{2}\varphi_{,k}\varphi^{,k}-\tfrac{1}{2}m^2\varphi^2,$$

which yields the modified Lagrangian

$$\mathfrak{L}_1'=\tfrac{1}{2}\mathfrak{H}(g^{\mu\nu}\varphi_{,\mu}\varphi_{,\nu}-m^2\varphi^2),$$

equivalent to $\mathfrak{L}_1$.[27] This should be contrasted with the second-order form of $\mathfrak{L}_2$, which is

$$\mathfrak{L}_2'=\tfrac{1}{2}\mathfrak{H}^{-1}(\mathfrak{H}h_i{}^{\mu}\varphi)_{;\mu}(\mathfrak{H}h^{i\nu}\varphi)_{;\nu}-\tfrac{1}{2}\mathfrak{H}m^2\varphi^2,$$

and clearly differs from $\mathfrak{L}_1'$ by a covariant divergence.

This seems to be a resonable criterion, but the arguments for it cannot be regarded as conclusive. For, although it is true that the spin tensor obtained from L_2 is nonzero, it is still true that the three space-space components of the total spin

$$s_{ij}=\int d_3x\, S^0{}_{ij}$$

are zero. Thus L_1 and L_2 differ only in the values of the

[26] J. Schwinger, Phys. Rev. 91, 713 (1953).

[27] Here $\mathfrak{L}_1$ is a "linearization" of $\mathfrak{L}_1'$ in the sense of T. W. B. Kibble and J. C. Polkinghorne, Nuovo cimento 8, 74 (1958).

spin part of the $(0i)$ components of angular momentum. Indeed, one easily sees that it is true in general that adding a divergence to L will change only the $(0i)$ components of $\mathfrak{s}_{ij}$. Since it is not at all clear what significance should be attached to the separation of these components into "orbital" and "spin" terms, it might be questioned whether one should expect the spin terms to vanish even for a spinless particle. Even so, the choice of L_1 seems in this case to be the most reasonable.

For a field of spin 1, the corresponding choice would be

$$L_1=-\tfrac{1}{2}f^{ij}(a_{i,j}-a_{j,i})+\tfrac{1}{4}f^{ij}f_{ij}+\tfrac{1}{2}m^2a_ia^i,$$

which is again equivalent to the choice of the second-order Lagrangian in terms of a_i only. It yields

$$S^k{}_{ij}=a_if_j{}^k-a_jf_i{}^k,$$

which is a reasonable definition of the spin density.[28] The modified Lagrangian may be expressed in terms of the world vector a_μ as

$$\mathfrak{L}=-\tfrac{1}{4}\mathfrak{H}g^{\mu\rho}g^{\nu\sigma}(a_{\mu;\nu}-a_{\nu;\mu})(a_{\rho;\sigma}-a_{\sigma;\rho})+\tfrac{1}{2}\mathfrak{H}m^2g^{\mu\nu}a_\mu a_\nu. \quad \text{(A.1)}$$

It should be noticed that the electromagnetic Lagrangian is not obtained simply by putting $m=0$ in (A.1). The difference is that the derivatives in (A.1) are covariant derivatives, and since $\Gamma^\lambda{}_{\mu\nu}$ is nonsymmetric the covariant curl is not equal to the ordinary curl (though both are of course tensors). In fact, (A.1) with $m=0$ would not be gauge invariant. The reason for the difference is that a_i is here treated simply as a component of χ, whereas A_μ is introduced along with the gravitational variables to ensure gauge invariance.[29]

For a spinor field ψ, symmetry between ψ and $\bar{\psi}$ appears to demand that one should choose the symmetrized Lagrangian

$$L=\tfrac{1}{2}(\bar{\psi}i\gamma^k\psi_{,k}-\bar{\psi}_{,k}i\gamma^k\psi)-m\bar{\psi}\psi,$$

which yields the spin density

$$S_{kij}=\tfrac{1}{2}\epsilon_{kijl}\bar{\psi}i\gamma^l\gamma_5\psi.$$

Since the Lagrangian $\mathfrak{L}$ must be Hermitian, one could not write the derivative on ψ alone. There remains, however, another possible choice: We could introduce a distinction between the left- and right-handed components, $\psi_\pm=\frac{1}{2}(1\pm i\gamma_5)\psi$, treating one of them line φ and the other like π^k. This gives the Lagrangian

$$L=\tfrac{1}{2}\bar{\psi}i\gamma^k(1+i\gamma_5)\psi_{,k}-\tfrac{1}{2}\bar{\psi}_{,k}i\gamma^k(1-i\gamma_5)\psi-m\bar{\psi}\psi.$$

This form of Lagrangian may seem rather unnatural, but it should be mentioned because there are other grounds for treating ψ_+ and ψ_- on a nonsymmetrical footing.[30]

[28] Compare footnote 18.

[29] This has the rather strange consequence that for the electromagnetic field the "spin" tensor $S^k{}_{ij}$ vanishes, since the Lagrangian is independent of $A^{ij}{}_\mu$.

[30] See R. P. Feynman and M. Gell-Mann, Phys. Rev. **109**, 193 (1958).

Solution of the Equations of Statistical Mechanics*

ROBERT M. LEWIS
Institute of Mathematical Sciences, New York University, New York, New York
(Received April 25, 1960)

The solution of the initial value problem for Bogoliubov's functional differential equation of nonequilibrium statistical mechanics is obtained. This solution is then expanded in an infinite power series in the density which has the advantage that the calculation of the leading terms requires the solution of s-body problems only for small values of s. A derivation of the equilibrium equations by reduction from the nonequilibrium equation is included. These results are applied to obtain a simple derivation of the Boltzmann equation.

1. INTRODUCTION

THE statistical mechanical treatment of a classical many-body system usually begins with an "n-particle function" D_n which is the solution of an initial value problem for Liouville's equation. There are, however, two major difficulties with this approach:

1. In the problems of interest the solution of Liouville's equation is equivalent to the solution of an n-body problem where n is very large, and is therefore not practical.
2. The initial conditions are, in general, unknown.

In an attempt to circumvent these difficulties, one introduces "s-particle density functions" F_s defined by appropriate integrals of D_n. Bogoliubov has shown[1] that for these functions, the Liouville equation can be replaced by a functional differential equation for a generating functional $L[u]$ which generates the functions F_s, and has obtained an expansion of the solution of the equation to first order in the density.

In Sec. 2 of this paper we derive the functional differential equation by a slight variation of Bogoliubov's method. The resulting Eq. (20) differs slightly from, but is equivalent to, the equation of Bogoliubov; however, the form of Eq. (20) facilitates a new method of solution.

Section 3 contains the main result of this paper. In that section we obtain the solution of the initial value problem for Eq. (20) by a method similar to the method devised by B. Zumino[2] for the equilibrium case. The solution is then expanded in an infinite power series in the density. In this form, it has the advantage that for small densities it may be approximated by a few terms of the expansion. Then to obtain an explicit expression for F_s where s is small, only certain k-body problems, where k is small, need to be solved. Furthermore, only the initial data for certain functions F_j, where j is small, are required. If these data are known, our expansion circumvents both of the difficulties enumerated previously.

Sections 4 and 5 are included for the sake of completeness. In Sec. 4 we carry out a suggestion of Zumino and derive the functional differential equation for the equilibrium case by reduction from the nonequilibrium equation. In Sec. 5 we solve the equilibrium equation by a slight simplification of Zumino's method. Section 6 is an application of the expansion obtained in Sec. 3. That expansion is used to obtain a very simple derivation of the Boltzmann equation.

2. DERIVATION OF THE FUNCTIONAL DIFFERENTIAL EQUATION

We consider a classical mechanical system of n identical monatomic particles contained in a finite volume, V. The dynamical state of the jth particle is described by the 6 component vector

$$x_j=(q_j,p_j)=(q_j^1,q_j^2,q_j^3,p_j^1,p_j^2,p_j^3),$$

where the q_j^α are the Cartesian coordinates of the particle, and the p_j^α are the conjugate momenta. x_j is a point in the phase-space Ω_V defined by the restriction that q_j is a point in the finite volume V. The Hamiltonian of the system is given by

$$\mathcal{H}_n=\sum_{i=1}^{n} h(x_i)+U_n, \tag{1}$$

$$h(x_i)=T(p_i)+u_V(q_i), \tag{2}$$

$$U_n=\sum_{1\le i<j\le n}\phi(|q_i-q_j|), \tag{3}$$

$$T(p_i)=\frac{p_i^2}{2m}=\sum_{\alpha=1}^{3}\frac{(p_i^\alpha)^2}{2m}, \tag{4}$$

where m denotes the mass of a particle, ϕ is the interparticle potential, and $u_V(q_i)$ is the potential due to the containing boundary. Thus $u_V(q)$ is constant inside V and rapidly approaches infinity at the boundary.

The statistical-mechanical behavior of the system is described by the n-particle "probability density" function, $D_n(t,x_1,\cdots,x_n)$ which is symmetric in the

* The research in this paper was supported by the U. S. Air Force under a contract monitored by the AF Office of Scientific Research of the Air Research and Development Command.

[1] (a) N. N. Bogoliubov, *Problems of a Dynamical Theory in Statistical Physics* (translated from Russian by E. K. Gora), Geophysics Research Dictorate, ASTIA Document No. AD-213317. Copies may be obtained by writing to the translator. (b) Much of the material in this reference appears in J. Phys. (U.S.S.R.) **10**, 257, 265 (1946).

[2] B. Zumino, Phys. Fluids **2**, 20 (1959); also see New York University Institute of Mathematical Sciences, Division of EM Research, Rept. No. HT-1.

variables $(x_1,\cdots,x_n)$, is normalized by the condition,

$$\int_{\Omega V} D_n dx_1\cdots dx_n=1, \tag{5}$$

and is a solution of Liouville's equation,

$$\frac{\partial D_n}{\partial t}=[\mathfrak{H}_n; D_n]=\sum_{i=1}^{n}\sum_{\alpha=1}^{3}\left\{\frac{\partial \mathfrak{H}_n}{\partial q_i^{\alpha}}\frac{\partial D_n}{\partial p_i^{\alpha}}-\frac{\partial \mathfrak{H}_n}{\partial p_i^{\alpha}}\frac{\partial D_n}{\partial q_i^{\alpha}}\right\}. \tag{6}$$

Let $S_t^{(n)}$ denote the solution operator of the n-particle mechanical system, i.e., if the system at time $t=0$ is represented by the state $\{x_1,\cdots,x_n\}$, at time t it will be represented by the state $\{x_1',\cdots,x_n'\}=S_t^{(n)}\{x_1,\cdots,x_n\}$. Under suitable conditions, the solution operator exists, but of course cannot be calculated explicitly except when n is very small. If g is a function of $(\tau,x_1,\cdots,x_{n+k})$ it is convenient to define $S_t^{(n)}g$ by the equation

$$S_t^{(n)}g(\tau,x_1,\cdots,x_{n+k})=g(\tau,S_t^{(n)}\{x_1,\cdots,x_n\}, x_{n+1},\cdots,x_{n+k}).^3 \tag{7}$$

In terms of the solution operator, one may express the solution of the initial value problem for Liouville's equation in the form

$$D_n(t,x_1,\cdots,x_n)=S_{-t}^{(n)}D_n(0,x_1,\cdots,x_n). \tag{8}$$

However, since $S_{-t}^{(n)}$ cannot be calculated, and since $D_n(0,x_1,\cdots,x_n)$ is in general unknown, the solution [Eq. (8)] is of no practical value.

We introduce the s-particle density functions

$$F_{n,s}(t,x_1,\cdots,x_s)=V^s\int_{\Omega V} D_n(t,x_1,\cdots,x_n)dx_{s+1}\cdots dx_n; \quad s=0,1,2,\cdots. \tag{9}$$

It follows that $F_{n,s}$ is symmetric in $(x_1,\cdots,x_s)$, $F_{n,0}=1$, and

$$\int_{\Omega V}\frac{1}{V^s}F_{n,s}dx_1\cdots dx_s=\int_{\Omega V} D_n dx_1\cdots dx_n=1; \quad s=1,2,\cdots. \tag{10}$$

We now set $v=V/n$ and introduce the functional

$$L_n[t,u]=\int_{\Omega V} D_n(t,x_1,\cdots,x_n)\times\prod_{i=1}^{n}[1+vu(x_i)]dx_1\cdots dx_n, \tag{11}$$

which is defined on the domain of functions $u(x)$ for which the integral converges. By functional differentiation[4] we obtain

$$\frac{\delta^s L_n}{\delta u(x_1)\cdots\delta u(x_s)}=\frac{v^s n!}{(n-s)!}\int_{\Omega V} D_n(t,x_1,\cdots,x_n)\times\prod_{i=s+1}^{n}[1+vu(x_i)]dx_{s+1}\cdots dx_n; \quad s=0,1,2,\cdots,n; \tag{12}$$

$$\left.\frac{\delta^s L_n}{\delta u(x_1)\cdots\delta u(x_s)}\right|_{u=0}=\frac{n!}{n^s(n-s)!}F_{n,s}(t,x_1,\cdots,x_s); \quad s=0,1,2,\cdots,n. \tag{13}$$

With the aid of Eq. (13), L_n may now be expressed as a (finite) series expansion around $u=0$:

$$L_n[t,u]=1+\sum_{s=1}^{n}\frac{1}{s!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{s-1}{n}\right)\times\int_{\Omega V}F_{n,s}(t,x_1,\cdots,x_s)u(x_1)\cdots u(x_s)dx_1\cdots dx_s. \tag{14}$$

A differential equation for L_n may be obtained by multiplying Eq. (6) by

$$\prod_{i=1}^{n}[1+vu(x_i)],$$

and integrating with respect to $x_1,\cdots,x_n$ over Ω_V. We obtain

$$\frac{\partial L_n}{\partial t}=\sum_{k=1}^{n}\int_{\Omega V}[1+vu(x_k)]\left\{h(x_k); D_n\prod_{\substack{i=1\\ i\neq k}}^{n}[1+vu(x_i)]\right\}dx_1\cdots dx_n+\sum_{1\le r<s\le n}\int_{\Omega V}[1+vu(x_r)][1+vu(x_s)]\left\{\phi(|q_r-q_s|); D_n\prod_{\substack{i=1\\ si\neq r,}}^{n}[1+vu(x_i)]\right\}dx_i\cdots dx_n. \tag{15}$$

By making use of the symmetry of D_n Eq. (15) becomes

$$\frac{\partial L_n}{\partial t}=n\int_{\Omega V}[1+vu(x_1)]\left\{h(x_1); \int_{\Omega V} D_n\prod_{i=2}^{n}[1+vu(x_i)]dx_2\cdots dx_n\right\}dx_1+\frac{n(n-1)}{2}\int_{\Omega V}[1+vu(x_1)][1+vu(x_2)]\left\{\phi(|q_1-q_2|); \int_{\Omega V} D_n\prod_{i=3}^{n}[1+vu(x_i)]dx_3\cdots dx_n\right\}dx_1dx_2 \tag{16}$$

$$=\frac{1}{v}\int_{\Omega V}[1+vu(x_1)]\left\{h(x_1); \frac{\delta L_n}{\delta u(x_1)}\right\}dx_1+\frac{1}{2v^2}\int_{\Omega V}[1+vu(x_1)][1+vu(x_2)]\left\{\phi(|q_1-q_2|); \frac{\delta^2 L_n}{\delta u(x_1)\delta u(x_2)}\right\}dx_1dx_2.$$

[3] Thus $S_t^{(n)}$ acts on the first n of the variables x_j appearing in g.

[4] V. Volterra, *Theory of Functionals* (Blackie and Son, Limited, London, England, 1931).

We now let $n \to \infty$ and $V \to \infty$ in Eqs. (14) and (16) in such a way that $v = V/n$ is finite. If we set

$$L[t,u] = \lim_{\substack{n\to\infty \\ V\to\infty}} L_n[t,u], \tag{17}$$

then from Eq. (14)

$$L[t,u] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \int F_s(t,x_1,\cdots,x_s) \times u(x_1)\cdots u(x_s)dx_1\cdots dx_s. \tag{18}$$

Here

$$F_s(t,x_1,\cdots,x_s) = \lim_{\substack{n\to\infty \\ V\to\infty}} F_{n,s}(t,x_1,\cdots,x_s). \tag{19}$$

From Eq. (16)

$$\frac{\partial L}{\partial t} = \int \left[u(x_1) + \frac{1}{v}\right]\left[T(p_1); \frac{\delta L}{\delta u(x_1)}\right]dx_1 + \tfrac{1}{2}\int \left[u(x_1)+\frac{1}{v}\right]\left[u(x_2)+\frac{1}{v}\right] \times \left[\phi(|q_1-q_2|); \frac{\delta^2 L}{\delta u(x_1)\delta u(x_2)}\right]dx_1dx_2. \tag{20}$$

It follows from Eq. (18) that

$$F_s(t,x_1,\cdots,x_s) = \left.\frac{\delta^s L}{\delta u(x_1)\cdots\delta u(x_s)}\right|_{u=0}. \tag{21}$$

Equation (20) is the functional differential equation which we shall solve in the next section. The solution L is called the "generating functional" because it generates the functions F_s by means of Eq. (21). We have derived Eq. (20) by the method of Bogoliubov[1] with a slight modification, and our equation apparently differs slightly from the corresponding Eq. (7.9) of Bogoliubov. However, the difference is only apparent. The two equations can be shown to be equivalent, and it will be seen that our form is more suggestive of how to proceed in solving the equation.

By applying the operator

$$\left.\frac{\delta^s}{\delta u(x_1)\cdots\delta u(x_s)}\right|_{u=0}$$

to Eq. (20) one can obtain the infinite system of "hierarchy" equations

$$\frac{\partial F_s}{\partial t} = [H_s; F_s] + \frac{1}{v}\int [\sum_{1\le i\le s} \phi(|q_i-q_{s+1}|); F_{s+1}]dx_{s+1}; \quad s=1,2,\cdots, \tag{22}$$

where

$$H_s = \sum_{i=1}^{s} T(p_i) + U_s; \quad U_s = \sum_{1\le i<j\le s} \phi(|q_i-q_j|); \quad s=1,2,3,\cdots. \tag{23}$$

The system (22) is equivalent to the single equation (20).

3. SOLUTION OF THE FUNCTIONAL DIFFERENTIAL EQUATION

In order to solve Eq. (20) we begin by examining the case of zero density, $1/v = 0$. We shall use superscripts "0" to denote this case. Thus[5]

$$L^0[t,w] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \int F_s^0(t,x_1,\cdots,x_s) \times w(x_1)\cdots w(x_s)dx_1\cdots dx_s, \tag{24}$$

$$\frac{\partial L^0}{\partial t} - \int w(x_1)\left[T(p_1); \frac{\delta L^0}{\delta w(x_1)}\right]dx_1 - \tfrac{1}{2}\int w(x_1)w(x_2) \times \left[\phi(|q_1-q_2|); \frac{\delta^2 L^0}{\delta w(x_1)\delta w(x_2)}\right]dx_1dx_2 = 0, \tag{25}$$

$$F_s^0(t,x_1,\cdots,x_s) = \left.\frac{\delta^s L^0}{\delta w(x_1)\cdots\delta w(x_s)}\right|_{w=0}, \tag{26}$$

and Eq. (22) reduces to

$$\frac{\partial F_s^0}{\partial t} = [H_s; F_s^0]; \quad s=1,2,\cdots. \tag{27}$$

The solution of Eq. (27) is immediately obtained in terms of the solution operator $S_t^{(s)}$ corresponding to the Hamiltonian, H_s. It is given by

$$F_s^0(t,x_1,\cdots,x_s) = S_{-t}^{(s)} F_s^0(0,x_1,\cdots,x_s); \quad s=1,2,\cdots. \tag{28}$$

On inserting this expression in Eq. (24) we obtain the solution of Eq. (25) subject to the initial conditions

$$L^0[0,w] = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} \int F_s^0(0,x_1,\cdots,x_s) \times w(x_1)\cdots w(x_s)dx_1\cdots dx_s, \tag{29}$$

where the $F_s^0(0,x_1,\cdots,x_s)$ are the given initial data.

In order to solve the general equation (20), we observe that the form of the latter suggests that we try a solution of the form

$$L[t,u] = L^0[t,w]; \quad w(x) = u(x) + (1/v). \tag{30}$$

Then

$$\frac{\delta L}{\delta u(x_1)} = \frac{\delta L^0}{\delta w(x_1)}, \quad \frac{\delta^2 L}{\delta u(x_1)\delta u(x_2)} = \frac{\delta^2 L^0}{\delta w(x_1)\delta w(x_2)}, \quad \frac{\partial L}{\partial t} = \frac{\partial L^0}{\partial t},$$

and inserting in Eq. (20) we see at once that that

[5] It is convenient now to denote the arbitrary testing functions by w instead of u.

equation is satisfied by virtue of the fact that L^0 satisfies Eq. (25).

But Eq. (20) must be solved subject to the initial conditions

$$L[0,u]=\mathfrak{L}[u]=1+\sum_{s=1}^{\infty}\frac{1}{s!}\int F_s(0,x_1,\cdots,x_s)\times u(x_1)\cdots u(x_s)dx_1\cdots dx_s. \quad (31)$$

The functions $F_s(0,x_1,\cdots,x_s)$ are the given initial data. In terms of Eq. (30) this becomes

$$L^0[0,w]=\mathfrak{L}[u];\quad u=w-(1/v). \quad (32)$$

The main result of this paper is the general solution of Eq. (20) defined by Eq. (30). If $L^0[t,w]$ is the solution of the initial value problem for Eq. (25) with initial conditions presented in Eq. (32), then $L[t,u]$ is the solution of the initial value problem for the general equation (20) with initial conditions presented in Eq. (31).

The method we have used in obtaining this solution closely resembles the method devised by Zumino[2] to solve the corresponding functional differential equation for the equilibrium case. This is discussed in Sec. 5.

We now proceed to obtain expansions of the functions $F_s(t,x_1,\cdots,x_s)$ in powers of the density, $1/v$. For this purpose it is convenient to introduce a functional of two variables[6]

$$Q[t,u,w]=1+\sum_{k=1}^{\infty}\frac{1}{k!}\int S_{-t}{}^{(k)}\frac{\delta^k\mathfrak{L}}{\delta u(x_1)\cdots\delta u(x_k)}\times w(x_1)\cdots w(x_k)dx_1\cdots dx_k. \quad (33)$$

Now from Eqs. (24) and (32)

$$F_k{}^0(0,x_1,\cdots,x_k)=\frac{\delta^k L^0[0,w]}{\delta w(x_1)\cdots\delta w(x_k)}\bigg|_{w=0}=\frac{\delta^k\mathfrak{L}}{\delta u(x_1)\cdots\delta u(x_k)}\bigg|_{u=-1/v}.$$

Hence from Eq. (28)

$$Q[t,u,w]\bigg|_{u=-1/v}=1+\sum_{k=1}^{\infty}\frac{1}{k!}\int F_k{}^0(t,x_1,\cdots x_k)\times w(x_1)\cdots w(x_k)dx_1\cdots dx_k=L^0[t,w]. \quad (34)$$

From Eqs. (21), (30), and (34)

$$F_s(t,x_1,\cdots,x_s)=\frac{\delta^s L^0[t,w]}{\delta w(x_1)\cdots\delta w(x_s)}\bigg|_{w=1/v}=\frac{\delta^s Q}{\delta w(x_1)\cdots\delta w(x_s)}\bigg|_{w=1/v,u=-1/v}. \quad (35)$$

[6] This functional is needed in the analysis in order to avoid expressions involving divergent integrals.

Now, from Eq. (33)

$$\frac{\delta^s Q}{\delta w(x_1)\cdots\delta w(x_s)}=\sum_{j=0}^{\infty}\frac{1}{j!}\int S_{-t}{}^{(j+s)}\frac{\delta^{j+s}\mathfrak{L}}{\delta u(x_1)\cdots\delta u(x_{j+s})}\times w(x_{s+1})\cdots w(x_{s+j})dx_{s+1}\cdots dx_{s+j}, \quad (36)$$

and from Eq. (31)

$$\frac{\delta^{j+s}\mathfrak{L}}{\delta u(x_1)\cdots\delta u(x_{j+s})}=\sum_{n=0}^{\infty}\frac{1}{n!}\int F_{n+j+s}(0,x_1,\cdots,x_{n+j+s})\times u(x_{j+s+1})\cdots u(x_{j+s+n})dx_{j+s+1}\cdots dx_{j+s+n}. \quad (37)$$

We now insert Eq. (37) in Eq. (36); the resulting double series can be rearranged and evaluated for $u=-w$. We obtain

$$\frac{\delta^s Q}{\delta w(x_1)\cdots\delta w(x_s)}\bigg|_{u=-w}=\sum_{k=0}^{\infty}\sum_{j=0}^{k}\frac{(-1)^{k-j}}{j!(k-j)!}\int S_{-t}{}^{(j+s)}F_{k+s}(0,x_1,\cdots,x_{k+s})\times w(x_{s+1})\cdots w(x_{s+k})dx_{s+1}\cdots dx_{s+k}.$$

In this integral we introduce the transformation

$$x_{j+s+1}=S_{-t}{}^{(1)}x_{j+s+1}',\cdots,x_{s+k}=S_{-t}{}^{(1)}x_{s+k}'.$$

We then interchange integration and summation over the index j, and set $w=1/v$. The result is

$$F_s(t,x_1,\cdots,x_s)=\sum_{k=0}^{\infty}\left(\frac{1}{v}\right)^k\int\left[\sum_{j=0}^{k}\frac{(-1)^{k-j}}{j!(k-j)!}T_{-t}{}^{(j+s)}\times F_{k+s}(0,x_1,\cdots,x_{k+s})\right]dx_{s+1}\cdots dx_{s+k}, \quad (38)$$

$$s=1,2,\cdots,$$

where the operator $T_t{}^{(s)}$ is defined by

$$T_t{}^{(s)}g(z,x_1,\cdots,x_{s+m})=g(z,S_t{}^{(s)}\{x_1,\cdots,x_s\},S_t{}^{(1)}x_{s+1},\cdots,S_t{}^{(1)}x_{s+m}). \quad (39)$$

This is our expansion of F_s as a power series in the density.

In order to check the results, one can show in a straightforward manner that Eq. (38) satisfies the system of Eqs. (22). To verify that the initial conditions are satisfied, we may set $t=0$ in Eq. (38). Since $T_0{}^{(s)}$ is the identity operator, the integrand in Eq. (38) reduces to

$$F_{k+s}(0,x_1,\cdots,x_{k+s})\sum_{j=0}^{k}\frac{(-1)^{k-j}}{j!(k-j)!}.$$

But

$$\sum_{j=0}^{k}\frac{(-1)^{k-j}}{j!(k-j)!}=\frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}=\frac{1}{k!}(1-1)^k=0,$$

for $k=1, 2, \cdots$. (40)

Thus every term in Eq. (38) vanishes, except the first, and the series reduces to $F_s(0,x_1,\cdots,x_s)$ as required.

The series expansion (38) is a very useful form of the solution. We observe that for small densities ($1/v\ll 1$), the function F_s is approximated by terminating the series after a few terms. Now the functions F_s of main interest are those for which s is small, and for these functions, the calculation of the leading terms of the expansion requires a knowledge only of solution operators $S_t^{(k)}$ where k is small and initial data $F_j(0,x_1,\cdots,x_j)$ where j is small.

The leading terms of Eq. (38) are given by

$$F_s(t,x_1,\cdots,x_s)=T_{-t}^{(s)}F_s(0,x_1,\cdots,x_s)+\frac{1}{v}\int[T_{-t}^{(s+1)}F_{s+1}(0,x_1,\cdots,x_{s+1})-T_{-t}^{(s)}F_{s+1}(0,x_1,\cdots,x_{s+1})]dx_{s+1}+O(1/v^2). \quad (41)$$

Bogoliubov[1] obtains the equation

$$F_s(t,x_1,\cdots,x_s)=S_{-t}^{(s)}F_s(0,x_1,\cdots,x_s)+\frac{1}{v}\int_0^t\left\{S_{\tau-t}^{(s)}\int[\sum_{1\le i\le s}\phi(|q_i-q_{s+1}|);S_{-\tau}^{(s+1)}\times F_{s+1}(0,x_1,\cdots,x_{s+1})]dx_{s+1}\right\}d\tau+O(1/v^2). \quad (42)$$

With a little manipulation it is possible to reduce Eq. (42) to the simpler form of Eq. (41).

4. DERIVATION OF THE EQUILIBRIUM EQUATION

In this section we shall derive the well-known functional differential equation for the equilibrium case by reduction from the general equation (20). The first step is to derive a new form of Eq. (20) by expanding the Poisson brackets that appear in that equation [as is done in Eq. (6)] and by using the following identity which is obtained by interchanging integration variables:

$$\int\left[u(x_1)+\frac{1}{v}\right]\left[u(x_2)+\frac{1}{v}\right]\frac{\partial\phi(|q_1-q_2|)}{\partial q_2^{\alpha}}\times\frac{\partial}{\partial p_2^{\alpha}}\frac{\delta^2L}{\delta u(x_1)\delta u(x_2)}dx_1dx_2=\int\left[u(x_1)+\frac{1}{v}\right]\left[u(x_2)+\frac{1}{v}\right]\frac{\partial\phi(|q_1-q_2|)}{\partial q_1^{\alpha}}\times\frac{\partial}{\partial p_1^{\alpha}}\frac{\delta^2L}{\delta u(x_1)\delta u(x_2)}dx_1dx_2. \quad (43)$$

With the aid of Eq. (43), Eq. (20) now becomes

$$\frac{\partial L}{\partial t}=\sum_{\alpha=1}^{3}\left\{-\frac{1}{m}\int\left[u(x_1)+\frac{1}{v}\right]p_1^{\alpha}\frac{\partial}{\partial q_1^{\alpha}}\frac{\delta L}{\delta u(x_1)}dx_1+\int\left[u(x_1)+\frac{1}{v}\right]\left[u(x_2)+\frac{1}{v}\right]\frac{\partial\phi(|q_1-q_2|)}{\partial q_1^{\alpha}}\times\frac{\partial}{\partial p_1^{\alpha}}\frac{\delta^2L}{\delta u(x_1)\delta u(x_2)}dx_1dx_2\right\}. \quad (44)$$

Let us now consider time-independent solutions

$$L[u]=1+\sum_{s=1}^{\infty}\frac{1}{s!}\int F_n(x_1,\cdots,x_s)u(x_1)\cdots\times u(x_s)dx_1\cdots dx_s. \quad (45)$$

Then $\partial L/\partial t=0$, and Eq. (44) will be satisfied if

$$\sum_{\alpha=1}^{3}\left\{-\frac{1}{m}p_1^{\alpha}\frac{\partial}{\partial q_1^{\alpha}}\frac{\delta L}{\delta u(x_1)}+\int\left[u(x_2)+\frac{1}{v}\right]\times\frac{\partial\phi(|q_1-q_2|)}{\partial q_1^{\alpha}}\frac{\partial}{\partial p_1^{\alpha}}\frac{\delta^2L}{\delta u(x_1)\delta u(x_2)}dx_2\right\}=0. \quad (46)$$

This equation is an identity in $x_1=(q_1,p_1)$. It is sufficient for Eq. (44), but not necessary.

Following a suggestion of Zumino,[2] let us now consider solutions of Eq. (46) for which

$$F_s(x_1,\cdots,x_s)=c^{-s}\exp\left[-\frac{1}{2m\theta}[p_1^2+\cdots+p_s^2]\right]f_s(q_1,\cdots,q_s), \quad (47)$$

where θ is a constant and

$$c=\int\exp[-p^2/2m\theta]dp. \quad (48)$$

Let $\langle L\rangle[u]$ denote the restriction of $L[u]$ to the domain of functions $u=u(q)$ which are independent of p. Then if Eq. (47) is assumed,

$$L[u]=1+\sum_{k=1}^{\infty}\frac{1}{k!}\int f_k(q_1,\cdots,q_k)\prod_{i=1}^{k}c^{-1}\times\exp\left[\frac{-p_i^2}{2m\theta}\right]u(x_i)dx_1\cdots dx_k \quad (49)$$

and

$$\langle L\rangle[u]=1+\sum_{k=1}^{\infty}\frac{1}{k!}\int f_k(q_1,\cdots,q_k)u(q_1)\cdots\times u(q_k)dq_1\cdots dq_k. \quad (50)$$

By functional differentiation of Eqs. (49) and (50) it is

easy to show that

$$\frac{\delta^s\langle L\rangle}{\delta u(x_1)\cdots\delta u(x_s)}=c^{-s}\prod_{i=1}^{s}\exp\left[-\frac{p_i^2}{2m\theta}\right]\frac{\delta^s\langle L\rangle}{\delta u(q_1)\cdots\delta u(q_s)}, \tag{51}$$

and

$$\frac{\partial}{\partial p_1{}^\alpha}\frac{\delta^2 L}{\delta u(x_1)\delta u(x_2)}=-\frac{p_1{}^\alpha}{m\theta}\frac{\delta^2 L}{\delta u(x_1)\delta u(x_2)}. \tag{52}$$

With the aid of Eq. (52), Eq. (46) becomes

$$\sum_{\alpha=1}^{3}p_1{}^\alpha\left\{\frac{\partial}{\partial q_1{}^\alpha}\frac{\delta L}{\delta u(x_1)}+\frac{1}{\theta}\int\left[u(x_2)+\frac{1}{v}\right]\times\frac{\partial\phi(|q_1-q_2|)}{\partial q_1{}^\alpha}\frac{\delta^2 L}{\delta u(x_1)\delta u(x_2)}dx_2\right\}=0. \tag{53}$$

If we restrict Eq. (53) to functions $u=u(q)$ and use Eq. (51) we obtain

$$c^{-1}\exp\left[-\frac{p_1^2}{2m\theta}\right]\sum_{\alpha=1}^{3}p_1{}^\alpha\left\{\frac{\partial}{\partial q_1{}^\alpha}\frac{\delta\langle L\rangle}{\delta u(q_1)}+\frac{1}{\theta}\int\left[u(q_2)+\frac{1}{v}\right]\frac{\partial\phi(|q_1-q_2|)}{\partial q_1{}^\alpha}\frac{\delta^2\langle L\rangle}{\delta u(q_1)\delta u(q_2)}\times\left[\int c^{-1}\exp\left[-\frac{p_2^2}{2m\theta}\right]dp_2\right]dq_2\right\}=0, \tag{54}$$

and since p_1 is arbitrary,

$$\frac{\partial}{\partial q_1{}^\alpha}\frac{\delta\langle L\rangle}{\delta u(q_1)}+\frac{1}{\theta}\int\left[u(q_2)+\frac{1}{v}\right]\frac{\partial\phi(|q_1-q_2|)}{\partial q_1{}^\alpha}\times\frac{\delta^2\langle L\rangle}{\delta u(q_1)\delta u(q_2)}dq_2=0;\quad\alpha=1,2,3. \tag{55}$$

Equation (55) is the well-known[1a,2] equation of equilibrium theory, where $\theta=kT$, k is the Boltzmann constant, and T is the absolute temperature. This equation is usually derived from an assumption about the explicit form of D_n. This form is given by

$$D_n=Z_n^{-1}\exp\left[-\frac{1}{\theta}H_n\right];\quad H_n=\sum_{j=1}^{n}T(p_j)+U_n; \tag{56}$$

where

$$Z_n=\int_{\Omega V}\exp\left[-\frac{1}{\theta}H_n\right]dx_1\cdots dx_n=Q_nc^n;$$
$$Q_n=\int_V\exp\left[-\frac{1}{\theta}U_n\right]dq_1\cdots dq_n. \tag{57}$$

The purpose of this section has been to show that the equilibrium equation (55) can be derived from the general equation (20) by using the assumption presented in Eq. (47). It is not surprising that this can be done, in view of the fact that Eq. (47) is a consequence of Eq. (56). To see this, we use Eqs. (19) and (9) to obtain

$$\begin{aligned}F_s(t,x_1,\cdots,x_s)&=\lim_{\substack{n\to\infty\\V\to\infty}}F_{n,s}=\lim_{\substack{n\to\infty\\V\to\infty}}V^s\int_{\Omega V}Z_n^{-1}\times\exp\left[-\frac{1}{\theta}H_n\right]dx_{s+1}\cdots dx_n\\&=\lim_{\substack{n\to\infty\\V\to\infty}}V^sc^{-s}\exp\left[-\frac{1}{2m\theta}(p_1^2+\cdots+p_s^2)\right]\times\int_V Q_n^{-1}\exp\left[-\frac{1}{\theta}U_n\right]dq_{s+1}\cdots dq_n.\end{aligned} \tag{58}$$

From Eq. (58) we see at once that Eq. (47) follows with

$$f_s(q_1,\cdots,q_s)=\lim_{\substack{n\to\infty\\V\to\infty}}V^s\int_V Q_n^{-1}\exp\left[-\frac{1}{\theta}U_n\right]dq_{s+1}\cdots dq_n. \tag{59}$$

Before proceeding to the solution of Eq. (55) we point out that that equation is also equivalent to an infinite system of equations, given by[1a]

$$\frac{\partial f_k}{\partial q_1{}^\alpha}+\frac{1}{\theta}\frac{\partial U_k}{\partial q_1{}^\alpha}f_k+\frac{1}{\theta v}\int\frac{\partial\phi(|q_1+q_{k+1}|)}{\partial q_1{}^\alpha}f_{k+1}dq_{k+1}=0,\qquad\alpha=1,2,3;\quad k=1,2,\cdots. \tag{60}$$

5. SOLUTION OF THE EQUILIBRIUM EQUATION

In this section we shall solve the equilibrium equation (55) by a slight simplification of a method by Zumino.[2] As in Sec. 3 we begin by examining the case of zero density, $1/v=0$. We again use superscripts "0" to denote this case. Thus

$$\langle L^0\rangle[w]=1+\sum_{k=1}^{\infty}\frac{1}{k!}\int f_k{}^0(q_1,\cdots,q_k)\times w(q_1)\cdots w(q_k)dq_1\cdots dq_k, \tag{61}$$

$$\frac{\partial}{\partial q_1{}^\alpha}\frac{\delta\langle L^0\rangle}{\delta w(q_1)}+\frac{1}{\theta}\int w(q_2)\frac{\partial\phi(|q_1-q_2|)}{\partial q_1{}^\alpha}\times\frac{\delta^2\langle L^0\rangle}{\delta w(q_1)\delta w(q_2)}dq_2=0;\quad\alpha=1,2,3, \tag{62}$$

$$f_s{}^0(q_1,\cdots,q_s)=\left.\frac{\delta^s\langle L^0\rangle}{\delta w(q_1)\cdots\delta w(q_s)}\right|_{w=0}, \tag{63}$$

and Eq. (60) reduces to

$$\frac{\partial f_k{}^0}{\partial q_1{}^\alpha}+\frac{1}{\theta}\frac{\partial U_k}{\partial q_1{}^\alpha}f_k{}^0=0;\quad\alpha=1,2,3;\quad k=1,2,\cdots. \tag{64}$$

In order to solve Eq. (64) ,set

$$f_k^0=C_k(q_1,\cdots,q_k)\exp\left[-\frac{1}{\theta}U_k\right];\quad k=1,2,\cdots. \tag{65}$$

From Eq. (64),

$$\partial C_k/\partial q_1{}^\alpha=0;\quad \alpha=1,2,3;\quad k=1,2,\cdots. \tag{66}$$

Since $C_k(q_1,\cdots,q_k)$ is symmetric in its arguments, it follows that C_k is a constant. In order to determine the constant, we observe that by letting $n\to\infty$, $V\to\infty$ in Eq. (10) we obtain

$$\lim_{V\to\infty}\frac{1}{V^s}\int_{\Omega V}F_s dx_1\cdots dx_s+1. \tag{67}$$

Now from Eq. (47)

$$\lim_{V\to\infty}\frac{1}{V^s}\int_V f_s dq_1\cdots dq_s=1, \tag{68}$$

and from Eq. (65)

$$\lim_{V\to\infty}\frac{1}{V^s}\int_V\exp\left[-\frac{1}{\theta}U_s\right]dq_1\cdots dq_s=\frac{1}{C_s}. \tag{69}$$

It is clear from Eq. (69) that $C_s=1$ for potentials $\phi(r)$ which vanish sufficiently rapidly as $r\to\infty$. We shall therefore impose as a condition on ϕ that the left-hand side of Eq. (69) be equal to 1 for $s=1, 2, 3, \cdots$. It follows now from Eq. (65) that

$$f_k^0(q_1,\cdots,q_k)=\exp\left[-\frac{1}{\theta}U_k\right];\quad k=1,2,\cdots. \tag{70}$$

The solution of Eq. (55) for nonzero density can be obtained from the zero density solution in a manner very similar to the procedure used in the nonequilibrium case. We begin with a trial form of the solution slightly more general than the one used in Sec. 3:

$$\langle L\rangle[u]=\langle L^0\rangle[w];\quad w=a(u+1/v);\quad a=\text{const.} \tag{71}$$

By functional differentiation we have

$$\frac{\delta^s\langle L\rangle}{\delta u(q_1)\cdots\delta u(q_s)}=a^s\frac{\delta^s\langle L^0\rangle}{\delta w(q_1)\cdots\delta w(q_s)}, \tag{72}$$

and substituting in Eq. (55) we see that the latter equation is satisfied because $\langle L^0\rangle$ satisfies Eq. (62).

In order to determine the constant, a, we observe first that since U_n is a function only of the coordinate differences (q_i-q_j), Eq. (59) implies that $f_1(q_1)$ is a constant, and Eq. (68) implies that the constant is unity. Thus

$$f_1(q_1)=1. \tag{73}$$

Now from Eqs. (73) and (50) it follows that

$$\left.\frac{\delta\langle L\rangle}{\delta u(q_1)}\right|_{u=0}=1;\quad \langle L\rangle[0]=1. \tag{74}$$

This in turn implies that

$$a\left.\frac{\delta\langle L^0\rangle}{\delta w(q_1)}\right|_{w=a/v}=1;\quad \langle L^0\rangle\left[\frac{a}{v}\right]=1. \tag{75}$$

We shall see that Eq. (75) suffices to determine the constant a.

Now from Eqs. (50) and (72),

$$f_s(q_1,\cdots,q_s)=\left.\frac{\delta^s\langle L\rangle}{\delta u(q_1)\cdots\delta u(q_s)}\right|_{u=0}=a^s\left.\frac{\delta^s\langle L^0\rangle}{\delta w(q_1)\cdots\delta w(q_s)}\right|_{w=a/v}. \tag{76}$$

It would appear that we need only differentiate Eq. (61) s times and set $w=a/v$ to obtain an explicit formula for $f_s(q_1,\cdots,q_s)$. However, this is incorrect because $f_k^0\approx 1$ for large $|q_i-q_j|$ and the integrals in Eq. (61) converge only for testing functions $w(q)$ which vanish sufficiently rapidly at infinity. For $w=a/v$, the integrals diverge. The difficulty is that Eq. (61) does not represent the functional $\langle L^0\rangle[w]$ in a sufficiently large domain of functions $w(q)$. What is needed is an "analytical continuation" of the representation of the functional.

Such an analytic continuation can be obtained by the following transformation which was suggested by Zumino.[2]

$$\langle L^0\rangle[w]=\exp[\langle M^0\rangle[w]],\quad \langle M^0\rangle[w]=\log\langle L^0\rangle[w], \tag{77}$$

where

$$\langle M^0\rangle[w]=\sum_{k=1}^{\infty}\frac{1}{k!}\int g_k^0(q_1,\cdots,q_k)\times w(q_1)\cdots w(q_k)dq_1\cdots dq_k. \tag{78}$$

Now

$$f_1^0(q_1)=\left.\frac{\delta\langle L^0\rangle}{\delta w(q_1)}\right|_{w=0}=\left[\exp[\langle M^0\rangle[w]]\frac{\delta\langle M^0\rangle}{\delta w(q_1)}\right]_{w=0}=\left.\frac{\delta\langle M^0\rangle}{\delta w(q_1)}\right|_{w=0}=g_1^0(q_1).$$

Proceeding in this manner, we may obtain the following

relations between the f_k^0 and g_k^0:

$$\begin{aligned}
f_1^0(q_1)&=g_1^0(q_1),\\
f_2^0(q_1,q_2)&=g_1^0(q_1)g_1^0(q_2)+g_2^0(q_1,q_2),\\
f_3^0(q_1,q_2,q_3)&=g_1^0(q_1)g_1^0(q_2)g_1^0(q_3)\\
&\quad+g_2^0(q_1,q_2)g_1^0(q_3)+g_2^0(q_2,q_3)g_1^0(q_1)\\
&\quad+g_2^0(q_1,q_3)g_1^0(q_2)+g_3^0(q_1,q_2,q_3),
\end{aligned}\tag{79}$$

$$\begin{aligned}
g_1^0(q_1)&=f_1^0(q_1),\\
g_2^0(q_1,q_2)&=f_2^0(q_1,q_2)-f_1^0(q_1)f_1^0(q_2),\\
g_3^0(q_1,q_2,q_3)&=f_3^0(q_1,q_2,q_3)-f_1^0(q_1)f_2^0(q_2,q_3)\\
&\quad-f_1^0(q_2)f_2^0(q_1,q_3)-f_1^0(q_3)f_2^0(q_1,q_2)\\
&\quad+2f_1^0(q_1)f_1^0(q_2)f_1^0(q_3),
\end{aligned}\tag{80}$$

etc. Functions g_k^0 related to the f_k^0 in this manner are known in statistical mechanics as Ursell functions. We observe that except for g_1^0 they vanish for large values of $|q_i-q_j|$.

Let

$$z=a/v. \tag{81}$$

From Eqs. (77) and (75),

$$\left.\frac{\delta\langle M^0\rangle}{\delta w(q_1)}\right|_{w=z}=\left.\frac{1}{\langle L^0\rangle[z]}\frac{\delta\langle L^0\rangle}{\delta w(q_1)}\right|_{w=z}=\frac{1}{a}; \tag{82}$$

and from Eqs. (76), (77), and (75)

$$\begin{aligned}
f_2(q_1,q_2)&=a^2\left.\frac{\delta^2\langle L^0\rangle}{\delta w(q_1)\delta w(q_2)}\right|_{w=z}\\
&=a^2\langle L^0\rangle[z]\left\{\frac{\delta\langle M^0\rangle}{\delta w(q_1)}\frac{\delta\langle M^0\rangle}{\delta w(q_2)}+\frac{\delta^2\langle M^0\rangle}{\delta w(q_1)\delta w(q_2)}\right\}\Bigg|_{w=z}\\
&=a^2\left\{\frac{1}{a^2}+\left.\frac{\delta^2\langle M^0\rangle}{\delta w(q_1)\delta w(q_2)}\right|_{w=z}\right\}.
\end{aligned}\tag{83}$$

Thus from Eq. (78)

$$\begin{aligned}
f_2(q_1,q_2)&=1+a^2\sum_{k=0}^{\infty}\frac{1}{k!}z^k\\
&\quad\times\int q_{k+2}^0(q_1,\cdots,q_{k+2})dq_3\cdots dq_{k+2},
\end{aligned}\tag{84}$$

and from Eq. (82)

$$1+\sum_{k=1}^{\infty}\frac{1}{k!}z^k\int g_{k+1}^0(q,q_1,\cdots,q_k)dq_1\cdots dq_k=\frac{1}{a}. \tag{85}$$

By virtue of the remark made at the end of the last paragraph, we see that the integrals appearing in Eqs. (84) and (85) are convergent. If we set

$$b_k=\frac{1}{k!}\int g_k^0(q,q_1,\cdots,q_{k-1})dq_1\cdots dq_{k-1};\qquad k=2,3,\cdots;\quad b_1=1; \tag{86}$$

then Eq. (85) takes the form

$$v\sum_{k=1}^{\infty}kb_kz^k=1. \tag{87}$$

The b_k are called "cluster integrals" and are independent of q because the f_k^0, and hence the g_k^0, are functions only of the coordinate differences. The quantity z is called the "activity." Equation (84) expresses f_2 as a power series in the activity, and the latter is related to the density $1/v$ by Eq. (87). In order to obtain an expression for f_2 as a power series in the density, we assume that

$$z=\sum_{j=1}^{\infty}\frac{a_j}{v^j} \tag{88}$$

and insert in Eq. (87). One obtains easily

$$z=\frac{1}{v}-\frac{2b_2}{v^2}+O(1/v^3), \tag{89}$$

and inserting in Eq. (84) we obtain

$$\begin{aligned}
f_2(q_1,q_2)&=1+g_2^0(q_1,q_2)+\frac{1}{v}\Big[\int g_3^0(q_1,q_2q_3)dq_3\\
&\quad-2g_2^0(q_1,q_2)\int g_2^0(q_1,q_3)dq_3\Big]+O(1/v^2).
\end{aligned}\tag{90}$$

It can easily be shown that this result agrees with previously given expressions for f_2, and can be used to obtain the virial expansion of the equation of state to order $1/v$. Formulas for f_s for $s>2$ can be obtained by an obvious generalization of the method used for f_2. However, the solution of the functional differential equation, Eq. (55), is in principle already given by Eq. (71), where $\langle L^0\rangle$ is given by Eqs. (61) and (70), and a is determined by Eq. (87).

6. BOLTZMANN EQUATION

The purpose of this section is to present a simple derivation of the Boltzmann equation based on the results of Sec. 3.

In recent years, several authors have given derivations of the Boltzmann equation based on the hierarchy equations (22). Kirkwood[7,9] has used the hierarchy equation

[7] I. Prigogine, *Proceedings of the International Symposium on Transport Processes in Statistical Mechanics Held in Brussels, August 27–31, 1956* (Interscience Publishers, Inc., New York, 1958).

for F_1, and by means of "phase space transformation functions" has shown that the Boltzmann equation is satisfied by a function obtained from F_1 by time-averaging. Bogoliubov[1a] has obtained the Boltzmann equation for F_1 by assuming that for $s>1$, F_s depends on the time only through a functional dependence on F_1. Green[8] has obtained the Boltzmann equation for the spatially uniform case by means of "Ursell function" expansions of solutions of the hierarchy equations.

For the present derivation we are indebted to all of these, but our task is greatly simplified by the fact that we have at our disposal the explicit solution of the initial value problem for the hierarchy equations obtained in Sec. 3. The leading term of the expansion of that solution consists of an integral which already closely resembles the Boltzmann collision integral. Following the method of Green, we simplify this integral by making use of the fact that in all but a small portion of the initial configuration space of two particles, two-body interactions can be described in terms of complete collisions. Under the assumptions usually made in deriving the Boltzmann equation we find, in agreement with Kirkwood, that that equation is satisfied by a time-averaged density function. It appears likely that by using further terms of our expansion, which involve interactions of more than two particles, the method used here can be extended to obtain generalizations of the Boltzmann equation to higher densities.

We begin by writing Eq. (41) for the case $s=1$:

$$F_1(t,x_1)=T_{-t}^{(1)}F_1(0,x_1)+\frac{1}{v}\int[T_{-t}^{(2)}F_2(0,x_1,x_2)-T_{-t}^{(1)}F_2(0,x_1,x_2)]dx_2+O(1/v^2). \quad (91)$$

Since initial data may be specified at an arbitrary time, Eq. (91) may be rewritten as

$$F_1(t+\tau,x_1)=T_{-\tau}^{(1)}F_1(t,x_1)+\frac{1}{v}\int[T_{-\tau}^{(2)}F_2(t,x_1,x_2)-T_{-\tau}^{(1)}F_2(t,x_1,x_2)]dx_2+O(1/v^2). \quad (92)$$

Since Eq. (92) is an identity in x_1, we may replace x_1 by $S_\tau^{(1)}x_1$. If we then introduce the transformation of the integration variable $x_2=S_\tau^{(1)}x_2'$, we obtain

$$\frac{1}{\tau}[F_1(t+\tau,S_\tau^{(1)}x_1)-F_1(t,x_1)]=\frac{1}{\tau v}\int[S_{-\tau}^{(2)}F_2(t,S_\tau^{(1)}x_1,S_\tau^{(1)}x_2)-F_2(t,x_1,x_2)]dq_2dp_2+O(1/v^2). \quad (93)$$

[8] M. S. Green, J. Chem. Phys. **25**, 836 (1956).
[9] J. G. Kirkwood, J. Chem. Phys. **15**, 72 (1947).

We now introduce a time average[10] of F_1 defined by

$$\langle F\rangle_1(t,x_1)=\frac{1}{\tau}\int_0^\tau F_1(t+s,S_s^{(1)}x_1)ds. \quad (94)$$

We shall show that $\langle F_1\rangle$ satisfies the Boltzmann equation under the following assumptions:

(I) $\phi(r)\approx 0$ for $r\geq r_1$.

(II) $F_2(t,x_1,x_2)\approx F_1(t,x_1)F_1(t,x_2)$ for $|q_2-q_1|\geq r_2$, for some $r_2>0$.

(III) $v\gg 1$.

(IV) $F_1(t,q_1+\Delta q_1,p_1)\approx F_1(t,q_1,p_1)$ where $|\Delta q_1|\gg r_1,r_2$.

Expression (II) is a generalized form of the molecular chaos assumption. This form of the assumption and the general technique which will be used to obtain the collision integral were suggested by a paper of Green.[8] A detailed discussion of the assumption, which is part of Green's "product condition" is contained in his article.[8] In (IV) we assume that F_1 does not change appreciably under translations Δq_1 where $|\Delta q_1|$ is large compared to both r_1 (the range of intermolecular forces) and r_2 (the correlation distance). F_1 may, however, change considerably over distances of the order of a mean free path. Conditions (I) and (III) are self-explanatory. Expressions (I)–(IV) are the assumptions usually made in deriving the Boltzmann equation. The earlier derivations required that (II) hold for all values of $|q_2-q_1|$.

Let

$$D_t\langle F\rangle_1(t,x_1)=\frac{\partial\langle F\rangle_1}{\partial t}+\sum_{\alpha=1}^{3}\frac{p_1^\alpha}{m}\frac{\partial\langle F\rangle_1}{\partial q_1^\alpha}. \quad (95)$$

Then, since $S_z^{(1)}x_1=S_z^{(1)}(q_1,p_1)=[q_1+(z/m)p_1,p_1]$ we have

$$\begin{aligned}D_t\langle F\rangle_1(t,x_1)&=\frac{d}{dz}\langle F\rangle_1(t+z,S_z^{(1)}x_1)|_{z=0}\\&=\frac{1}{\tau}\int_0^\tau\frac{d}{dz}F_1(t+s+z,S_{s+z}^{(1)}x_1)ds|_{z=0}\\&=\frac{1}{\tau}\int_0^\tau\frac{d}{ds}F_1(t+s,S_s^{(1)}x_1)ds\\&=\frac{1}{\tau}[F_1(t+\tau,S_\tau^{(1)}x_1)-F_1(t,x_1)]. \quad (96)\end{aligned}$$

We recognize Eq. (96) as the left-hand side of Eq. (93).

In order to evaluate the right-hand side of Eq. (93) we choose fixed values of q_1, p_1, and p_2 and examine the integration with respect to q_2. The region of integration may be conveniently analyzed as in Fig. 1.

[10] The operation that leads from F_1 to $\langle F\rangle_1$ is sometimes called "coarse-graining" or "smoothing." The latter term would appear to be preferable. See footnote 7.

In Fig. 1, A, B, and C together form a cylindrical region with spherical end surfaces. Outside of this region $S_{-\tau}^{(2)}\{S_\tau^{(1)}x_1, S_\tau^{(1)}x_2\} = \{x_1, x_2\}$ and the integrand in Eq. (93) vanishes. For points in C, the operator $S_\tau^{(1)}$ maps q_2 into q_2'' (the position of $S_\tau^{(1)}q_2$ relative to fixed q_1) and $S_{-\tau}^{(2)}$ maps q_2'' into q_2'. For all points in C having the same orthogonal projection (given by the vector b) onto the plane P, the momenta p_1' and p_2' are the same, i.e., p_1' and p_2' are functions of p_1, p_2 and the "impact vector" b. From Eqs. (9) and (12) we have

$$D_t\langle F\rangle_1(t,q_1,p_1) = -\frac{1}{v}\int dp_2 \int_P dA[F_1(t,q_1,p_1')F_1(t,q_1,p_2') - F_1(t,q_1,p_1)F_1(t,q_1,p_2)] \times \left|\frac{p_2-p_1}{m}\right| + \frac{1}{\tau v}I + O\left(\frac{1}{v^2}\right). \quad (97)$$

The integral in Eq. (97) represents the contribution to the integral in Eq. (93) from the region C except for an error due to the fact that the ends of the cylinder C are not planes but spherical segments. This error, plus the contributions from the regions A and B, determine the term $(1/\tau v)I$. In Eq. (97) we have clearly made use of assumptions (II) and (IV). The volume element dq_2 has been replaced by $(\tau/m)|p_2-p_1|dA$ where dA is an element of area on the plane P. Now from Eq. (93)

$$F_1(t+s, S_s^{(1)}x_1) = F_1(t,x_1) + O(1/v). \quad (98)$$

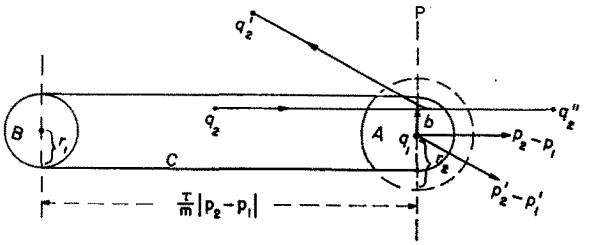


FIG. 1. Configuration space for a binary collision.

By inserting Eq. (98) in Eq. (94) we obtain

$$F_1(t,x_1) = \langle F\rangle_1(t,x_1) + O(1/v). \quad (99)$$

This result enables us to replace F_1 by $\langle F\rangle_1$ in Eq. (97), the only change being in the term of order $1/v^2$.

Let u_0 be the average relative velocity of two particles. If we now choose τ such that $\tau u_0 \gg r_j (j=1,2)$, the region C will be much larger than the regions A and B and the term $(1/\tau v)I$ may be neglected. By (III) we may omit the term of order $1/v^2$ and the resulting equation is

$$D_t\langle F\rangle_1(t,q_1,p_1) = -\frac{1}{v}\int dp_2 \int_P dA[\langle F\rangle_1(t,q_1,p_1')\langle F\rangle_1(t,q_1,p_2') - \langle F\rangle_1(t,q_1,p_1)\langle F\rangle_1(t,q_1,p_2)]\left|\frac{p_2-p_1}{m}\right|. \quad (100)$$

This equation is the Boltzmann equation.

Determination of Thermodynamic Green's Functions

GORDON BAYM*† AND N. DAVID MERMIN*
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts
(Received July 7, 1960)

In the study of thermodynamic correlation functions or Green's functions, one is naturally led to a calculation of values of the Fourier transform of the Green's function on a discrete set of points in the complex energy plane. It is shown that even though these points do not in general possess a limit point within the region of analyticity, one may still uniquely determine the Fourier transform of the Green's function directly from its values at these points.

RECENTLY, several authors[1] have approached the quantum-mechanical many-body problem with the aid of thermodynamic time-dependent correlation functions, or Green's functions. These are determined in actual calculation by integro-differential equations, derived from the field equations, together with a particular boundary condition in time. One of the features of this approach is that by a suitable extension of these functions to complex times and temperatures, the boundary condition becomes a condition of periodicity along a particular line in the complex time plane. This is taken into account by expressing the Green's function along this line in a Fourier series, and determining the Fourier coefficients. These turn out to be evaluations at a particular set of points of the extension of the Fourier transform of the Green's function to the complex energy plane. At this stage one would like to obtain the Fourier transform from a knowledge of the Fourier coefficients. If the Green's function is analytic at infinity in the complex energy plane, one is assured of a unique analytic continuation of the Fourier coefficients, because of the theorem that an analytic function is determined by its values on a set of points possessing a limit point within the region of analyticity. In general, however, the Green's function is not analytic at infinity. It is our purpose to provide the criteria by means of which this continuation may be uniquely inferred, in spite of the possibility of nonanalytic behavior at infinity.

We shall discuss only the density autocorrelation function, since this case illustrates all the essential features of our argument. First we extend this function to complex temperature and time, in order to state the usual boundary condition. We next review the argument relating the Fourier transform to the Fourier coefficients. Then we are in a position to show how the transform may be inferred from the coefficients. Our results are applicable to other Green's functions with only minor modifications in the discussion.

The thermodynamic density autocorrelation function is given by

$$F^{i\tau}(\mathbf{r}_1 t_1,\mathbf{r}_2 t_2)=\frac{\mathrm{Tr}e^{-i\tau(H-\mu N)}(\rho(\mathbf{r}_1 t_1)\rho(\mathbf{r}_2 t_2))_+}{\mathrm{Tr}e^{-i\tau(H-\mu N)}}, \quad (1)$$

where $\rho(\mathbf{r}t)$ is the density operator, $\psi\dagger(\mathbf{r}t)\psi(\mathbf{r}t)$, μ is the chemical potential, and $(\ )_+$ denotes the time ordered product. When $i\tau$ is real and positive and equals $\beta=1/KT$, the trace in Eq. (1) is just the grand canonical average, which we assume converges absolutely. As a result of this absolute convergence, F is analytic in the lower-half τ plane, so that if we know F in a suitable region of the lower-half τ plane, we can infer the grand canonical average for real β by analytic continuation. For convenience we restrict τ to the fourth quadrant, i.e., $\mathrm{Re}\tau>0$, $\mathrm{Im}\tau<0$.

For an isolated system F is a function only of the time difference t_1-t_2. We shall denote this difference by t, and shall also suppress explicit reference to the space variables. F is conveniently analyzed by separately considering the functions

$$\begin{aligned} F_>^{i\tau}(t)&=\frac{\mathrm{Tr}e^{-i\tau(H-\mu N)}\rho(t)\rho(0)}{\mathrm{Tr}e^{-i\tau(H-\mu N)}} \\ &=\frac{\mathrm{Tr}e^{-i\tau(H-\mu N)}e^{itH}\rho(0)e^{-itH}\rho(0)}{\mathrm{Tr}e^{-i\tau(H-\mu N)}}, \\ F_<^{i\tau}(t)&=\frac{\mathrm{Tr}e^{-i\tau(H-\mu N)}\rho(0)\rho(t)}{\mathrm{Tr}e^{-i\tau(H-\mu N)}} \\ &=\frac{\mathrm{Tr}e^{-i\tau(H-\mu N)}e^{-itH}\rho(0)e^{itH}\rho(0)}{\mathrm{Tr}e^{-i\tau(H-\mu N)}}, \end{aligned} \quad (2)$$

so that

$$F^{i\tau}=\begin{cases} F_>^{i\tau}(t), & t>0 \\ F_<^{i\tau}(t), & t<0. \end{cases} \quad (3)$$

Under the assumption of absolute convergence of the trace, $F_<^{i\tau}(t)$ converges absolutely in the closed region III+IV (see Fig. 1) of the complex t plane, and is, therefore, analytic and bounded in the open region III+IV. Similarly, $F_<^{i\tau}(t)$ converges absolutely in the closed region I+II, and is analytic in the open region I+II. We make the natural complex extension of F

* National Science Foundation Predoctoral Fellow.
† Present address: Institute for Theoretical Physics, Copenhagen, Denmark.
[1] P. C. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959); A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskii, J. Exptl. Theoret. Phys. (U.S.S.R.) **36**, 900 (1959) [English translation: Soviet Phys. JETP **36**, 636 (1959)]; E. S. Fradkin, Nuclear Phys. **12**, 465 (1959).

to the region II+IV defined by

$$F^{i\tau}(t)=\begin{cases}F_>{}^{i\tau}(t), & t \text{ in III+IV}\\ F_<{}^{i\tau}(t), & t \text{ in II},\end{cases} \tag{4}$$

i.e., we take the time ordering in Eq. (1) to be governed by the real part of t.

From the cyclic property of the trace [$\mathrm{Tr}(ABC)=\mathrm{Tr}(BCA)$, etc., we then obtain the boundary condition,

$$F_>{}^{i\tau}(t)=F_<{}^{i\tau}(t-\tau), \quad t \text{ in III+IV}. \tag{5}$$

The corresponding condition on the extended F is

$$F^{i\tau}(t)=F^{i\tau}(t-\tau), \quad t \text{ in IV}, \quad \mathrm{Re}t<\mathrm{Re}\tau. \tag{6}$$

Since $F_>{}^{i\tau}(t)$ is bounded in III+IV, its Fourier transform (in the generalized sense) exists[2] along every line in this region parallel to the real axis, and along the boundaries. Similarly, $F_<{}^{i\tau}(t)$ has a Fourier transform along every line in I+II parallel to the real axis, and along the boundaries. We then have

$$\begin{aligned}F_>{}^{i\tau}(t)&=\int_{-\infty}^{\infty}e^{-i\omega t}f_>{}^{i\tau}(\omega)\frac{d\omega}{2\pi}, \quad t \text{ in III+IV},\\ F_<{}^{i\tau}(t)&=\int_{-\infty}^{\infty}e^{-i\omega t}f_<{}^{i\tau}(\omega)\frac{d\omega}{2\pi}, \quad t \text{ in I+II}.\end{aligned} \tag{7}$$

The boundary condition of Eq. (5) implies that the Fourier transforms satisfy

$$f_>{}^{i\tau}(\omega)=e^{i\omega\tau}f_<{}^{i\tau}(\omega). \tag{8}$$

It is useful to introduce

$$A^{i\tau}(\omega)=i(1-e^{-i\omega\tau})f_>{}^{i\tau}(\omega), \tag{9}$$

in terms of which

$$F_>{}^{i\tau}(t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi i}e^{-i\omega t}\frac{A^{i\tau}(\omega)}{1-e^{-i\omega\tau}}, \quad t \text{ in III+IV}, \tag{10}$$

$$F_<{}^{i\tau}(t)=-\int_{-\infty}^{\infty}\frac{d\omega}{2\pi i}e^{-i\omega t}\frac{A^{i\tau}(\omega)}{1-e^{i\omega\tau}}, \quad t \text{ in I+II}, \tag{11}$$

and

$$F^{i\tau}(t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}e^{-i\omega t}\left[\int_{-\infty}^{\infty}\frac{d\omega'}{2\pi}\frac{A^{i\tau}(\omega')}{\omega+i\epsilon-\omega'}\frac{1}{1-e^{-i\omega'\tau}}+\int_{-\infty}^{\infty}\frac{d\omega'}{2\pi}\frac{A^{i\tau}(\omega')}{\omega-i\epsilon-\omega'}\frac{1}{1-e^{i\omega'\tau}}\right], \quad t \text{ real}. \tag{12}$$

From Eqs. (8) and (9) we note that

$$A^{i\tau}(\omega)=i\int_{-\infty}^{\infty}e^{i\omega t}\frac{\mathrm{Tr}e^{-i\tau(H-\mu N)}[\rho(t),\rho(0)]}{\mathrm{Tr}e^{-i\tau(H-\mu N)}}dt. \tag{13}$$

[2] M. J. Lighthill, *Introduction to Fourier Analysis and Generalised Functions* (Cambridge University Press, New York, 1958), Sec. 2.3.

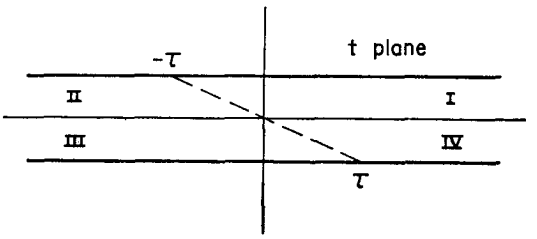


FIG. 1. The complex t plane.

Since

$$1/(\omega\pm i\epsilon)=P(1/\omega)\mp\pi i\delta(\omega),$$

Eq. (12) is equivalent to

$$F^{i\tau}(t)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}e^{-i\omega t}\left[\frac{\Phi^{i\tau}(\omega+i\epsilon)}{1-e^{-i\omega\tau}}+\frac{\Phi^{i\tau}(\omega-i\epsilon)}{1-e^{i\omega\tau}}\right], \quad t \text{ real}, \tag{14}$$

where the function Φ of the complex variable z is defined by

$$\Phi^{i\tau}(z)=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{A^{i\tau}(\omega)}{z-\omega}. \tag{15}$$

By virtue of the absolute convergence of the grand canonical averages, the integral of the absolute value of $A(\omega)$ exists, so that from Eq. (15) we have

(A) Φ is analytic off the real axis;

(B) Φ goes to zero as z approaches infinity along any straight line in the upper or lower half-plane.

The boundary condition of Eq. (6) is most simply taken into account by determining F along the line from $-\tau$ to τ (the dotted line in Fig. 1) in the form of a Fourier series with period τ:

$$F^{i\tau}(t)=\frac{1}{\tau}\sum_{\nu=\text{integer}}e^{-2\pi i\nu t/\tau}f_\nu{}^{i\tau}, \quad t/\tau \text{ real}, \quad -1\le t/\tau\le 1. \tag{16}$$

The Fourier coefficient $f_\nu{}^{i\tau}$ is given by

$$f_\nu{}^{i\tau}=\int_0^\tau e^{2\pi i\nu t/\tau}F^{i\tau}(t)dt=\int_0^\tau e^{2\pi i\nu t/\tau}F_>{}^{i\tau}(t)dt. \tag{17}$$

From Eq. (10) we then have

$$f_\nu=\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{A^{i\tau}(\omega)}{(2\pi\nu/\tau)-\omega}=\Phi^{i\tau}(2\pi\nu/\tau). \tag{18}$$

It should be noted that the Fourier coefficients $\{f_\nu{}^{i\tau}\}$ determine F everywhere, for they determine it on the line from $-\tau$ to τ, and hence on lines contained in its regions of analyticity. Therefore, F is determined in II+IV and, in particular, its boundary value on the real line is determined. In principle then, one can determine F from $\{f_\nu\}$ by summing the Fourier series and performing the necessary analytic continuations. However, it is also possible to take advantage of the

fact that f_ν is simply $\Phi(z)$ evaluated at $z=2\pi\nu/\tau$, and infer $\Phi(z)$ directly from its values at these points. Since Φ is analytic off the real axis, the problem is that of finding the correct analytic continuation from these values. Finding analytic continuations is not a problem in practice, since the result of a calculation is to give $\{f_\nu\}$ not as a sequence of numbers, but in a functional form $\bar{\Phi}(z)$ evaluated at $z=2\pi\nu/\tau$, ν a positive or negative integer. But the function $\bar{\Phi}(z)$ is clearly not unique since, for example, $\bar{\Phi}(z)$ and $\bar{\Phi}(e^{in\tau}z)$ take on the same values at $z=2\pi\nu/\tau$.

We do know, however, that the desired continuation, $\Phi(z)$, satisfies A and B. We shall show that these two requirements are sufficient to single out the correct continuation; i.e., if a $\bar{\Phi}(z)$ satisfies A and B, it is indeed the desired $\Phi(z)$.[3]

Suppose then that there were two functions, $\Phi_1(z)$ and $\Phi_2(z)$, both satisfying A and B and such that $\Phi_1(2\pi\nu/\tau)=\Phi_2(2\pi\nu/\tau)$, for all integral ν. The function $\varphi(z)=\Phi_1(z)-\Phi_2(z)$ therefore satisfies A and B and in addition $\varphi(2\pi\nu/\tau)=0$ for all integral ν. We know, however, from Carleman's theorem[4] that if a function $g(z)$ is analytic and bounded in the upper half-plane including the real axis, and has zeros in the upper half plane at $r_ne^{i\theta_n}$ with multiplicity α_n, but is not identically zero, then the series

$$S=\sum_{n=1}^{\infty}(\sin\theta_n)\alpha_n/r_n$$

is convergent. From A and B it follows that $\varphi(z)$ is bounded in any region bounded away from the real axis. Therefore, the function $g(z)=\varphi(z+2\pi/\tau)$ is analytic and bounded in the upper half-plane and on the real axis, and has zeros at $r_\nu e^{i\theta_\nu}=2\pi\nu/\tau$, $\nu=1, 2, \cdots$. We note that $(\sin\theta)/r$ is positive for $re^{i\theta}$ in the upper half-plane. Thus, if g is not to be identically zero, the sum S must also converge, if we sum over any subset of the zeros of g, and ignore multiplicities. But summing over the set of zeros at

$$r_\nu e^{i\theta_\nu}=(2\pi\nu/|\tau|)e^{-i(\arg\tau)},$$

we have

$$\sum_{\nu=1}^{\infty}\frac{\sin(\arg\tau)}{\nu}\frac{|\tau|}{2\pi}=\frac{|\tau|}{2\pi}\sin(\arg\tau)\sum_{\nu=1}^{\infty}\frac{1}{\nu}.$$

But the sum on the right is divergent. Therefore $g(z)$ and, consequently, $\varphi(z)$ must be identically zero in the upper half plane. By a similar argument, we may show that φ is identically zero in the lower half plane. Therefore, the criteria A and B are necessary and sufficient conditions for the continuation of the Fourier coefficients to be the correct one.

It should be added that if one calculates only

$$\lim_{\mathrm{Im}\tau\to 0}[\Phi^{i\tau}(2\pi\nu/\tau)],$$

in some open set of the real τ axis, then one may still infer the desired continuation, $\Phi^{i\tau}(z)$, for τ in the lower half plane. This is because $\Phi^{i\tau}(2\pi\nu/\tau)$ is analytic in the lower-half τ plane for each fixed ν, so that a knowledge of its boundary value on an open set of the real τ axis determines it in the lower half-plane.[5] Since we know that A and B suffice to characterize the correct continuation $\Phi^{i\tau}(z)$ when τ is in the lower half plane,[6] $\Phi^{i\tau}(z)$ is inferred by first continuing the boundary value Fourier coefficients to the lower-half τ plane, and then continuing in z, under the requirement that A and B be satisfied.

[3] If the weight function $A(\omega)$ is known to be zero outside of a bounded region on the real axis, then the function $\Phi(z)$ is analytic at infinity, and A and B clearly determine a unique continuation. In general, however, $A(\omega)$ will not be identically zero in a neighborhood of infinity, and Φ need not be analytic at infinity.

[4] E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, New York, 1939), 2nd ed., p. 131.

[5] This follows from the Schwarz reflection principle. See, e.g., footnote 4.

[6] Note that A and B do not suffice to characterize the correct z continuation of the boundary value function

$$\lim_{\mathrm{Im}\tau\to 0}\Phi^{i\tau}(2\pi\nu/\tau),$$

since, for example, we may add to any continuation the function

$$\Delta(z)=\begin{cases}(1-e^{iz\tau})/(z+i), & z \text{ in upper half-plane}\\ (1-e^{-iz\tau})/(z-i), & z \text{ in lower half-plane,}\end{cases}$$

which for real positive τ satisfies A and B, and is zero when $z=2\pi\nu/\tau$.

Almost Periodicity and the Quantal H Theorem

IAN C. PERCIVAL
University College, London, England*
(Received August 19, 1960)

It is proved that the H theorem for an ensemble of isolated quantal systems with a discrete energy spectrum is false provided the systems satisfy certain broad conditions: the theorem is false for bounded many-particle systems with potential interaction, provided that interaction contains no repulsive singularities stronger than r^{-2} and no attractive singularities stronger than r^{-1}. Ensembles of such systems behave almost periodically, in the sense of H. Bohr. The entropy and the probability of finding an observable in a given range are both almost periodic functions of time.

I. INTRODUCTION

THE H theorem for an ensemble of isolated quantal systems states that the coarse-grained entropy of the ensemble increases to its equilibrium value and stays there.

The H theorem was originally proposed by Boltzmann to provide a theoretical basis for the irreversibility of thermodynamic systems and in particular for the increase of entropy with time. Boltzmann's original statement of the theorem, by which the entropy was defined as a function of the state of an individual system, was contraverted by the well-known[1] objections of Loschmidt and Zermelo. The recurrence objection of Zermelo uses a theorem of Poincaré by which the state of a finite bounded system of particles recurs over and over again—if not exactly, at least to within any arbitrary positive error.

The H theorem was therefore restated by Gibbs in terms of ensembles of systems, the entropy being defined as a property of the ensemble. This form of the theorem was generalized to quantum mechanics by Pauli[2] and by others. We shall be concerned with Pauli's form of the H theorem, which is also that of Tolman.[3]

Pauli worked through a master equation whose derivation was based on the assumption of random phases at all times after some initial instant t_0, but this is untenable except for equilibrium systems. The assumption was removed by van Hove,[4] who required initial random phases only. However, van Hove's derivation depends on an approximation which may be removed only by passing to the limit in which the number of particles becomes infinite, whereas Gibbs' form of the H theorem was supposed to apply to ensembles of systems which contain a finite though large number of particles.

Experimentally we are not able to distinguish each individual quantum state of a macroscopic system, only a large though finite set of states. Gibbs and Pauli define the entropy in terms of the coarse-grained probabilities of finding a system of the ensemble in one of these sets.

It might be supposed that coarse-graining would be sufficient to ensure the irreversibility of ensembles of isolated finite many-particle quantal systems, and that there would be no need to allow the systems to become infinite. We shall show that this supposition is incorrect, and that for reasonable interaction potentials Pauli's form of the quantal H theorem is false when the systems are finite.

A set of conditions for the falsity of the theorem is given at the end of Sec. V. The disproof depends on the theory of almost periodic functions, due largely to H. Bohr.[5] The entropy is shown to be an almost periodic function of time. Except for the trivial case of equilibrium ensembles, it cannot also be nondecreasing, as will be seen from the definition of almost periodicity which follows.

II. ALMOST PERIODICITY

In Schrödinger representation, the state vector of a quantal system in a pure stationary state is a periodic function of time. If the system has a discrete spectrum, then some of the features of periodicity remain, even though the system may be in a nonstationary state or in a mixed state. The features which remain are those of almost periodicity, which is a generalization of pure periodicity.

The amplitude of a single classical harmonic oscillator of unit angular frequency has the form

$$A(t)=A_0\sin(t+\delta_0). \tag{1}$$

Now consider a pair of coupled oscillators, in which the proper vibrations of the pair have angular frequencies 1 and $\sqrt{2}$. Then the amplitude of one of the oscillators has the form

$$A(t)=A_0\sin(t+\delta_0)+A_1\sin(\sqrt{2}t+\delta_1).$$

* This work was largely performed while the author was at Stanford Research Institute, Menlo Park, California, with the support of the National Aeronautics and Space Administration, and partly while at the International Summer School of the University of Grenoble at Les Houches.

[1] D. ter Haar, Revs. Modern Phys. 27, 289 (1955), gives a general review of the history and foundations of the H theorem.

[2] W. Pauli in *Probleme der Modernen Physik*, Sommerfeld Festschrift, edited by P. Debye (Hirzel, Leipzig, 1928), p. 30.

[3] R. C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, New York, 1938).

[4] L. van Hove, Physica 21, 517 (1955); 23, 441 (1957).

[5] H. Bohr, *Collected Mathematical Works*, edited by E. Følner and B. Jessen (Danish Mathematical Society, Copenhagen, 1952), Vols. 2 and 3; see also A. S. Besicovitch, *Almost Periodic Functions* (Dover Publications, New York, 1954), reissue.

Generally this amplitude is not periodic: There is no nonzero value of τ which satisfies the equation

$$A(t)=A(t+\tau) \quad (\text{all } t). \tag{2}$$

But we can establish the existence of values of τ for which this equation is approximately satisfied, and the approximation can be made arbitrarily close by appropriate choice of the values of τ.

Almost periodicity of a function $f(t)$ is defined by the following property: For an arbitrarily small positive error ϵ, the inequality

$$|f(t+\tau)-f(t)|<\epsilon \quad (\text{all } t), \tag{3}$$

is satisfied by infinitely many values of τ, these values being spread over the whole range $-\infty$ to $+\infty$ so as not to leave arbitrarily long empty intervals. This defines what Besicovitch terms a "uniformly almost periodic function (uap)," but since we shall not need to consider any other types of almost periodicity we allow the preceding definition.

Every almost periodic function $f(t)$ has a Fourier series which converges uniformly to the function

$$f(t)=\sum_{n=0}^{\infty} A_n e^{i\omega_n t}, \quad (-\infty<t<+\infty), \tag{4}$$

where the ω_n are real. Conversely, every uniformly convergent Fourier series converges to an almost periodic function. The Fourier transform of such a function is a sequence of delta functions.

III. DENSITY MATRICES

The density matrices which represent mixed states or ensembles of quantal systems span a Hilbert space,[6] which we shall term "density space." In a discrete representation the scalar product of two elements $\mathbf{A}$ and $\mathbf{B}$ of this space is defined as follows:

$$(\mathbf{A},\mathbf{B})=\sum_{mm'} A_{mm'}{}^{*}B_{mm'}=\mathrm{Tr}(\mathbf{A}^{\dagger}\mathbf{B}). \tag{5}$$

This quantity is unchanged under unitary transformations in state vector space. The length $\|\varrho\|=(\varrho,\varrho)^{\frac{1}{2}}$ of a density matrix is never greater than 1, and if the system is isolated it remains constant in time.

Almost periodicity was generalized by Bochner[7] to include functions whose values are elements of a metric space, such as Hilbert space. This definition depends on the length used to specify the "error" which appears in the definition of almost periodicity. All the usual properties of almost periodicity carry over to this generalization. In particular, finite sums and uniformly convergent series of almost periodic functions are themselves almost periodic, and uniformly continuous functions of almost periodic functions of time are themselves almost periodic in time.

Almost periodicity of solutions of the classical wave equation has been treated by Muckenhoupt[8] and Bochner.[9] We have considered almost periodicity of density matrices, using length in density space to define the error.

A density matrix is an almost periodic function of time if for an arbitrarily small positive error ϵ the inequality

$$\|\varrho(t+\tau)-\varrho(t)\|<\epsilon \quad (\text{all } t) \tag{6}$$

is satisfied by infinitely many values of τ, these values being spread over the whole range $-\infty$ to $+\infty$ so as not to leave arbitrarily long empty intervals. Almost periodicity of the elements of a density matrix is not alone sufficient to ensure almost periodicity of the matrix, although the converse is true.

Let $\varrho(t)$ be the density matrix of a system with a discrete set of stationary states, labeled $n=0, 1, 2, \cdots$, with energies E_n, some of which may be equal if there are degeneracies. In energy representation the matrix elements are

$$\begin{aligned}\rho_{nn'}(t)&=\langle n|\varrho(t)|n'\rangle\\ &=\rho_{nn'}(0)\exp[i\hbar^{-1}(E_{n'}-E_n)t],\end{aligned} \tag{7}$$

and each is a periodic function of time. Let $\mathbf{T}_n=n\rangle\langle n$ be the projection operator onto the nth stationary state: then

$$\varrho^{nn'}(t)=\mathbf{T}_n\varrho(t)\mathbf{T}_{n'}, \tag{8}$$

is the matrix which in energy representation has only one nonzero element, equal to $\rho_{nn'}(t)$ and in the location (n,n'). These matrices are orthogonal in density space:

$$(\varrho^{nn'}(t),\varrho^{n''n'''}(t))=\delta_{nn''}\delta_{n'n'''}|\rho_{nn'}(t)|^2, \tag{9}$$

and

$$\begin{aligned}\varrho(t)&=\sum_{n=0}^{\infty}\sum_{n'=0}^{\infty}\varrho^{nn'}(t)\\ &=\sum_{n=0}^{\infty}\sum_{n'=0}^{\infty}\varrho^{nn'}(0)\exp(i\omega_{nn'}t),\end{aligned} \tag{10}$$

where $\omega_{nn'}=(E_{n'}-E_n)/\hbar\cdots\varrho(t)$ is almost periodic if this Fourier series, with coefficients in density space, converges uniformly. At this stage the discreteness of the energy spectrum is essential, for if the spectrum were continuous we should in general have a Fourier integral, and no almost periodicity.

Consider the finite sum

$$\sigma^{NN'}(t)=\sum_{n=0}^{N}\sum_{n'=0}^{N'}\varrho^{nn'}(t) \tag{11}$$

[6] U. Fano, Revs. Modern Phys. **29**, 74 (1957).
[7] S. Bochner, Acta. Math. **61**, 149 (1933).
[8] C. F. Muckenhoupt, J. Math. and Phys. **8**, 163 (1929).
[9] S. Bochner, Acta Math. **62**, 227 (1934).

as an approximation to $\varrho(t)$. The square of the error is

$$\|\varrho(t)-\sigma^{NN'}(t)\|^2=\|\sum_{n=N+1}^{\infty}\sum_{n'=N'+1}^{\infty}\varrho^{nn'}(t)\|^2$$

$$=\sum_{n=N+1}^{\infty}\sum_{n'=N'+1}^{\infty}\|\varrho^{nn'}(t)\|^2$$

$$=\sum_{n=N+1}^{\infty}\sum_{n'=N'+1}^{\infty}|\rho_{nn'}(0)|^2. \quad (12)$$

The second equality follows from the orthogonality of the $\varrho^{nn'}(t)$. Since the error is independent of time, $\sigma^{NN'}(t)$ converges uniformly to $\varrho(t)$, and $\varrho(t)$ is almost periodic.

IV. PROBABILITY AND ENTROPY

The probability of finding a system of the ensemble in a typical state m of a complete discrete set of states is

$$p_m(t)=\mathrm{Tr}\,[\varrho(t)\mathbf{T}_m], \quad (13)$$

where $\mathbf{T}_m=m\rangle\langle m$ projects onto the state m. $\mathrm{Tr}\,[\varrho\mathbf{T}_m]$ is a uniformly continuous function of ϱ, so $p_m(t)$ is an almost periodic function of t.

For systems of many particles it is not possible to distinguish the individual states m, but only to determine whether the state of the system lies in some range R_M containing a finite number ν_M of states. We shall suppose that the ranges R_M do not overlap and together include all states m. This simplified representation of experimental error is analogous to coarse-graining in classical mechanics. The probability of finding the system in the range R_M is

$$P_M(t)=\sum_{m\in R_M}p_m(t), \quad (14)$$

and

$$\sum_{M=0}^{\infty}P_M(t)=1. \quad (15)$$

The entropy $S(t)$ is defined in terms of these sets of states by the equation

$$S(t)=-k\sum_M P_M(t)\ln[\nu_M^{-1}P_M(t)], \quad (16)$$

where k is Boltzmann's constant.

$f(x)=x\ln x$ is a uniformly continuous function of x when $0\leqslant x\leqslant 1$; consequently each of the terms in (16) is almost periodic in time, and if the sum were finite, then $S(t)$ would also be almost periodic.

If it is assumed that the same finite region of phase space makes a dominant contribution to the entropy at all times, and that this implies that the sum in (16) is effectively finite, the proof of the almost periodicity of the entropy is already complete. The next section provides a fairly detailed proof, and the conditions necessary to carry it through.

V. UNIFORM CONVERGENCE

The mean energy $E=\langle\mathbf{H}\rangle=\mathrm{Tr}(\varrho\mathbf{H})$ of the ensemble is conserved. Suppose for the present that we could measure m sufficiently accurately to distinguish individual pure states. Each state m has mean energy

$$E_{(m)}=\langle m|\mathbf{H}|m\rangle, \quad (17)$$

and measurements of the variable m enable us to measure the mean energy of the ensemble:

$$E_{\mathrm{est}}=\sum_m p_m E_{(m)}$$
$$=\sum_m \rho_{mm}H_{mm}$$
$$=\langle\mathbf{H}\rangle-\sum_m\sum_{m'\neq m}\rho_{mm'}H_{m'm}. \quad (18)$$

The error in this estimate is frequently assumed to be negligible. We impose the weaker condition that it should be bounded:

$$E-E'\leqslant E_{\mathrm{est}}\leqslant E+E'. \quad (19)$$

Now return to the coarse measurements of ranges R_M. Let

$$E_{(M)}=\nu_M^{-1}\sum_{m\in R_M}E_{(m)}. \quad (20)$$

We suppose that the error in

$$E_{\mathrm{Est}}=\sum_M P_M E_{(M)}, \quad (21)$$

the energy estimated from the coarse measurements, is bounded; then

$$E-E''<E_{\mathrm{Est}}<E+E''. \quad (22)$$

Generally, $E_{\mathrm{Est}}\neq E_{\mathrm{est}}$.

Let the ranges M be labeled in order of increasing energy $E_{(M)}$, so that $E_{(M_1)}\leqslant E_{(M_2)}$ when $M_1<M_2$. We seek an upper bound to the contribution

$$S_{M_0}=-k\sum_{M=M_0}^{\infty}P_M\ln(\nu_M^{-1}P_M) \quad (23)$$

to the entropy from states $M\geqslant M_0$, where M_0 is large and the P_M are subject to the energy condition (22). Such a bound may be obtained by putting all systems with $M<M_0$ in the state $m=0$ of the lowest energy $E_{(m=0)}$ and allowing E_{Est} to take on its maximum possible value, so that

$$\sum_{M=M_0}^{\infty}P_M E_{(M)}=E+E''. \quad (24)$$

The energy scale has been chosen with $E_{(m=0)}=0$. Since for sufficiently large M_0, $E_{(M_0)}>E+E'$, the equality (24) ensures that

$$\sum_{M=M_0}^{\infty}P_M<1.$$

S_{M_0} may be maximized by Lagrange's method, whereby

$$P_M = \nu_M \exp[-(1+\alpha E_{(M)})], \quad (M \geqslant M_0) \tag{25}$$

and α is determined by the equation

$$\sum_{M=M_0}^{\infty} \nu_M E_{(M)} \exp[-(1+\alpha E_{(M)})] = E + E''. \tag{26}$$

Therefore for any ensemble whose mean energy is E,

$$\begin{aligned} S_{M_0} &\leqslant \sum_{M=M_0}^{\infty} \nu_M (1+\alpha E_{(M)}) \exp[-(1+\alpha E_{(M)})] \\ &\leqslant \sum_{M=M_0}^{\infty} \nu_M (E_{(M_0)}{}^{-1} E_{(M)} \\ &\qquad + \alpha E_{(M)}) \exp[-(1+\alpha E_{(M)})] \\ &= (E_{(M_0)}{}^{-1} + \alpha)(E + E''). \end{aligned} \tag{27}$$

Assuming there is no finite energy range containing an infinite number of $E_{(M)}$,

$$\lim_{M_0 \to \infty} E_{(M_0)}{}^{-1} = 0.$$

Also if

$$\sum_{M=M_0}^{\infty} \nu_M E_{(M)} \exp[-(1+\alpha_0 E_{(M)})] \tag{28}$$

converges for all fixed $\alpha_0 > 0$, however small, then

$$\lim_{M_0 \to \infty} \sum_{M=M_0}^{\infty} \nu_M E_{(M)} \exp[-(1+\alpha_0 E_{(M)})] = 0, \tag{29}$$

and α must be less than α_0 to satisfy Eq. (26) if M_0 is sufficiently large. Therefore

$$\lim_{M_0 \to \infty} \alpha = 0 \quad \text{and} \quad \lim_{M_0 \to \infty} S_{M_0} = 0. \tag{30}$$

Therefore, we can approximate as close as we please to the entropy of all ensembles with fixed mean energy E by replacing the infinite sum (16) by a finite sum over the *same* values of M for every such ensemble. By the conservation of energy an ensemble has the same E at all times, so the infinite sum (16) converges uniformly to $S(t)$, and $S(t)$ is almost periodic.

The assumptions made in the proof were

(A1) Each system of the ensemble is isolated.

(A2) The energy spectrum E_n is discrete.

(A3) The energy "spectrum" $E_{(m)} = H_{mm}$ of the complete set of states used to define the entropy is discrete, has a minimum energy level and no limit points (points of accumulation).

(A4) The density of the "measured energy spectrum" $E_{(M)}$ associated with ranges R_M each containing a finite number ν_M of states m does not increase so rapidly with M that

$$\sum_{M=0}^{\infty} \nu_M E_{(M)} \exp(-\alpha_0 E_{(M)})$$

diverges for any $\alpha_0 > 0$.

(A5) The "measured mean energy" E_{Est} is never greater than some maximum value $E = E''$.

VI. SYSTEMS OF INTERACTING PARTICLES

It is conventional to use the momenta of the interacting particles for the measured coordinate m. More generally m labels the eigenfunctions of a complete set of commuting operators which themselves commute with the Hamiltonian $\mathbf{H}_0$ which the particles would have in the absence of interaction between them. By the usual choice of zero of energy, $\mathbf{H}_0$ is a positive definite operator. Typically

$$\mathbf{H}_0 = \tfrac{1}{2} \textstyle\sum_i \mathbf{p}_i^2/\mu_i, \tag{31}$$

where $\mathbf{p}_i$ is the momentum and μ_i the mass of particle i. Suppose the interaction $\mathbf{V}$ can be represented by a potential function $V(X)$, where X represents the coordinates of all the particles. Then the total Hamiltonian $\mathbf{H}$ is

$$\mathbf{H} = \mathbf{H}_0 + \mathbf{V}. \tag{32}$$

A system of a finite number of independent particles contained in a box of finite volume with perfectly reflecting or periodic boundary conditions has a discrete energy spectrum. When the particles interact with any reasonable potential $V(X)$, the discreteness of the spectrum is retained. A petit ensemble of such systems then satisfies conditions (A1)–(A3).

Generally the ranges R_M are taken to be small regions in the space of the momenta of all the particles. Typically no momentum varies by more than a constant q within each R_M, so that the energy varies by no more than

$$\begin{aligned} \Delta E_{(M)} &= \textstyle\sum_i \mathbf{p}_i \cdot \Delta \mathbf{p}_i/\mu_i \\ &\leqslant \textstyle\sum_i |\mathbf{p}_i| q/\mu_i. \end{aligned} \tag{33}$$

For sufficiently large M, $\Delta E_{(M)}$ is negligible in comparison with $E_{(M)}$ so that for (A4) and (A5), $E_{(M)}$ may be replaced by $E_{(m)}$, ν_M by 1 and E_{Est} by E_{est}. Assumption (A4) is then satisfied by the spectrum $E_{(m)}$.

For (A5) we require to know

$$\begin{aligned} E_{\text{est}} - E &= \sum_m \sum_{m' \neq m} \rho_{mm'} H_{m'm} \\ &= \sum_m \sum_{m' \neq m} \rho_{mm'} V_{m'm} \\ &= \operatorname{Tr}(\varrho \mathbf{V}) - \sum_m \rho_{mm} V_{mm}. \end{aligned} \tag{34}$$

If the interaction has no singularities stronger than r^{-2}, where r is an interparticle distance, then V_{mm} is bounded

and the second term in (34) is bounded. We have assumed that the region which contains the system is finite in volume.

If the potential energy of interaction is always greater than some constant V_0, then

$$E \geqslant \langle \mathbf{V} \rangle = \mathrm{Tr}\,(\varrho \mathbf{V}) \geqslant V_0 \tag{35}$$

since the kinetic energy $\langle \mathbf{H}_0 \rangle$ is positive. Therefore Tr $(\varrho \mathbf{V})$ is bounded.

It remains to be shown that Tr$(\varrho \mathbf{V})$ is bounded in the presence of attractive Coulomb singularities. We give a nonrigorous demonstration. Suppose it were unbounded for the case of the interaction of two oppositely charged particles. Then for the magnitude of Tr $(\varrho \mathbf{V})$ to be very large, part of

$$\rho(\mathbf{r}) = \sum_{mm'} \langle \mathbf{r} | m \rangle \rho_{mm'} \langle m' | \mathbf{r} \rangle \tag{36}$$

must be concentrated in a very small volume Ω, say a sphere of radius R. For approximately uniform distribution in the sphere

$$\int_\Omega \rho(\mathbf{r}) d^3r \leqslant 1; \quad \rho(\mathbf{r}) \leqslant 3/(4\pi R^3). \tag{37}$$

If the charges are eZ_1 and $-eZ_2$, with positive Z_1 and Z_2, then

$$\begin{aligned} |\langle V \rangle| &\leqslant (3e^2/4\pi R^3) Z_1 Z_2 \int_\Omega d^3r/r \\ &= 3e^2 Z_1 Z_2/(2R) \\ &= \mathcal{O}(R^{-1}) \quad (\text{small } R). \end{aligned} \tag{38}$$

By the uncertainty relation the relative momentum is

$$\langle |\mathbf{p}| \rangle = \hbar \mathcal{O}(R^{-1}) \tag{39}$$

and the kinetic energy of relative motion is

$$\langle \mathbf{H}_0' \rangle = (\hbar^2/2\mu) \mathcal{O}(R^{-2}). \tag{40}$$

Therefore if the magnitude of the potential energy becomes very large, and the potential energy negative, the kinetic energy of relative motion becomes even larger and positive. For $\langle \mathbf{V} \rangle$ to be unbounded $\langle \mathbf{H} \rangle = \langle \mathbf{H}_0' \rangle + \langle \mathbf{V} \rangle$ would have to be unbounded, and this contradicts our initial assumptions. Any redistribution of $\rho(\mathbf{r})$ within Ω merely increases the kinetic energy even more than it decreases the potential energy. The argument is unaffected if the number of particles becomes large, so long as it remains finite. Thus $\langle \mathbf{V} \rangle$ is bounded.

Therefore $S(t)$ is an almost periodic function of time when the particles interact through potentials which have repulsive singularities no stronger than r^{-2} and attractive singularities no stronger than r^{-1}, and the H theorem is then false.

VII. DISCUSSION

Isolated finite quantal ensembles are essentially almost periodic, and not irreversible; this almost periodicity is not removed by coarse-graining. There is a reccurence for ensembles of quantal systems similar to Poincaré's recurrence for individual classical systems. For example, an ensemble of quantal systems each consisting of N identical particles all of which are in one half of an infinitely heavy perfectly reflecting cylinder at time zero, will be found after some long interval of time T in a condition in which all the particles are in the same half of the cylinder for nearly every member of the ensemble, and this *same* time T will be sufficient however many systems there may be in the ensemble. For systems of macroscopic size this time will be very long indeed, and much longer than the time of an ordinary experiment. However, we have shown that ensembles of finite isolated quantal systems can tend to equilibrium only over a finite length of time, and no theory of their irreversibility can be entirely satisfactory if it does not take this into account.

For sufficiently small systems it might be possible to detect the periods experimentally. They are essentially quantal periods.

The disproof of the H theorem breaks down for systems with continuous spectra in general, and for classical systems in particular, unless the latter are linear and finite. There are some classical ensembles at least which approach equilibrium and then stay there. For instance consider an ensemble in which each member system is a single anharmonic oscillator. At time $t=0$ let the system have a Bolzmann energy distribution, but a nonequilibrium distribution in the two-dimensional phase space. For instance in the case of simple pendulums with finite amplitude they could all be vertical at time $t=0$ with a Gaussian velocity distribution. The representative point in phase space of each system then cycles around a suitably chosen origin with a period depending on its energy. This is mathematically equivalent to Gibbs' model of a cylinder of liquid.[10] The coarse-grained entropy increases to its equilibrium value, and stays there. For the quantal analog of the foregoing classical ensemble, the entropy is almost periodic, whatever the initial distribution. Thus there is a basic difference between the classical and quantal statistical mechanics of isolated finite systems.

Although we are unable to come to any general conclusions regarding the classical H theorem, it can be stated that any "proof" of the theorem which can be generalized to quantal statistics is necessarily invalid for finite systems and arbitrarily long times.

ACKNOWLEDGMENT

The author would like to thank L. van Hove for stimulating discussion.

[10] J. S. Gibbs, *Elementary Principles in Statistical Mechanics* (Charles Scribner's Sons, New York, 1902), Chap. XII.

On Analytic Continuation, Multiplication, and Fourier Transformations of Schwartz Distributions

H. J. BREMERMANN* AND L. DURAND, III
The Institute for Advanced Study, Princeton, New Jersey
(Received February 9, 1960)

It is shown that a class of Schwartz distributions on the real axis can be continued to holomorphic functions in the upper and lower complex half-planes such that the "jump" on the real axis represents the distribution. Many operations with distributions can be reduced to operations with the associated holomorphic functions which is of particular interest for the convolution product and for Fourier transforms. By means of the continuations several kinds of multiplications for distributions are being defined, which is of interest for quantum field theory.

I. INTRODUCTION

THE theory of distributions of Schwartz[1] generalizes the notion of (ordinary) "function" such as to include measures and other "generalized functions." The notion of "distribution" comprises various quantities occurring in physics that are not functions in the ordinary sense; for example, Dirac's δ function, causal functions, and the vacuum expectation values in quantum field theory. Schwartz's theory serves well for the problems of physics except in the case where products of these quantities occur (for instance products of causal functions) for which it does not account. (Compare Akhiezer and Berestesky[2] and Bogoliubov and Shirkov.[3])

Schwartz defines a distribution space as the dual of a linear space of "test functions," such as the space ($\mathcal{E}$) of all (C^∞) functions or the space ($\mathfrak{D}$) of all (C^∞) functions with compact support.

Another approach (compare Lighthill[4] and the bibliography there) is to define distributions as the limit quantities of sequences of functions [$\delta(x)$, for instance, is represented by the approximating functions $(n/\pi)^{\frac{1}{2}} \exp(-nx^2)$].

In this paper we suggest a third approach: We associate with a distribution on the real axis a pair of holomorphic functions in the complex plane, one function holomorphic in the (open) upper half-plane, the other holomorphic in the lower half-plane. The limit of the sum of these two functions at $x+i\epsilon$ and $x-i\epsilon$, $\epsilon \to 0$, x on the real axis, represents the distribution. Every distribution with compact support (and certain others) can be represented in this way. (We denote the associated holomorphic functions as "analytic continuation" of the distribution.)

This technique has an advantage: operations with distributions can in many cases be replaced by operations with concrete analytic functions. For example, a distribution T (defined on the real axis) applied to a test function ϕ can be written only symbolically in the form of an integral

$$T \cdot \phi = \int_{-\infty}^{+\infty} T(x)\phi(x)dx,$$

where $T(x)$ has no independent meaning. By using the analytic continuation [denoted by $\mathbf{T}^1(z)$] one can replace the symbolic integral $\int_{-\infty}^{+\infty} T(x)\phi(x)dx$ by the ordinary contour integral $\int_{C_p} \mathbf{T}^1(z)\phi(z)dz$ [for holomorphic test functions $\phi(z)$, C_p consisting of a line above and a line below the real axis].

The analytic continuations of several distributions of interest in physics such as Dirac's δ function and its derivatives, the δ_+ function, and $P(1/x)$ (principal part of $1/x$), are given as examples and various commonly used identities connecting these distributions are easily established rigorously. The idea to continue functions and distributions to analytic functions is an old one. A beautiful exposition (for functions only) is given in Carleman.[5] Schwartz[6] continues Laplace transforms of distributions. Similarly the n-fold vacuum expectation values studied by Wightman[7,8] (Wightman functions) are Fourier transforms of quantities vanishing outside the light cone which implies that the Fourier transform converges not only on real space-time but in the "forward tube" as well. It defines a particular type of analytic continuation of a distribution. Similar techniques are used in dispersion relations (compare Bogoliubov, Medvedev, and Polivanov[9] and Bremermann, Oehme, and Taylor.[10]

* The work was finished at the University of California, Berkeley with support of the Office of Naval Research.

[1] L. Schwartz, Théorie des distributions, Vol. 1 and 2. Actualités sci. et ind. No. 1091 (1950) and 1122 (1951).

[2] A. I. Akhiezer and V. B. Berestetsky, *Kvantovaya Elektrodinamika* (Moscow, 1953) (English translation: U. S. Atomic Energy Commission).

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

[4] M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions* (Cambridge University Press, New York, 1958).

[5] T. Carleman, *L'integrale de Fourier et questions qui s'y rattachent* (Almqvist & Wiksells Boktryckeri-A.-B., Uppsala, 1944), Chap. II.

[6] L. Schwartz, Medd. Lunds Univ. Mat. Sem. Suppl. M. Riesz 196 (1952).

[7] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

[8] A. S. Wightman and D. Hall, Kgl. Danske Videnskab. Selskab Mat.-fys. Medd. **31**, No. 5 (1957).

[9] N. N. Bogoliubov, B. V. Medvedev, and M. K. Polivanov, *Problems in the Theory of Dispersion Relations* (Fizmatgiz, Moscow, 1958).

[10] H. J. Bremermann, R. Oehme, and J. G. Taylor, Phys. Rev. **109**, 2178 (1958).

Also, Köthe[11] observed in 1952 that, given a finite closed curve C in the complex plane and a distribution on the curve, a pair of holomorphic functions can be associated with the distribution, one function holomorphic in the interior of the domain bounded by the curve, one holomorphic in the exterior. Tillmann[12] generalized this theory to unbounded domains (such as the half plane). Similar results have also been obtained by Sato.[13]

Having applications to physics in mind, we limit ourselves to distributions on the real axis and use straightforward methods rather than the elegant language and techniques of Banach spaces extensively employed by Köthe, Tillmann, and Sato.

The product of two arbitrary distributions is in general undefined in the theory of Schwartz, the symbolic integral $\int_{-\infty}^{+\infty} S(x)T(x)\phi(x)dx$ having a well-defined meaning only if in the neighborhood of each singularity of $S(x)$, $T(x)$ is "smooth" of at least the order of the singularity of $S(x)$. By using the analytic continuation of a distribution, various ways to define a multiplication for arbitrary distributions suggest themselves. These possibilities appear to have applications relative to the divergence difficulties in quantum field theory arising from multiplication of singular distributions (compare Bogoliubov and Shirkov[3] and Bogoliubov and Parasiuk[14]). These applications are to be discussed in a forthcoming paper by Bremermann,[15] together with applications of results on Fourier transforms obtained here.

Following Carleman,[5] we define the Fourier transform as follows: We split the integration into $\int_0^\infty$ and $\int_{-\infty}^0$. The first integral gives a function holomorphic in the upper half-plane, the second a function holomorphic in the lower half-plane. An important result is: If f is a square integrable function, then this pair of holomorphic functions coincides with the analytic continuation of the ordinary Fourier transform of f (and this result can be extended to include tempered distributions). Various results are established and the Fourier transforms of such functions as $\delta_+(x)$, $P(x^{-n})$, $\epsilon(x)$ are computed as examples. Carleman's work, which was done before Schwartz's theory was developed, is limited to L^2 and L^p functions. The methods become more powerful when applied to distributions. Our results of Sec. IX are closely related to Schwartz.[6]

This paper is limited to the case of distributions on the real axis. An extension to higher dimensions (involving functions of several complex variables) is possible, but some new difficulties arise. The authors believe that the method of analytic continuation holds great potential for applications, some of which (to differential equations and in particular to electric networks) will be discussed in a future paper.

[11] G. Köthe, Math. Z. **57**, 13 (1952).
[12] H. G. Tillmann, Math. Z. **59**, 61 (1953).
[13] M. Sato, Proc. Japan Acad. **35**, 126 (1958).
[14] N. N. Bogoliubov and O. S. Parasiuk, Acta Math. **97**, 227 (1957).
[15] H. J. Bremermann, ONR Rept. No. 8 (1959).

II. SCHWARTZ DISTRIBUTIONS

In the following we summarize some of the basic definitions of Schwartz's "Théorie des distributions".[1]

Distributions are defined as linear functionals operating on a space of "test functions." Different spaces of test functions give rise to different distribution spaces.

In any case the test functions are complex valued functions $\phi(x)=\phi(x_1,\cdots x_n)$ of n real variables that are m times continuously differentiable ["of class (C^m)"], with $1\leq n<\infty$ and $0\leq m\leq\infty$. In most cases the test functions are taken to be indefinitely differentiable [(C^∞) functions].

The variables $x_1, \cdots x_n$ range over the n-dimensional Euclidean space R^n. The test functions $\phi(x)$ are defined on all of R^n. The complement of the largest open set where $\phi(x)$ is zero is called the "support of $\phi(x)$."

Schwartz considers also the case in which the $\phi(x)$ are functions on a differentiable manifold. In this paper we will limit ourselves to the R^n.

Space of Test Functions $(\mathfrak{D})$

$(\mathfrak{D})$ is the vector space of all (C^∞) functions on R^n that have compact support. Convergence is defined as follows: A sequence $\{\phi_j\}$ is said to converge to zero if all the ϕ_j have their support contained in a fixed compact subset of R^n and if the ϕ_j as well as all their derivatives converge uniformly to zero. (Uniform convergence is required for each fixed order of the derivatives, not for all orders collectively.)

Space of Distributions $(\mathfrak{D}')$

A functional T on $(\mathfrak{D})$ is an operation that associates with every $\phi\in(\mathfrak{D})$ a complex number. We denote this associated number by $T\cdot\phi$.

A functional T on $(\mathfrak{D})$ is *linear* if:

(a) $T\cdot(\phi_1+\phi_2)=T\cdot\phi_1+T\cdot\phi_2$ for every $\phi_1, \phi_2\in(\mathfrak{D})$;

(b) $T\cdot(k\phi)=kT\cdot\phi$ for every $\phi\in(\mathfrak{D})$ and every complex number k.

A functional T is *continuous* if $T\cdot\phi_j$ converges to zero for any sequence $\{\phi_j\}$ of functions $\phi_j\in(\mathfrak{D})$ that converges to zero in $(\mathfrak{D})$ (as defined earlier).

A distribution, as defined by Schwartz, *is a continuous linear functional on* $(\mathfrak{D})$. *The space of all distributions is denoted by* $(\mathfrak{D}')$ *and is the dual space of* $(\mathfrak{D})$.

Support of a Distribution

A distribution T is said to be *zero* in an open set Ω of R^n if $T\cdot\phi=0$ for all test functions $\phi\in(\mathfrak{D})$ whose support is contained in Ω.

The *support* of a distribution is the complement of the largest open set in which T is zero.

Space of Distributions ($\mathcal{E}'$)

Let $(\mathcal{E})$ be the vector space of all (C^∞) functions on R^n with arbitrary support. Convergence in $(\mathcal{E})$ is defined differently from convergence in $(\mathfrak{D})$, as follows: A sequence $\{\phi_j\}$, $\phi_j \in (\mathcal{E})$, converges to zero if the ϕ_j converge to zero uniformly in every compact subset of R^n, and the same is true for the derivatives of any order. (Again uniform convergence is required only for fixed order of the derivatives, not for all orders collectively.)

This notion of convergence is weaker than the convergence defined for $(\mathfrak{D})$ and it generates a true topology as follows: Let ϕ_0 be a fixed function in $(\mathcal{E})$. Then we define as neighborhoods the sets $\{\phi \mid |\phi^{(p)}(x) - \phi_0^{(p)}(x)| < \epsilon$ for $x \in K$, $p \leq m\}$, where $\epsilon > 0$ and K is a compact set and m an integer >0 ($\phi^{(p)}$ denotes the pth derivative). It is easy to see that this definition generates a topology, and a sequence is convergent in this topology if, and only if, it is convergent as just defined.

The dual space of $(\mathcal{E})$, that is, the space of continuous linear functionals on $(\mathcal{E})$ is denoted by $(\mathcal{E}')$. Schwartz has shown: $(\mathcal{E}')$ *consists exactly of those distributions in* $(\mathfrak{D}')$ *that have compact support.*

Generalized Function Notion

It is sometimes convenient to write a distribution as an integral over a "generalized function":

$$T \cdot \phi = \int_{-\infty}^{\infty} T(x)\phi(x)dx,$$

where the integral and the "symbolic kernel," or "generalized function" $T(x)$ are defined by this equation. The Dirac δ function and the causal functions in quantum field theory are usually written in this form. On the other hand, if $T(x)$ is given as a summable function, then the integral $\int_{-\infty}^{+\infty} T(x)\phi(x)dx$ defines a distribution, which, again, we denote by $T \cdot \phi$.

Principal Part of x^{-n}

The function x^{-1} is not integrable at the origin and therefore does not define a distribution in $(\mathfrak{D}')$. However, it is possible to define a "principal part of x^{-1}," denoted by $P(x^{-1})$ as follows:

$$P\int_{-\infty}^{+\infty} \frac{1}{x}\phi(x)dx = \lim_{\epsilon \to 0} \tfrac{1}{2}\int_{-\infty}^{+\infty} \phi(x)\left(\frac{1}{x+i\epsilon} + \frac{1}{x-i\epsilon}\right)dx$$

for all $\phi(x) \in (\mathfrak{D})$.

To show that this defines a distribution in $(\mathfrak{D}')$ let us first observe that due to the fact that any $\phi \in (\mathfrak{D})$ has compact support, we can limit the integration from $-\infty$ to $+\infty$ to integration over any finite interval $\{a,b\}$ that contains the support of ϕ. Let us choose $a,b \neq 0$.

Since ϕ is a C^∞ function we can develop it into a finite Taylor series

$$\phi(x) = \phi(0) + \phi'(0)x + [\phi''(\vartheta x)/2!]x^2, \quad 0 < \vartheta < 1.$$

We see that

$$[\phi(x) - \phi(0)]/x$$

is a continuous (even C^∞) function. Hence

$$\lim_{\epsilon \to 0} \tfrac{1}{2}\int_a^b [\phi(x) - \phi(0)]\left(\frac{1}{x+i\epsilon} + \frac{1}{x-i\epsilon}\right)dx = \int_a^b \frac{\phi(x) - \phi(0)}{x}dx$$

exists as an ordinary integral. Thus only

$$P\int_a^b \frac{\phi(0)}{x}dx$$

remains to be determined. We obtain

$$\lim_{\epsilon \to 0} \phi(0)\tfrac{1}{2}\int_a^b \left(\frac{1}{x+i\epsilon} + \frac{1}{x-i\epsilon}\right)dx$$

$$= \lim_{\epsilon \to 0} \frac{\phi(0)}{2}[\log(b+i\epsilon) + \log(b-i\epsilon) - \log(a+i\epsilon) - \log(a-i\epsilon)]$$

$$= \frac{\phi(0)}{2}\log\frac{b^2}{a^2}.$$

Hence $P\int_a^b (1/x)\phi(x)dx$ is well defined for every $\phi \in (\mathfrak{D})$, and the integral depends linearly and continuously upon ϕ, thus $P(1/x)$ defines a distribution in $(\mathfrak{D}')$.

The integral $P\int_{-\infty}^{+\infty} (1/x)\phi(x)dx$ converges even if $\phi(x)$ vanishes only linearly at infinity (rather than having compact support). This can be seen by splitting $\phi(x) = \phi_1(x) + \phi_2(x)$, where ϕ_1 and ϕ_2 are both (C^∞) functions, ϕ_1 with compact support and ϕ_2 vanishing in a neighborhood of the origin. In particular we get

$$\frac{1}{2\pi i}P\int_{-\infty}^{+\infty} \frac{1}{x}\frac{1}{x-z}dx = \begin{cases} 1/2z & \text{for } \mathrm{Im}z > 0, \\ -(1/2z) & \text{for } \mathrm{Im}z < 0. \end{cases}$$

This follows from the fact that for $\epsilon > 0$

$$\frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{1}{x+i\epsilon}\frac{1}{x-z}dx = \begin{cases} 1/(z+i\epsilon) & \text{for } \mathrm{Im}z > 0, \\ 0 & \text{for } \mathrm{Im}z < 0; \end{cases}$$

and

$$\frac{1}{2\pi i}\int_{-\infty}^{+\infty} \frac{1}{x-i\epsilon}\frac{1}{x-z}dx = \begin{cases} 0 & \text{for } \mathrm{Im}z > 0, \\ -[1/(z-i\epsilon)] & \text{for } \mathrm{Im}z < 0, \end{cases}$$

(Cauchy's integral formula).

Completely analogously a principal part can be defined for x^{-n}, n an integer and $n>0$ by

$$P\int_{-\infty}^{+\infty}\frac{1}{x^n}\phi(x)dx = \lim_{\epsilon\to 0}\tfrac{1}{2}\int_{-\infty}^{+\infty}\left[\frac{1}{(x+i\epsilon)^n}+\frac{1}{(x-i\epsilon)^n}\right]\phi(x)dx.$$

In particular, we get

$$\frac{1}{2\pi i}P\int_{-\infty}^{+\infty}\frac{1}{x^n}\frac{1}{x-z}dx=\begin{cases}1/(2zn) & \text{for } \mathrm{Im}z>0,\\ -[1/(2zn)] & \text{for } \mathrm{Im}z<0.\end{cases}$$

III. ANALYTIC CONTINUATION OF DISTRIBUTIONS

The Cauchy integral has the "reproducing property"

$$\frac{1}{2\pi i}\int_{\partial D}f(\zeta)\frac{1}{\zeta-z}d\xi=\begin{cases}f(z) & \text{for } z\in D,\\ 0 & \text{for } z\notin D,\end{cases}$$

D a domain in the complex z plane, $f(z)$ holomorphic in D, continuous in $\bar{D}$. If under the integral we substitute for the holomorphic function f an arbitrary function g, then

$$F(z)=(2\pi i)^{-1}\int_{\partial D}g(\zeta)\frac{1}{\zeta-z}d\zeta$$

is still a holomorphic function in D, but in general $F(z)\not\equiv g$, that is, the integral no longer reproduces g. In the following we will study the relationship between $F(z)$ and g, not only if g is an arbitrary function, but if g is a distribution. As ∂D we will take the real axis, hence D is the upper or lower half-plane.

Theorem 1. *Let T be a distribution with compact support. Then*

$$\mathbf{T}^0(z)\equiv(1/2\pi i)T\cdot(x-z)^{-1}$$

exists and is a holomorphic function of z in the whole z plane minus the support of T. For $z\to\infty$, $T(z)$ tends to zero.

We will call $\mathbf{T}^0(z)$ the "Cauchy intgral of T."

Proof. (This result can also be derived from Schwartz,[1] Vol. 2, theorem 11.) For $\mathrm{Im}z\neq 0$ the function $(x-z)^{-1}$ is a (C^∞) function with respect to x, and is thus a function in $(\mathcal{S})$. Hence, $T\cdot(x-z)^{-1}$ exists for any z with $\mathrm{Im}z\neq 0$.

To show that it exists also for z in the complement of the support of T we make use of a remark by Schwartz (Vol. I, Chap. III, Sec. 7):

Let $\alpha(x)$ be a (C^∞) function such that $\alpha(x)=1$ for x in the support of T. Then

$$T\cdot(\alpha\phi)=T\cdot\phi$$

for every $\phi\in(\mathcal{S})$.

Let N be a neighborhood of the support of T, $C(N)$ the complement of N on the x axis. Let $\alpha(x)$ be a (C^∞) function such that

$$\alpha(x)=0 \quad \text{for } x\in C(N)$$
$$\alpha(x)=1 \quad \text{for } x \text{ in support of } T.$$

[Such $\alpha(x)$ obviously exist.] Then

$$\alpha(x)(x-z)^{-1}$$

is a (C^∞) function for every $z\notin N$. Hence

$$T\cdot(x-z)^{-1}=T\cdot[\alpha(x)\cdot(x-z)^{-1}]$$

exists for every $z\notin N$.

Secondly, we have to show that $\mathbf{T}^0(z)$ is holomorphic. We have

$$\frac{\partial}{\partial z}\mathbf{T}^0(z)=\lim_{h\to 0}h^{-1}[\mathbf{T}^0(z+h)-\mathbf{T}^0(z)]$$
$$=\lim_{h\to 0}\frac{1}{2\pi i}\frac{1}{h}T\cdot\alpha(x)\left[\frac{1}{x-z-h}-\frac{1}{x-z}\right]$$
$$=\lim_{h\to 0}\frac{1}{2\pi i}T\cdot\frac{\alpha(x)}{(x-z-h)(x-z)}.$$

If $z\notin N$, then the function

$$\phi_h=\alpha(x)/(x-z-h)(x-z)$$

converges for $h\to 0$ uniformly on the whole x axis to $\alpha(x)(x-z)^{-2}$. Similarly, $(d/dx)\phi_h$ converges uniformly to $(d/dx)\alpha(x)(x-z)^{-2}$, and the higher derivatives converge correspondingly.

Therefore, $T\cdot\phi_h$ converges to $T\cdot\alpha(x)(x-z)^{-2}=T\cdot(x-z)^{-2}$. This shows that the complex derivative of $\mathbf{T}^0(z)$ exists for every $z\notin N$, hence that $\mathbf{T}^0(z)$ is holomorphic in the whole z plane minus the real axis. But since N was an arbitrary neighborhood of the support of T, $\mathbf{T}^0(z)$ exists and is holomorphic in the whole z plane minus the support of T.

For $z\to\infty$ the function $\alpha(x)(x-z)^{-1}$ tends to zero uniformly, together with all its derivatives. Consequently, because of continuity, $\mathbf{T}^0(z)$ tends to zero for $z\to\infty$. This completes the proof.

Corollary. *If $T\in(\mathcal{S}')$, then $\mathbf{T}^0(z)$ vanishes at least as $A|z|^{-1}$ for $|z|\to\infty$.*

Indeed, $\mathbf{T}^0(z)$ is holomorphic outside a sufficiently large circle, hence has a Laurent development:

$$\mathbf{T}^0(z)=a_0+a_1\frac{1}{z}+a_2\frac{1}{z^2}+\cdots.$$

But from theorem 1, limit as $z\to\infty$ $T(z)=0$. This implies that $a_0=0$, and the corollary follows immediately.

Theorem 2. *The nth derivative of the Cauchy integral of a distribution $T\in(\mathcal{S}')$ is equal to the Cauchy integral of the nth derivative of T:*

$$\frac{\partial^n}{\partial z^n}\mathbf{T}^0(z)=\mathbf{T}^{0(n)}(z)=\frac{n!}{2\pi i}T\cdot(x-z)^{-n-1}.$$

Here $T^{(n)}$ denotes the nth derivative of T in the sense of Schwartz, and $\mathbf{T}^{0(n)}(z)=(1/2\pi i)T^{(n)}\cdot(x-z)^{-1}$.

In the proof of theorem 1 we have already shown that

$$\frac{\partial}{\partial z}\mathbf{T}^0(z)=\frac{1}{2\pi i}T\cdot(x-z)^{-2}.$$

Similarly, one obtains for the nth derivative

$$\frac{\partial^n}{\partial z^n}\mathbf{T}^0(z)=\frac{n!}{2\pi i}T\cdot(x-z)^{-n-1}.$$

On the other hand, the derivative of T in Schwartz's theory is defined by

$$T'\cdot\phi=-T\cdot\phi'.$$

Thus

$$T'\cdot(x-z)^{-1}=-T\cdot(d/dx)(x-z)^{-1}=T\cdot(x-z)^{-2},$$

and by iteration we obtain

$$T^{(n)}\cdot(x-z)^{-1}=n!T\cdot(x-z)^{-n-1}.$$

This proves our theorem.

Relation between T and $T^0(z)$

If $T(x)$ is the restriction of a function f holomorphic in the upper half-plane which behaves like $|z|^{-\alpha}$ for $z\to\infty$ for some $\alpha>0$, then $\mathbf{T}^0(z)=f(z)$ in the upper half-plane and $\mathbf{T}^0(z)=0$ in the lower half-plane. [In theorem 1 we had assumed that T has compact support. This is, of course, a sufficient but not necessary condition that $\mathbf{T}^0(z)$ exist. We will discuss the general case of noncompact support later.] If $f(z)$ is holomorphic in the lower half-plane, we find similarly that $\mathbf{T}^0(z)=0$ for $\mathrm{Im}z>0$ and $\mathbf{T}^0(z)=-f(z)$ for $\mathrm{Im}z<0$. Thus, if $T(x)$ is a function analytic in either half-plane, then the "jump" of $T(z)$ on the real axis is equal to $T(x)$. We will show that this latter property holds not only when $T(x)$ is analytic, but that it holds for general distributions.

Theorem 3. *If $T(x)$ is a (C^n) function with compact support, then for $\epsilon\to 0+$ $[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]$ converges uniformly to $T(x)$ on the whole real axis, and the same is true for the derivatives up to nth order. If T is a distribution in $(\mathcal{E}')$, then $[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]$ converges for $\epsilon>0$, $\epsilon\to 0$ to T in the following sense:*

$$T\cdot\phi=\lim_{\epsilon\to 0}\int_{-\infty}^{+\infty}[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]\phi(x)dx$$

for every test function $\phi\in(\mathcal{D})$.

Proof. Let us first consider the case in which $T(x)$ is a continuous function. To demonstrate the uniform convergence, we need an estimate of the quantity $|\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)-T(x)|$ for $\epsilon\to 0$. We have by definition

$$\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}T(\xi)\left[\frac{1}{\xi-x-i\epsilon}-\frac{1}{\xi-x+i\epsilon}\right]d\xi.$$

We split the range of integration as follows:

$$\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)=\frac{1}{2\pi i}\left\{\int_{-\infty}^{x-\delta}+\int_{x-\delta}^{x+\delta}+\int_{x+\delta}^{\infty}\right\}T(\xi)\frac{2i\epsilon}{|\xi-x+i\epsilon|^2}d\xi,$$

where δ is an arbitrary positive constant. Since $T(x)$ is continuous and has compact support $[T\in(\mathcal{E}')]$, it is bounded, $|T(x)|<M$ for all x. Let L be the length of the support of T. Then the integrals over the intervals $\{-\infty, x-\delta\}$, $\{x+\delta, +\infty\}$ are together smaller in absolute value than $(2ML\epsilon)/\delta^2$.

In order to discuss the integral over the interval $\{x-\delta, x+\delta\}$, we split $T(\xi)$ into real and imaginary parts, $T(\xi)=\mathrm{Re}T(\xi)+i\,\mathrm{Im}T(\xi)$. Then

$$\int_{x-\delta}^{x+\delta}\mathrm{Re}T(\xi)\frac{2i\epsilon}{|\xi-x+i\epsilon|^2}d\xi=\mathrm{Re}T(\xi_0)\int_{x-\delta}^{x+\delta}\frac{2i\epsilon}{|\xi-x+i\epsilon|^2}d\xi$$

for some $\xi_0\in\{x-\delta, x+\delta\}$. We can now write

$$\int_{x-\delta}^{x+\delta}\left[\frac{1}{\xi-x-i\epsilon}-\frac{1}{\xi-x+i\epsilon}\right]d\xi=\int_\Gamma\frac{1}{z-x}dz,$$

where Γ is the contour consisting of the directed line from $x-\delta-i\epsilon$ to $x+\delta-i\epsilon$ and the line from $x+\delta+i\epsilon$ to $x-\delta+i\epsilon$. In order to make Γ a closed curve Γ^*, we add the integrals from $x-\delta+i\epsilon$ to $x-\delta-i\epsilon$ and from $x+\delta-i\epsilon$ to $x+\delta+i\epsilon$. These two integrals are together in absolute value less than $4\epsilon/\delta$. Now $\int_{\Gamma^*}[z-x]^{-1}dz=2\pi i$. Therefore,

$$\left|\mathrm{Re}T(x)-\frac{1}{2\pi i}\int_{x-\delta}^{x+\delta}\mathrm{Re}T(\xi)\frac{2i\epsilon}{|\xi-x-i\epsilon|^2}d\xi\right|<|\mathrm{Re}[T(x)-T(\xi_0)]|+\frac{2M}{\pi}\frac{\epsilon}{\delta},$$

and analogously we obtain the same estimate for $\mathrm{Im}T(x)$.

ξ_0 lies in the interval $\{x-\delta, x+\delta\}$. Let $m(\delta)$ be the largest variation of $T(x)$ in any interval of length 2δ. Since $T(x)$ is continuous and of compact support, $m(\delta)\to 0$ as $\delta\to 0$. Thus $|\mathrm{Re}[T(x)-T(\xi_0)]|<m(\delta)$, and we obtain as a total estimate

$$|\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)-T(x)|<\frac{2M\cdot L\cdot\epsilon}{\pi\delta^2}+2m(\delta)+\frac{4M}{\pi}\frac{\epsilon}{\delta}.$$

If we choose $\delta=\epsilon^{\frac{1}{2}}$, then the right-hand side tends to zero for $\epsilon\to 0$. The estimate is independent of x. Hence $\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)$ converges to $T(x)$ uniformly for all x, as asserted previously.

If now $T(x)$ is a (C^n) function, then we can apply the result just proved to the derivatives $T^{(r)}$ for $r\leq n$. On the other hand, we have by theorem 2

$$\frac{\partial^r}{\partial z^r}[\mathbf{T}^0(z)]=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}T^{(r)}(x)\frac{1}{x-z}dx.$$

Consequently,

$$\lim_{\epsilon\to 0}\left[\frac{\partial^r}{\partial z^r}\mathbf{T}^0(x+i\epsilon)-\frac{\partial^r}{\partial z^r}\mathbf{T}^0(x-i\epsilon)\right]=T^{(r)}(x)$$

and the convergence is uniform on the whole real axis. This proves our theorem for the case where $T(x)$ is a (C^n) function.

Proof of the general case. If T is an arbitrary distribution of compact support, then the limit in the ordinary sense in general does not exist. Instead we have to show that

$$\lim_{\epsilon\to 0}\int_{\infty}^{+\infty}[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]\phi(x)dx=T\cdot\phi$$

for every test function $\phi\in(\mathfrak{D})$. Since $\mathbf{T}^0(z)$ is holomorphic for $\mathrm{Im}z\neq 0$, and ϕ has compact support, this integral exists. Let

$$\begin{aligned}I(\epsilon)&=\int_{-\infty}^{+\infty}[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}_0(x-i\epsilon)]\phi(x)dx\\&=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\phi(x)T_\xi\cdot\left[\frac{1}{\xi-x-i\epsilon}-\frac{1}{\xi-x+i\epsilon}\right]dx\\&=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\phi(x)T_\xi\cdot\left(\frac{2i\epsilon}{|\xi-x+i\epsilon|^2}\right)dx\\&=\lim_{N\to\infty}\frac{1}{2\pi i}\sum_{k=1}^{N}\Delta x_k\phi(x_k)T_\xi\cdot\left(\frac{2i\epsilon}{|\xi-x_k+i\epsilon|^2}\right).\end{aligned}$$

We would like now to exchange the integration and the application of T. Because T is linear we can write

$$I(\epsilon)=\lim_{N\to\infty}T_\xi\cdot\left[\frac{1}{2\pi i}\sum_{k=1}^{N}\phi(x_k)\Delta x_k\left(\frac{2i\epsilon}{|\xi-x_k+i\epsilon|^2}\right)\right].$$

The bracket to which T is applied is a test function in $(\mathcal{E})$ which converges uniformly to $\phi^0(\xi+i\epsilon)-\phi^0(\xi-i\epsilon)$, where $\phi^0(z)$ is the Cauchy integral of $\phi(x)$. The derivatives with respect to ξ likewise converge uniformly. Therefore, from the definition of continuity in $(\mathcal{E})$, we obtain

$$I(\epsilon)=T\cdot[\phi^0(\xi+i\epsilon)-\phi^0(\xi-i\epsilon)].$$

$\phi(x)$ is a (C^∞) function; consequently $[\phi^0(\xi+i\epsilon)-\phi^0(\xi-i\epsilon)]$ converges uniformly to $\phi(\xi)$, together with all derivatives, as we have proved. This permits us to exchange once more the application of T and the limit:

$$\lim_{\epsilon\to 0}I(\epsilon)=\lim_{\epsilon\to 0}T\cdot[\phi^0(\xi+i\epsilon)-\phi^0(\xi-i\epsilon)]=T\cdot\phi.$$

Thus $I(\epsilon)$ converges to $T\cdot\phi$ for every $\phi\in(\mathfrak{D})$. This proves our theorem.

Examples.

1. Dirac's δ function is defined by

$$\int_{-\infty}^{+\infty}\delta(x)\phi(x)dx=\phi(0)\quad\text{for all }\phi\in(\mathcal{E}).$$

The Cauchy integral of $\delta(x)$ gives

$$\delta^0(z)=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\delta(x)\frac{1}{x-z}dx=-\frac{1}{2\pi iz}.$$

2. Derivatives of the δ function: $\delta^{0(n)}(z)=(d^n/dz^n)\times\delta^0(z)$ [theorem 2]. Thus,

$$\delta^{0(n)}(z)=\frac{(-1)^{n+1}n!}{2\pi i}\frac{1}{z^{n+1}}.$$

3. Finite step function:

$$\Theta_{a,b}(x)=1\quad\text{for }a<x<b;$$

$$\Theta_{a,b}(x)=0\quad\text{for }x<a\text{ and }b<x.$$

$$\Theta^0{}_{a,b}(z)=\frac{1}{2\pi i}\int_a^b\frac{1}{x-z}dx=\frac{1}{2\pi i}\log\frac{b-z}{a-z}.$$

$\Theta^0{}_{a,b}(z)$ is analytic in the cut z plane, with the cut connecting a to b. $\Theta^0{}_{a,b}(z)$ can be continued holomorphically across the cut from a to b, except at the end points. This continuation leads, of course, to a multiple valued function. However, we must choose that branch of $\log[(b-z)/(a-z)]$ that approaches zero at infinity [theorem 1].

4. Let T be a distribution whose support consists of an isolated point a. Then, according to Schwartz[1] T is a finite linear combination of Dirac's δ functions and its derivatives:

$$T(x)=\sum_{\nu=1}^{N}a_\nu\delta^{(\nu-1)}(x-a).$$

The Cauchy integral is consequently given by

$$\mathbf{T}^0(z)=\sum_{\nu=1}^{N}a_\nu\frac{(-1)^\nu(\nu-1)!}{2\pi i}(z-a)^{-\nu}=\sum_{\nu=1}^{N}a_\nu^*(z-a)^{-\nu},$$

if we denote $a_\nu(-1)^\nu(\nu-1)!(2\pi i)^{-1}$ by a_ν^*. Thus $\mathbf{T}^0(z)$ is a finite polynomial in $1/(z-a)$.

5. To every distribution T with compact support there corresponds a function holomorphic in the whole z plane minus the support of T, and tending to zero at infinity.

It is natural to ask whether the converse is true: Given a function $f(z)$, holomorphic in the z plane minus a compact set σ on the real axis and satisfying $f(z) \to 0$ for $z \to \infty$, does there exist a distribution with support σ such that $\mathbf{T}^0(z) = f(z)$?

The answer is negative. The function $e^{1/z}-1$ is a counter example: $e^{1/z}-1$ tends to zero for $z \to \infty$, and is holomorphic except for the origin. If there would exist a distribution T such that $\mathbf{T}^0(z) = e^{1/z}-1$, then T would have to have as support the origin, that is an isolated point. Hence, according to the preceding example,

$$\mathbf{T}^0(z) = \sum_{\nu=1}^{N} a_\nu^* z^{-\nu},$$

where N is finite. On the other hand, we have

$$e^{1/z}-1 = \sum_{\nu=1}^{\infty} \frac{1}{\nu!} z^{-\nu},$$

where the sum goes to infinity. Hence there does not exist a distribution T such that $\mathbf{T}^0(z) = e^{1/z}-1$.

[It can be shown, making use of Schwartz[6] that in general $\mathbf{T}^0(z)$ cannot have worse than "polar behavior" on the real axis.]

IV. "TOTAL VALUE" OF A DISTRIBUTION AS A GENERALIZED RESIDUE OF $T^0(z)$

$T \cdot 1$, that is a distribution applied to the test function $\phi(x) \equiv 1$, is called the "total value" or "integral" of T (Schwartz,[1] p. 88).

As a first demonstration of the usefulness of $\mathbf{T}^0(z)$ we have:

Theorem 4. *Let T be a distribution with compact support. Then*

$$T \cdot 1 = \int_{\infty}^{+\infty} T(x)dx = \int_{C0} \mathbf{T}^0(z)dz,$$

where C_0 is any simple closed curve circling the support of T clockwise.

If T has compact support $\sum$, then we can restrict the integration from $-\infty$ to $+\infty$ to an integration over $\sum$. Intuitively theorem 4 means the following:

If we replace $T(x)$ by $\mathbf{T}^0(z)$, then by theorem 3 $\int_{\Sigma} T(x)dx$ is replaced by two integrals: One integral over the "upper value" {limit as $\epsilon \to 0+$ $\mathbf{T}^0(x+i\epsilon)$} integrated over $\sum$ in positive direction plus the integral over the "lower value" {limit as $\epsilon \to 0+$ $\mathbf{T}^0(x-i\epsilon)$} integrated along $\sum$ in the negative direction.

The integration appears thus as an integration over a curve shrunk to an upper and lower layer along $\sum$. But since $\mathbf{T}^0(z)$ is holomorphic in the whole z plane minus $\sum$, we may deform the curve without changing the integral.

This idea can be made rigorous as follows. Let $\sum'$ be a neighborhood of $\sum$. We take $\sum'$ to consist of finitely many intervals. We then can treat each one separately. We thus will assume in the following that $\sum'$ is an interval: $\sum' = \{a,b\}$.

$$\int_a^b T(x)dx = \lim_{\epsilon \to 0+} \int_a^b [\mathbf{T}^0(x+i\epsilon) - \mathbf{T}^0(x-i\epsilon)]dx.$$

Thus for $\epsilon > 0$ sufficiently small,

$$\left| \int_a^b T(x)dx - \int_a^b [\mathbf{T}^0(x+i\epsilon) - \mathbf{T}^0(x-i\epsilon)]dx \right| < \frac{\delta}{2}$$

for any $\delta > 0$. However, we can write the integral $\int_a^b [\mathbf{T}^0(x+i\epsilon) - \mathbf{T}^0(x-i\epsilon)]dx$ as $\int_{C(\epsilon)} \mathbf{T}^0(z)dz$, where $C_{(\epsilon)}$ consists of the lines from $a+i\epsilon$ to $b+i\epsilon$ and $b-i\epsilon$ to $a-i\epsilon$. To obtain a closed contour, we add the integral over $T(z)$ from $a-i\epsilon$ to $a+i\epsilon$ and $b+i\epsilon$ to $b-i\epsilon$. Since $\mathbf{T}^0(z)$ is holomorphic, hence bounded along these paths, this addition will be smaller than $\delta/2$, if ϵ is only made small enough. Thus with $C_{(\epsilon)}^*$ the completed contour,

$$\left| \int_a^b T(x)dx - \int_{C(\epsilon)^*} T(z)dz \right| < \delta$$

for sufficiently small ϵ. But since $\mathbf{T}^0(z)$ is holomorphic in the {z plane} $-\sum$, we have at once

$$\int_{C(\epsilon)^*} \mathbf{T}^0(z)dz = \int_{C0} \mathbf{T}^0(z)dz$$

for any given curve C_0 in ({z plane} $-\sum$) that is homologous to $C_{(\epsilon)}^*$, that is for any simple closed curve circling $\sum$ clockwise.

Therefore,

$$\left| \int_a^b T(x)dx - \int_{C(\epsilon)^*} T(z)dz \right| < \delta$$

does not depend upon ϵ. It is smaller than any $\delta > 0$. Hence the two integrals are equal.

Theorem 4 is obviously a generalized residue theorem.

Examples.

1. $T(x) = \delta(x) + \delta'(x)$. Then

$$\mathbf{T}^0(z) = -\frac{1}{2\pi i z} + \frac{1}{2\pi i z^2}.$$

The support consists of the origin. Hence

$$T \cdot 1 = -\frac{1}{2\pi i} \int_{C0} \left(\frac{1}{z} - \frac{1}{z^2} \right) dz = \operatorname{Res} \left(\frac{1}{z} - \frac{1}{z^2} \right) = 1.$$

2. Let T have support at a finite number of isolated points $a_1, \cdots a_m$. Then (compare example 4 of Sec. 3)

$$\mathbf{T}^0(z)=-\frac{1}{2\pi i}\sum_{\mu=1}^{m}\sum_{\nu=1}^{N} a_{\mu\nu}(z-a_\mu)^{-\nu}.$$

Then

$$T\cdot 1=\text{sum of residues of } \mathbf{T}^0(z)=\sum_{\mu=1}^{m} a_{\mu 1}.$$

V. DISTRIBUTIONS WITH NONCOMPACT SUPPORT

The function $(x-z)^{-1}$ (for $\mathrm{Im}z\neq 0$) belongs to $(\mathcal{E})$ but not to $(\mathcal{D})$, and consequently $T\cdot(x-z)^{-1}$ is not defined for all $T\in(\mathcal{D}')$. On the other hand, $T\cdot(x-z)^{-1}$ does exist for certain $T\in(\mathcal{D}')$ that are not in $(\mathcal{E}')$, and that includes important cases such as.

$$T(x)=(x+i\epsilon)^{-1};\quad T(x)=P(x^{-n}),\quad n>0.$$

Since neither $(\mathcal{D}')$ nor $(\mathcal{E}')$ nor any of the other distribution spaces defined by Schwartz suits the problem of studying $\mathbf{T}^0(z)$, we will introduce in the following a new space:

Definition 1. *Let (V) be the subspace of all $\phi\in(\mathcal{E})$ which have the following properties:*

$$\phi(x)|x|\leq K_0 \quad \text{for } |x|\to\infty,$$
$$\phi^{(n)}(x)|x|\leq K_n \quad \text{for } x\to\infty,$$

where K_0, K_1, $K_2\cdots$ are constants. Let convergence be defined as in $(\mathcal{E})$.

The corresponding distribution space, which we denote by $(\mathcal{U}')$, is the dual space of $(\mathcal{U})$.

Remark. In the definition of $(\mathcal{U})$ we include the boundedness condition for the derivatives $\phi^{(n)}$ of ϕ to insure that $T^{(n)}\cdot\phi=(-1)^nT\cdot\phi^{(n)}$ is defined.

Lemma 1. *Theorems 1, 2, and 3 remain true if $T\in(\mathcal{E}')$ is replaced by $T\in(\mathcal{U}')$, and (in theorem 3) "uniform convergence on the whole real axis" is replaced by "uniform convergence on every compact interval of the real axis."*

Theorem 1 asserts the existence and analyticity of $\mathbf{T}^0(z)=(1/2\pi i)T\cdot(x-z)^{-1}$. Since $(x-z)^{-1}\in(\mathcal{U})$ for $\mathrm{Im}z\neq 0$, and since $(x-z)^{-1}\to 0$ uniformly for $|z|\to\infty$ and x in a compact subset of the real axis, $\mathbf{T}^0(z)$ exists for all $T\in(\mathcal{U}')$, and $\mathbf{T}^0(z)\to 0$ for $|z|\to\infty$. The analyticity of $\mathbf{T}^0(z)$ outside the [possibly unbounded] support of T follows exactly as for $T\in(\mathcal{E}')$. Theorem 2 follows as before and from the observation that $(x-z)^{-n}\in(\mathcal{U})$ for all $n\geq 1$, $\mathrm{Im}z\neq 0$.

To extend theorem 3 we have to show that for continuous functions $T(x)\in(\mathcal{U}')$

$$[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)-T(x)]$$

converges uniformly to zero for x in any compact interval on the real axis. If this has been established, then it follows immediately that if $T(x)$ is a (C^n) function in $(\mathcal{U}')$, then also the derivatives up to nth order converge uniformly.

No change of the proof in the "general case": $\int_{-\infty}^{+\infty}[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]\phi(x)dx\to T\cdot\phi$ is necessary since we assume that $\phi\in(\mathcal{D})$.

To show the uniform convergence of $\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)$ for continuous $T(x)$ we proceed as follows: Let $\alpha_R(x)$ be a (C^∞) function with the following property:

$$\alpha_R(x)\equiv 1 \quad \text{for } |x|\leq R$$
$$0\leq\alpha_R(x)\leq 1 \quad \text{for } R<|x|<2R$$
$$\alpha_R(x)\equiv 0 \quad \text{for } |x|\geq R.$$

Let

$$T_R(x)=\alpha_R(x)T(x),$$

then $T_R(x)$ has compact support.

$$T(x)-T_R(x)=T(x)[1-\alpha_R(x)].$$

Let

$$\beta_R(x)=1-\alpha_R(x).$$

Then

$$\beta_R(x)\equiv 0 \quad \text{for } |x|\leq R$$
$$0\leq\beta_R(x)\leq 1 \quad \text{for } R<|x|<2R$$
$$\beta_R(x)\equiv 1 \quad \text{for } |x|\geq R.$$

We have

$$\mathbf{T}^0(z)-\mathbf{T}_R{}^0(z)=(2\pi i)^{-1}T\cdot\beta_R(x)(x-z)^{-1}.$$

Let D be any compact domain of the z plane (which may or may not intersect the real x axis). Then for all sufficiently large R, $\beta_R(x)(x-z)^{-1}$ is a (C^∞) function in $(\mathcal{U}')$ for all $z\in D$. For $R\to\infty$ the functions $\beta_R(x)\times(x-z)^{-1}$ converge to zero, uniformly in x and z, for x in any compact interval of the real axis, and for $z\in D$. Hence

$$\mathbf{T}^0(z)-\mathbf{T}_R{}^0(z)$$

converges to zero for $R\to\infty$, uniformly for all $z\in D$.

Now, since T_R has compact support,

$$\mathbf{T}_R{}^0(x+i\epsilon)-\mathbf{T}_R{}^0(x-i\epsilon)\to T_R(x)$$

uniformly on the whole real axis. Also, given any compact interval on the real axis, then $T_R(x)=T(x)$ in the interval for all sufficiently large R. By combining the two uniform convergences, we obtain that

$$\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)\to T(x)$$

uniformly in every compact interval on the real axis, which was to be proved.

Definition 2. *Let $T\in(\mathcal{U}')$. Then we say: "$T(x)$ vanishes for $|x|\to\infty$" if for every $\epsilon>0$ there exists an R such that $|T\cdot\phi|<\epsilon$ for every $\phi\in(\mathcal{D})$ which has the following properties: (1) $\phi(x)\geq 0$ for all x; (2) the support of ϕ is contained in $\{x\,|\,|x|>R\}$, and (3) $\int_{-\infty}^{+\infty}\phi(x)dx=1$.*

If $T(x)$ is a function which vanishes for $|x|\to\infty$, then we can find for every $\epsilon>0$ an R such that for $|x|>R$, $|T(x)|<\epsilon$. Then, if ϕ is a test function with

the three properties stated,

$$\left|\int_{-\infty}^{+\infty} T(x)\phi(x)dx\right| < \epsilon \int_{-\infty}^{+\infty} \phi(x)dx = \epsilon;$$

and conversely, if a function $T(x)$ vanishes in the sense of definition 2, then it vanishes in the ordinary sense. Our definition of the meaning of the "vanishing of $T(x)$ at infinity" is therefore an extension of the usual notion for functions.

Corollary. *If $T\in(\mathcal{U}')$, then $T(x)$ vanishes for $|x|\to\infty$.*

By theorem 3 we have for $\phi\in(\mathfrak{D})$

$$\int_{-\infty}^{+\infty} T(x)\phi(x)dx = \lim_{\delta\to 0}\int_{-\infty}^{+\infty}[\mathbf{T}^0(x+i\delta)-\mathbf{T}^0(x-i\delta)]\phi(x)dx.$$

From theorem 1 it follows that $\mathbf{T}^0(z)\to 0$ for $z\to\infty$. Hence we can find for every $\epsilon>0$ an R such that $|\mathbf{T}^0(z)|<\epsilon/2$ for $|z|>R$. It follows that

$$|T\cdot\phi| \leqslant \epsilon\int_{-\infty}^{+\infty}\phi(x)dx = \epsilon$$

for ϕ with the properties previously described. Hence $T(x)$ vanishes in the sense of our definition.

While $T(x)\to 0$ for $|x|\to\infty$ is a necessary condition that $T(x)$ belongs to $(\mathcal{U}')$, it is not sufficient, as the following example shows:

$$T(x)=1/\log(1+|x|)$$

tends to zero for $|x|\to\infty$. However,

$$\int_{-R}^{+R}\frac{dx}{|x|\,\log(1+|x|)}$$

does not converge for $R\to\infty$, since $\log(1+|x|)$behaves as $\log|x|$ for large $|x|$, and $\int 1/x\log x dx=\log\log x$.

Sufficient condition. Let $T(x)$ be a summable function such that there exists an $\alpha>0$ and a constant A such that

$$|x|^{\alpha}|T(x)|<A \quad \text{for } |x|\to\infty.$$

Then T belongs to $(\mathcal{U}')$. The proof is immediate. Since $T(x)$ behaves like $A|x|^{-\alpha}$ for large $|x|$, $T(x)\phi(x)$ behaves like $AK_0|x|^{-1-\alpha}$ for $\phi(x)\in(\mathcal{U})$ [definition 1]. Thus $\int_{-\infty}^{+\infty}T(x)\phi(x)dx$ converges, and $T\in(\mathcal{U}')$.

Examples. A number of distributions which are encountered frequently in the problems of physics and which are not contained in $(\mathcal{E}')$ are contained in $(\mathcal{U}')$. The following are examples.

1. $T(x)=P(x^{-n})$, $n\geq 1$, n integer. Then

$$\mathbf{T}^0(z)=\frac{1}{2\pi i}P\int_{-\infty}^{+\infty}\frac{1}{x^n}\frac{1}{x-z}dx$$
$$=1/2z^n \quad \text{for } \mathrm{Im}z>0$$
$$=-1/2z^n \quad \text{for } \mathrm{Im}z<0$$

(compare Sec. II, example 2).

2. $T(x)=1/(x+i\epsilon)$, $\epsilon>0$.

$T(x)$ is the restriction to the real axis of a function which is holomorphic for all z except for $z=-i\epsilon$. Thus,

$$\mathbf{T}^0(z)=1/(z+i\epsilon) \quad \text{for } \mathrm{Im}z>0,$$

and

$$\mathbf{T}^0(z)=0 \quad \text{for } \mathrm{Im}z<0.$$

3. $\delta_+(x)$ is defined as follows:

$$\delta_+(x)=-\frac{1}{i\pi}\lim_{\epsilon\to 0+}\left(\frac{1}{x+i\epsilon}\right).$$

From the preceding example, we see that

$$\boldsymbol{\delta}_+^0(z)=-(1/i\pi)(1/z) \quad \text{for } \mathrm{Im}z>0,$$
$$\boldsymbol{\delta}_+^0(z)=0 \quad \text{for } \mathrm{Im}z<0.$$

4. In physics one makes use of the "symbolic identities"

$$-i\pi\delta_+(x)=\lim_{\epsilon\to 0+}\frac{1}{x+i\epsilon}=P\left(\frac{1}{x}\right)-i\pi\delta(x)$$

$$i\pi\delta_+(-x)=\lim_{\epsilon\to 0+}\frac{1}{x-i\epsilon}=P\left(\frac{1}{x}\right)+i\pi\delta(x).$$

To prove these identities rigorously for the analytic continuations we only have to make use of the examples 1–3.

5. From examples 1-3 we also immediately obtain the identities

$$\delta_+(x)+\delta_+(-x)=2\delta(x),$$

and

$$\delta_+(x)-\delta_+(-x)=-(2/i\pi)P(1/x).$$

VI. ANALYTIC TEST FUNCTIONS

In computing $T\cdot 1$, the "total value" of a distribution, we were able to replace the integral over the real axis by a contour C_0 circling the support of T. It would be convenient if we could do the same in general, that is, if we could write

$$T\cdot\phi=\int_{-\infty}^{+\infty}T(x)\phi(x)dx=\int_{C_0}\mathbf{T}^0(z)\boldsymbol{\phi}^0(z)dz. \quad (?)$$

This equality, however, does not hold in general. Indeed

$$T\cdot\phi=\lim_{\epsilon\to 0}\int_{-\infty}^{+\infty}[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]\times[\boldsymbol{\phi}^0(x+i\epsilon)-\boldsymbol{\phi}^0(x-i\epsilon)]dx,$$

while

$$\int_{C_0}\mathbf{T}^0(z)\boldsymbol{\phi}^0(z)dz = \int_{-\infty}^{+\infty}[\mathbf{T}^0(x+i\epsilon)\boldsymbol{\phi}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)\boldsymbol{\phi}^0(x-i\epsilon)]dx$$

(independently of ϵ).

The two integrals are obviously not equal in general. They are equal if $\phi(x)$ is the restriction of an analytic function $\phi(z)$ and if we replace $\phi^0(z)$ by $\phi(z)$. [Note that $\phi^0(z) \not\equiv \phi(z)$, if $\phi^0(z)$ is the Cauchy integral of the restriction of $\phi(z)$ to the real axis.]

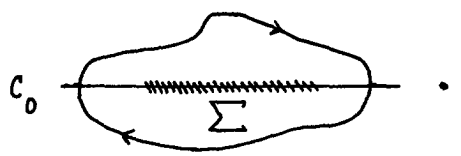
This leads us to consider distributions defined with respect to the class of analytic test functions.

Definition 3. *We denote by* $(\mathcal{A})$ *the set of all entire functions, by* $(\mathcal{A}_b)$ *the set of all functions that are holomorphic in the strip* $\{z \mid |Imz| < b\}$, *where* $0 < b \leq \infty$.

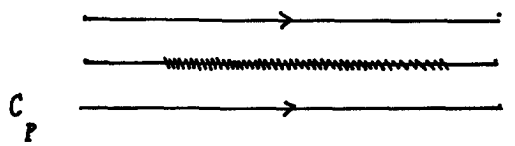
We have $(\mathcal{A}_\infty) = (\mathcal{A})$, and for all b: $(\mathcal{A}) \subset (\mathcal{A}_b) \subset (\mathcal{E})$.

In the following sections, repeated use will be made of several contours in the complex plane; we therefore introduce the following notation.

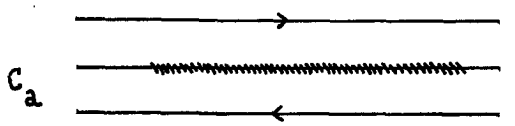
Definition 4. *By* C_0 *we denote any simple closed curve that circles the support of the given distribution clockwise:*



By C_P *we will denote two lines parallel to the real axis, both directed like the real axis, one above and one below:*



By C_a *we will denote two lines, one in the upper half-plane and directed parallel to the real axis, and one in the lower half-plane, directed opposite (antiparallel) to the real axis:*



Finally we will denote the strip $\{z \mid |Im\, z| < b\}$ *by* S_b.

Theorem 5. *Let* $T \in (\mathcal{E}')$, $\phi \in (\mathcal{A}_b)$, $0 < b \leq \infty$. *Then*

$$T \cdot \phi = \int_{-\infty}^{+\infty} T(x)\phi(x)dx = \int_0 \mathbf{T}^0(z)\phi(z)dz,$$

for any $C_0 \subset S_b$.

The proof is analogous to the proof of theorem 4, as is the proof of the following:

Theorem 5a. *Let* $T \in (\mathcal{U}')$, $\phi \in (\mathcal{A}_b)$, $0 < b < \infty$, *and* $|\phi| \leq A|z|^{-1}$ *for* $z \to \infty$. *Then*

$$T \cdot \phi = \int_{C_a} \mathbf{T}^0(z)\phi(z)dz$$

for any $C_a \subset S_b$.

We note that $|\phi| \leq A|z|^{-1}$ for $z \to \infty$ cannot be satisfied by any $\phi \in (\mathcal{A}) = (\mathcal{A}_\infty)$.

Example. Let $T(x) = \delta^{(n)}(x)$, $\phi(x) \in (\mathcal{A}_b)$. Then $\mathbf{T}^0(z) = (-1)^{n+1} n!/2\pi i z^{n+1}$, and

$$\int_{-\infty}^{\infty} \delta^{(n)}(x)\phi(x)dx = \frac{(-1)^{n+1}n!}{2\pi i} \int_{C_0} z^{-n-1}\phi(z)dz$$

$$= (-1)^n n!\, \mathrm{Res}[z^{-n-1}\phi(z)]_{z=0} = (-1)^n \phi^{(n)}(0)$$

in agreement with the usual result.

Convergence in $(\mathcal{A}_b)$. We define convergence in $(\mathcal{A}_b)$, $0 < b \leq \infty$, as follows: *A sequence* $\{\phi_j(z)\}$ *of functions* $\phi_j(z) \in (\mathcal{A}_b)$ *converges to zero if it converges to zero uniformly in every compact subset in* S_b.

The dual space of $(\mathcal{A}_b)$ *we denote, as usual, by* $(\mathcal{A}_b')$. It has also been studied by Ehrenpreis.[16]

Since $(\mathcal{A}_b) \subset (\mathcal{E})$ properly, the associated dual space $(\mathcal{A}_b')$ is larger than $(\mathcal{E}')$. A similar situation was encountered with respect to the set $(\mathcal{U})$ of linearly vanishing test functions, $(\mathcal{U}) \subset (\mathcal{E})$, the dual space $(\mathcal{U}') \supset (\mathcal{E}')$ containing many distributions of practical interest. It is useful also in the present case to consider the subset of linearly vanishing functions in $(\mathcal{A}_b)$. This subspace consists only of the constant 0 for $b = \infty$, but it is nontrivial for $0 < b < \infty$.

Definition 5. *We denote by* $(\mathcal{A}_b{}^V)$ *the subspace of* $(\mathcal{A}_b)$ *consisting of all* $\phi \in (\mathcal{A}_b)$ *which vanish linearly for* $z \to \infty$:

$$|\phi^{(n)}(z)| \leq K_n |z|^{-1} \quad \text{for } z \to \infty, \quad n = 0, 1, 2 \cdots.$$

Convergence is defined as in $(\mathcal{A}_b)$.

Obviously, $(\mathcal{A}_b{}^{V'}) \supset (\mathcal{A}_b')$ and $(\mathcal{A}_b{}^{V'}) \supset (\mathcal{U}')$. Theorems 1–3 and 5a hold for $T \in (\mathcal{A}_b{}^{V'})$ just as for $T \in (\mathcal{U}')$. We note also that the condition $\mathbf{T}^0(z) \to 0$ for $|z| \to \infty$ is again a necessary condition for $T \in (\mathcal{A}_b{}^{V'})$, $0 < b < \infty$, while $|\mathbf{T}^0(z)| \leq A|z|^{-\alpha}$, $|z| \to \infty$, $\alpha > 0$ is a sufficient condition.

VII. MULTIPLICATION OF DISTRIBUTIONS

To define a multiplication for distributions, it would seem natural to write for a product ST

$$ST \cdot \phi = \int_{-\infty}^{+\infty} S(x)T(x)\phi(x)dx.$$

However, this definition leads to difficulties. For example, $T(x) = |x|^{-\frac{1}{2}}$ is summable at the origin and defines a distribution in $(\mathcal{D}')$, but $[T(x)]^2$ is not summable. Even worse is $\int_{-\infty}^{+\infty} \delta(x)\delta(x)dx$. Since $\int_{-\infty}^{+\infty} \delta(x)\phi(x)dx = \phi(0)$, one might interprete $\int_{-\infty}^{+\infty} \delta(x) \times \delta(x)dx = \delta(0)$, but $\delta(0)$ is undefined. If $\delta(x)$ is approximated by ordinary functions, then the sequence diverges.

Schwartz[1] has observed that ST is well defined if locally S is "more regular" than T is "irregular." If T is a general distribution, then this condition means that S has to be a (C^∞) function. In physical applications, on the other hand, ill-defined integrals over

[16] Leon Ehrenpreis, Ann. Math. 63, 129 (1956).

products of δ-like causal functions occur, which lead to difficulties, notably in quantum field theory.

Schwartz[17] has shown that it is impossible to have a multiplication for distributions (not necessarily commutative, but such that the product of two distributions is well defined) that contains x, $P(x^{-1})$, and 1, and that is associative. König has shown[18,19] that if one gives up some of these requirements, then there are many possible "multiplication theories." Bogoliubov and Parasiuk[14] have defined a multiplication for "causal functions." It imitates the "subtraction procedures" used in practical calculations of Feynman diagrams in perturbation theory. The multiplication prescription is rather complicated and limited to the special class of "causal functions."

In a further paper (Bremermann[15]) a multiplication is defined where the product of two distributions, in general, contains arbitrary constants (like constants of integration). It is based on Fourier transforms (and on the results of this paper). (For multiplication based on convolution compare also Ehrenpreis.[20])

In the following we want to demonstrate that the analytic continuations of distributions lend themselves readily to define various multiplications in a rather natural way. The products are well defined, they are distributions not on the spaces $(\mathfrak{D})$ or $(\mathcal{E})$ of (C^∞) functions but operate on spaces of analytic test functions.

1. *Let S, T be two distributions with compact support.* Then $\mathbf{S}^0(z)$ and $\mathbf{T}^0(z)$ exist and are holomorphic outside of the supports of S and T. *One can define a multiplication as follows:*

$$(S\textcircled{1}T)\cdot\phi=\int_{C_0}\mathbf{S}^0(z)\mathbf{T}^0(z)\phi(z)dz,$$

where C_0 circles the union of the supports of S and T, and $\phi(z)\in(\mathfrak{A}_b)$. Since $\mathbf{S}^0(z)\mathbf{T}^0(z)$ is holomorphic outside the union of the supports of S and T, and since $\phi(z)$ is holomorphic and C_0 is bounded, the integral exists for every $\phi(z)\in(\mathfrak{A}_b)$. (We assume that C_0 lies in the strip $\{z|\operatorname{Im}(z)<b\}$). The integral does not depend upon the curve C_0, and $S\textcircled{1}T\cdot\phi$ depends linearly and continuously upon $\phi(z)$, and hence is a distribution in $(\mathfrak{A}_b')$. The multiplication is associative and commutative.

This can be extended to distributions with non-compact support. If the product $\mathbf{S}^0(z)\mathbf{T}^0(z)$ vanishes like $A|z|^{-\alpha}$, $\alpha>0$ and if $\phi(z)\in(\mathfrak{A}_b{}^V)$, then $S\textcircled{1}T\cdot\phi=\int_{C_a}\mathbf{S}^0(z)\mathbf{T}^0(z)\phi(z)dz$ is defined (C_a consists of one line parallel to the real axis and one antiparallel). The condition $|\mathbf{S}^0(z)\mathbf{T}^0(z)|<A|z|^{-\alpha}$ for $|z|\to\infty$ is trivially satisfied if one of the factors has compact support. If S has compact support, then $S(z)$ vanishes at least linearly at infinity while $T(z)$ vanishes of unspecified order, hence $S(z)T(z)$ vanishes at least linearly.

This multiplication is no contradiction to Schwartz's "theorem of impossibility."[17] Neither $T(x)=1$, nor $S(x)=x$ possesses a "Cauchy integral," and secondly, we have analytic test functions rather than (C^∞) test functions.

2. We observe that $\mathbf{f}^0(z)$ equals $f(z)$ for $\operatorname{Im}(z)>0$ [if $f(z)$ is holomorphic in the upper half-plane], while $\mathbf{f}^0(z)=-f(z)$ for $\operatorname{Im}(z)<0$ [if $f(z)$ is holomorphic in the lower half-plane]. Thus one could argue that one should multiply $f(z)$ with a factor (-1) in the lower half-plane. Also, in the section on Fourier transforms, we will see that this new quantity, which we will call the "analytic continuation" has a special meaning for Fourier transforms.

Definition 6. *Let T be a distribution such that $\mathbf{T}^0(z)$ exists. Let*

$$\mathbf{T}^1(z)=\begin{cases}\mathbf{T}^0(z) & \text{for } \operatorname{Im}(z)>0\\ -\mathbf{T}^0(z) & \text{for } \operatorname{Im}(z)<0.\end{cases}$$

We call $\mathbf{T}^1(z)$ the "analytic continuation of the distribution T."

We have

$$T\cdot\phi=\int_{C_a}\mathbf{T}^0(z)\phi(z)dz=\int_{C_p}\mathbf{T}^1(z)\phi(z)dz$$

for analytic test functions $\phi(z)$. Thus we can define a multiplication:

$$S\textcircled{2}T\cdot\phi=\int_{C_p}\mathbf{S}^1(z)\mathbf{T}^1(z)\phi(z)dz.$$

If at least one of the factors has compact support and if $\phi(z)$ vanishes linearly for $|z|\to\infty$, then the existence of the integral is assured, and obviously it depends linearly and continuously upon $\phi(z)$. Thus $S\textcircled{2}T\cdot\phi$ is a distribution in $(\mathfrak{A}_b{}^{V'})$. The multiplication is associative and commutative.

Lemma 2. *Let S, T have compact support, then both $S\textcircled{1}T\cdot 1=0$ and $S\textcircled{2}T\cdot 1=0$.*

Proof. We have $\mathbf{S}^0(z)\mathbf{T}^0(z)=\mathbf{S}^1(z)\mathbf{T}^1(z)$, $\mathbf{S}^0(z)$ and $\mathbf{T}^0(z)$ are holomorphic outside their compact support and vanish at infinity. Therefore each one has a Laurent development that has no constant term. Therefore the product $\mathbf{S}^0(z)\mathbf{T}^0(z)$ vanishes at least quadratically at infinity. Hence

$$\int_{C_0}\mathbf{S}^0(z)\mathbf{T}^0(z)dz=\operatorname{Res}S(z)T(z)=0.$$

For the integral

$$\int_{C_p}\mathbf{S}^1(z)\mathbf{T}^1(z)dz$$

we can, thanks to the quadratic behavior of $\mathbf{S}^1(z)\mathbf{T}^1(z)$ at infinity, close the contour at infinity for the two lines of which C_p exists, and it follows that the integral is zero.

The lemma shows that neither of the two multiplica-

[17] L. Schwartz, Compt. rend. **239**, 847 (1954).
[18] H. König, Math. Ann. **128**, 420 (1954).
[19] H. König, Abhandl. bayer. Akad. Wiss. Math. Naturw. Kl. No. 82 (1957).
[20] Leon Ehrenpreis, Am. J. Math. **76**, 883 (1954); part II, **77**, 286 (1955); part III, **78**, 685 (1956).

tions defined coincides with the ordinary multiplication in the case where $S(x)$ and $T(x)$ are, for instance, continuous functions with compact support.

3. In the following we define a third multiplication which does not suffer from this discrepancy. It reduces to ordinary multiplication for continuous functions. However, it is defined only for a rather limited class of distributions.

Definition 7. *Let S and T be such that $\mathbf{S}^0(z)$ and $\mathbf{T}^0(z)$ exist. Let $\phi\in(\mathfrak{D}')$. Then*

$$S\textcircled{3}T\cdot\phi=\lim_{\epsilon\to 0+}\int_{-\infty}^{+\infty}[\mathbf{S}^0(x+i\epsilon)-\mathbf{S}^0(x-i\epsilon)]\times[\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)]\phi(x)dx.$$

If $S(x)$ and $T(x)$ are continuous functions, then $\mathbf{S}^0(x+i\epsilon)-\mathbf{S}^0(x-i\epsilon)$ converges to $S(x)$ uniformly in every compact interval on the real axis, and in the same way $\mathbf{T}^0(x+i\epsilon)-\mathbf{T}^0(x-i\epsilon)$ converges to $T(x)$. Consequently, the foregoing limit exists and equals $\int_{-\infty}^{+\infty}S(x)T(x)\phi(x)dx$. Thus for continuous functions this multiplication reduces to the ordinary multiplication. If $S(x)$ and $T(x)$ are arbitrary distributions, then the limit may or may not exist.

Examples.

1.
$$\int_{-\infty}^{+\infty}\delta(x)\delta(x)\phi(x)dx$$
$$=\lim_{\epsilon\to 0+}-(2\pi)^{-2}\int_{-\infty}^{+\infty}[(x+i\epsilon)^{-1}-(x-i\epsilon)^{-1}]^2\phi(x)dx$$
$$=\lim_{\epsilon\to 0+}-(2\pi)^{-2}\int_{-\infty}^{+\infty}\{(x+i\epsilon)^{-2}+(x-i\epsilon)^{-2}+(i\epsilon)^{-1}[(x+i\epsilon)^{-1}-(x-i\epsilon)^{-1}]\}\phi(x)dx$$
$$=-(2\pi)^{-2}\int_{-\infty}^{+\infty}2P(x^{-2})\phi(x)dx-\lim_{\epsilon\to 0+}(2\pi\epsilon)^{-1}\times\int_{-\infty}^{+\infty}[(x+i\epsilon)-(x-i\epsilon)]\phi(x)dx.$$

Since the second term contains the factor $(\epsilon)^{-1}$, the limit does not exist.

2.
$$\int_{-\infty}^{+\infty}\delta(x)P(x^{-1})\phi(x)dx$$
$$=\lim_{\epsilon\to 0+}-\frac{1}{4\pi i}\int_{-\infty}^{+\infty}\left[\frac{1}{x+i\epsilon}-\frac{1}{x-i\epsilon}\right]\times\left[\frac{1}{x+i\epsilon}+\frac{1}{x-i\epsilon}\right]\phi(x)dx$$
$$=\lim_{\epsilon\to 0+}-\frac{1}{4\pi i}\int_{-\infty}^{+\infty}\left[\frac{1}{(x+i\epsilon)^2}-\frac{1}{(x-i\epsilon)^2}\right]\phi(x)dx$$
$$=-\frac{1}{2}\int_{-\infty}^{+\infty}\delta'(x)\phi(x)dx=\tfrac{1}{2}\phi'(0).$$

Thus in this particular case the limit does exist.

The examples show that for $\delta(x)\delta(x)$ the multiplication is not defined. One could attempt to make it defined by taking a "finite part" of the integral, for instance by declaring the finite part in example (1) to be the first term [which is equal to $-(1/2\pi^2)P(x^{-2})$]. However, we will not carry out this possibility in this paper.

A more satisfactory solution of the multiplication problem can be obtained by making use of the Fourier transforms of the factors. This has been carried out in Bremermann[15] and has been applied to problems in quantum field theory. In the remaining part of this paper we will provide a basis for Bremermann[15] by studying Fourier transforms and convolutions.

Restriction of the Support of a Distribution (Multiplication with a Step-Function)

The step-function $\Theta(x)$ is defined as follows:

$$\Theta(x)=\begin{cases}1 & \text{for } x>0\\ 0 & \text{for } x<0.\end{cases}$$

For $x=0$, $\Theta(x)$ is undefined [sometimes $\Theta(0)=\frac{1}{2}$ is used]. Note that $\Theta(x)$ is the limit of the "finite step function $\Theta_{0,b}(x)$" (compare example 3, Sec. III), however, the Cauchy integral of $\Theta_{0,b}(x)$ does not converge for $b\to\infty$. $\Theta(x)$ has no Cauchy integral [though in an improper sense we can associate $-(2\pi i)^{-1}\times\log(-z)$ as "improper Cauchy integral" with $\Theta(x)$].

If $T(x)$ is a continuous function, then we may multiply $T(x)$ with $\Theta(x)$ and the product is well defined. If $T(x)$ is only a generalized function, then we encounter difficulties. For example, what is the meaning of $\delta(x)\Theta(x)$? Formally one could write $\int_{-\infty}^{+\infty}\delta(x)\Theta(x)\times\phi(x)dx=\Theta(0)\phi(0)$, however, $\Theta(0)$ is undefined.

For any distribution T the following is true: If ϕ_1 is a test function with support in $x>0$, then

$$\int_{-\infty}^{+\infty}T(x)\Theta(x)\phi_1(x)dx=\int_{-\infty}^{+\infty}T(x)\phi_1(x)dx,$$

while if $\phi_2(x)$ has support in $x<0$, then

$$\int_{-\infty}^{+\infty}T(x)\phi_2(x)dx=0.$$

These properties hold because $\Theta(x)\phi_1(x)$ and $\Theta(x)\phi_2(x)$ are (C^∞) test functions.

To define $T(x)\Theta(x)$ we approximate $\Theta(x)$ by (C^∞) test functions $\psi_j(x)$ with support in $x>0$. Since the convergence is not uniform, the limit $T\cdot\psi_j\phi$ need not exist. If, however, this limit does exist for all ϕ, then we denote it by

$$\int_{-\infty}^{+\infty}T(x)\Theta(x)\phi(x)dx.$$

Correspondingly we define the limit

$$\int_{-\infty}^{+\infty} T(x)\Theta(-x)\phi(x)dx.$$

The values of $\int_{-\infty}^{+\infty}T(x)[\Theta(-x)+\Theta(x)]\phi(x)dx$ and $\int_{-\infty}^{+\infty}T(x)\phi(x)dx$ need not agree in general, however, the two values agree for every test function whose support does not contain the origin. Hence

$$T(x)[1-\Theta(-x)-\Theta(x)]$$

is a distribution whose support consists at most of the origin. Consequently it is a finite linear combination of $\delta(x)$ and its derivatives. Let us denote this distribution by Q. Then

$$T(x)=T(x)\Theta(-x)+T(x)\Theta(x)+Q(x).$$

Examples.

1. $T(x)=\delta(x+1)+\delta(x)+\delta(x-1)$,

then

$$T(x)\Theta(-x)=\delta(x+1)$$
$$T(x)\Theta(x)=\delta(x-1)$$
$$\Theta(x)=\delta(x).$$

2. $T(x)=\delta(x)$.

In this case the limit $T(x)\Theta(x)$ does not exist.

Definition 8. *If T is a distribution such that $T(x)\Theta(x)$ and $T(x)\Theta(-x)$ exist, then we call T "a distribution with a polelike singularity at the origin."*

VIII. CONVOLUTIONS

The convolution product of two functions, $f(x)$ and $g(x)$, denoted by $h=f*g$, is a new function defined by the formula

$$h(x)=\int_{-\infty}^{+\infty} f(x-t)g(t)dt=\int_{-\infty}^{+\infty} g(x-t)f(t)dt.$$

While this product does not converge for arbitrary f and g, convergence is assured if the functions are bounded, and if one of the functions has compact support. On the other hand, if f and g are distributions, this definition is not suitable; we require, rather, the definition of $h\cdot\phi$ for test functions $\phi\in(\mathfrak{D})$ [or some other appropriate space]. By formally interchanging the order of the integrations in $h\cdot\phi$, one obtains

$$h\cdot\phi=\int_{-\infty}^{+\infty} h(x)\phi(x)dx=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f(x-t)g(t)dt\right\}\phi(x)dx$$
$$\rightarrow\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)g(t)\phi(x+t)dxdt.$$

This form is used by Schwartz to define a convolution for distributions. Convergence of $h\cdot\phi$ is not in general assured unless the support of $f(x)g(t)\phi(x+t)$ is compact in the (x,t) plane. We note that the support of $\phi(x+t)$ is not compact in the plane even for $\phi(x)\in(\mathfrak{D})$, so that the classes of distributions for which a convolution product is defined are somewhat restricted. We will not, however, discuss the finer convergence properties here but will proceed immediately to consider the Cauchy integral of a convolution product.

The Cauchy integral of $T*S$ is defined as

$$(T*^0S)(z)\equiv\frac{1}{2\pi i}(T*S)\cdot(x-z)^{-1}$$
$$=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} T(x)S(t)\frac{1}{x+t-z}dxdt.$$

If we formally integrate first over x, and if $\mathbf{T}^0$ exists, we obtain

$$(T*^0S)(z)=\int_{-\infty}^{+\infty}\mathbf{T}^0(z-t)S(t)dt.$$

This has the appearance of the convolution product as defined for functions, except that z is now complex and the distribution $T(x-t)$ has been replaced by $\mathbf{T}^0(z-t)$. We will take the above equation as a definition.

Definition 9. *Let S, $T\in(\mathfrak{A}_b{}')$, $0<b\leq\infty$. Then we define the convolution of T, S by*

$$(T*^0S)(z)=\int_{-\infty}^{+\infty}\mathbf{T}^0(z-t)S(t)dt,$$

provided that the integral converges. We note that S, $T\in(\mathfrak{A}_b{}')$ guarantees the existence of $\mathbf{T}^0(z)$, $\mathbf{S}^0(z)$.

Lemma 3. *If for $|z|\rightarrow\infty$, $|\mathbf{S}^0(z)\mathbf{T}^0(-z)|<A|z|^{-1-\alpha}$, $\alpha>0$, where A is a constant, then $(T*^0S)(z)$ exists.*

According to theorem 5a we can replace $\int_{-\infty}^{+\infty}\mathbf{T}^0(z-t)\times S(t)dt$ by an integral over a curve C_a as long as as $\mathbf{T}^0(z-t')$, t' complex, remains holomorphic in the strip with boundary C_a; that is, the replacement is valid for z outside C_a:

$$\int_{-\infty}^{+\infty}\mathbf{T}^0(z-t)S(t)dt=\int_{C_a}\mathbf{T}^0(z-t')\mathbf{S}^0(t')dt',$$
$$|\mathrm{Im}z|>|\mathrm{Im}t'|.$$

$\mathbf{T}^0(z-t')$ behaves for large t' as $\mathbf{T}^0(-t')$, hence

$$|\mathbf{T}^0(z-t')\mathbf{S}^0(t')|<A|t|^{-1-\alpha}$$

for $|t'|\rightarrow\infty$, and the integral converges.

Theorem 6. *If* $(T*^0S)(z)$ *exists, then*

$$\int_{-\infty}^{+\infty}\mathbf{T}^0(z-t)S(t)dt=\int_{C_a}\mathbf{T}^0(z-t')\mathbf{S}^0(t')dt'=\frac{1}{2\pi i}\int_{C_a\times C_a}\mathbf{T}^0(x')\mathbf{S}^0(t')\frac{1}{x'+t'-z}dx'dt'$$

$$=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}T(x)S(t)\frac{1}{x+t-z}(dxdt)=\int_{C_a}\mathbf{S}^0(z-x')\mathbf{T}^0(x')dx'=\int_{-\infty}^{+\infty}\mathbf{S}^0(z-x)T(x)dx,$$

where C_a *is such that* $z/2$ *lies outside the strip bounded by* C_a.

Corollary. $(T*^0S)(z)=(S*^0T)(z)$.

Proof. If the first integral exists, then it can be replaced by the second according to theorem 5a, where C_a has to be chosen such that $\mathbf{T}^0(z-t)$ is holomorphic in the strip bounded by C_a. This is the case if and only if z lies outside the strip, and is therefore true if $z/2$ lies outside.

We then substitute for $\mathbf{T}^0(z-t')$ the integral

$$\frac{1}{2\pi i}\int_{-\infty}^{+\infty}T(x)\frac{1}{x-(z-t')}dx,$$

which again can be replaced by

$$\frac{1}{2\pi i}\int_{C_a}\mathbf{T}^0(x')\frac{1}{x'-(z-t')}dx',$$

where x' is complex and $z-t'$ outside C_a. The last condition is satisfied if $z/2$ lies outside C_a.

Since we now integrate only over well-behaved holomorphic functions, we can write the iterated integral as a double integral over $C_a\times C_a$ (integral III). By contracting C_a to the real axis and by an argument analogous to that of theorem 4, we obtain the double integral IV. But by writing the integral over $C_a\times C_a$ again as two successive integrals, but now in different order, and reversing the previous steps, we obtain integral V, which leads immediately to integral VI.

Lemma 4. *If either* T *or* S *is a distribution with compact support, then* $(T*^0S)(z)$ exists.

We can assume that S has compact support, which trivially implies the existence of

$$\int_{-\infty}^{+\infty}\mathbf{T}^0(z-t)S(t)dt.$$

Theorem 7. *The convolution product* $(T*^0S)(z)$ *is a holomorphic function for all* z *with* $\mathrm{Im}\,z\neq 0$. *For* $|z|\to\infty$ *the function* $(S*^0T)(z)$ *tends to zero. It defines a distribution* $T*S$ *for all* $\phi\in(\mathfrak{A}_b)$, $0<b\leq\infty$ *which vanish for* $|z|\to\infty$ *as* $A|z|^{-1-\alpha}$, $\alpha>0$, A constant:

$$T*S\cdot\phi=\int_{C_a}(T*^0S)(z)\phi(z)dz.$$

Proof. To show that $(T*^0S)(z)$ is holomorphic, we make use of theorem 7 to write $(T*^0S)(z)$ as the double integral:

$$(T*^0S)(z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}T(x)S(t)\frac{1}{x+t-z}(dxdt).$$

$T(x)S(t)$, the direct product of T and S, is a distribution on R^2, the Euclidean space of two real variables, while $[x+t-z]^{-1}$ is a (C^∞) function on R^2. The further argument that the foregoing integral is a holomorphic function of z for $\mathrm{Im}\,z\neq 0$ is exactly analogous to the proof of theorem 1.

For $|z|\to\infty$ the function $1/(x+t-z)$ tends to zero unformly in every compact subset of R^2, together with all its derivatives. Hence, by continuity, $(T*^0S)(z)$ tends to zero. This implies that $\int_{C_a}(S*^0T)(z)\phi(z)dz$ converges for $\phi(z)$ that behave as $A|z|^{-1-\alpha}$, $\alpha>0$, at infinity. Hence $T*S\cdot\phi=S*T\cdot\phi$ is defined for every ϕ in question. Linearity and continuity are obvious, and the theorem is proved.

Reproducing property of convolutions with $\delta(x)$:

$$\mathbf{T}^0(z)=\int_{-\infty}^{+\infty}\mathbf{T}^0(z-x)\delta(x)dx=\int_{-\infty}^{+\infty}T(x)\boldsymbol{\delta}^0(z-x)dx \qquad \text{for } \mathrm{Im}\,z\neq 0.$$

The first equality is obvious, since for $\mathrm{Im}\,z\neq 0$, $\mathbf{T}^0(z-x)$ is a holomorphic function of x for all x. The second equality follows from theorem 7, but may be verified directly: $\boldsymbol{\delta}^0(z-x)=(1/2\pi i)(1/x-z)$; thus, $\int_{-\infty}^{+\infty}T(x)\times\boldsymbol{\delta}^0(z-x)dx=(1/2\pi i)T\cdot[x-z]^{-1}=\mathbf{T}^0(z)$, and we see that *the Cauchy integral* $\mathbf{T}^0(z)$ *is the same as the convolution product of* T *and* δ,

$$\mathbf{T}^0(z)=(T*^0\delta)(z).$$

The convolution product can obviously be generalized to three or more distributions, for if $T*S=S*T$ exists, and if U has compact support, we may define

$$[(S*T)*\mathfrak{U}]^0=\int_{-\infty}^{\infty}(S*^0T)(z-t)U(t)dt$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{S}^0(z-x-t)T(x)U(t)dxdt.$$

The product is associative and commutative.

IX. FOURIER TRANSFORMS

The Fourier transform of a function is defined as

$$\int_{-\infty}^{+\infty} f(x)e^{ipx}dx,$$

and for a distribution:

$$T\cdot e^{ipx}.$$

Remark. Sometimes a factor 2π is included in the exponent of e: $e^{2\pi ipx}$. We will use the transform defined without this factor.

Theorem 8. *If T is a distribution with compact support, then $T\cdot e^{ipx}$ exists and is a holomorphic function of p in the whole complex p plane. We will write*

$$T\cdot e^{ipx}=\mathfrak{F}(T,p).$$

The function e^{ipx} is for every p a test function in $(\mathcal{E})$, hence $T\cdot e^{ipx}$ exists for $T\in(\mathcal{E}')$. That $T\cdot e^{ipx}$ is a holomorphic function of p for all p follows as in theorem 1 by showing that it is complex differentiable with respect to p. (Schwartz[1] has proved the following stronger result: Necessary and sufficient for T to have compact support is that $\mathfrak{F}(T,p)$ is an entire function of exponential type $\leq C$, where C is some constant; see footnote 1, Vol. 2, theorem XVI).

Lemma 5. *Let $T\in(\mathcal{E}')$. Then*

$$\mathfrak{F}(T,p)=T\cdot e^{ipx}=\int_{C_0}\mathbf{T}^0(z)e^{ipz}dz.$$

This is an immediate consequence of theorem 5. We have nevertheless stated it as a different lemma because it permits us to compute some Fourier transforms as *residues*.

Examples.

1. $$\int_{-\infty}^{+\infty}\delta(x)e^{ipx}dx=e^0=1.$$

This can also be computed as a residue:

$$\int_{-\infty}^{+\infty}\delta(x)e^{ipx}dx=-\frac{1}{2\pi i}\int_{C_0}\frac{1}{z}e^{ipz}dz$$

$$=-\frac{1}{2\pi i}\int_{C_0}\frac{1}{z}\left(1+\frac{ipz}{1!}+\frac{(ipz)^2}{2!}+\cdots\right)dz=1.$$

2. For the derivatives of the δ function we have

$$\int_{-\infty}^{+\infty}\delta^{(n)}(x)e^{ipx}=(-1)^n\frac{d^n}{dx^n}e^{ipx}\Big|_0=(-1)^n(ip)^n.$$

This, again, can be computed as a residue.

3. *If $T\in(\mathcal{E}')$ has point support at $x=0$, then $\mathfrak{F}(T,p)$ is a polynomial in p.*

If T has point support at $x=0$, then $T(x)$ is a finite linear combination of $\delta(x)$ and its derivatives. The result follows then from example 2.

Theorem 9. *Let $T\in(\mathcal{E}')$. Then*

$$T^{(n)}\cdot e^{ipx}=(-ip)^nT\cdot e^{ipx}=(-ip)^n\mathfrak{F}(T,p),$$

where $T^{(n)}$ is the nth derivative of T.

We have by definition: $T^{(n)}\cdot e^{ipx}=(-1)^nT\cdot(d^n/dx^n)\times e^{ipx}=(-ip)^nT\cdot e^{ipx}$.

Noncompact Support

If the support of T is not compact, then the situation is quite different. For ordinary functions, Plancherel's theorem states that, if $f(x)$ is square integrable from $-\infty$ to $+\infty$, then $\mathfrak{F}(f)=g(p)=\int_{-\infty}^{+\infty}f(x)e^{ipx}dx$ exists and is square integrable. In particular the inverse Fourier transform $\mathfrak{F}^{\text{inv}}(g)=f^*(x)=\int_{-\infty}^{+\infty}g(p)e^{-ipx}dp$ exists, and $f(x)=(1/2\pi)f^*(x)$. {If one writes the factor 2π in the exponent, then $\mathfrak{F}^{\text{inv}}[\mathfrak{F}(f)]=f$, without the factor $(2\pi)^{-1}$.} However, as was seen before, the Fourier transform of the nth derivative of the δ function is $(-ip)^n$, $n\geq 0$; for this function the inverse Fourier integral obviously does not converge in the ordinary sense.

One possibility to deal with functions that behave like a polynomial for $|x|\to\infty$ has been elaborated by Bochner.[21] In taking the Fourier transform of $f(x)$, the function is divided by $1+|x|^k$, where k is larger than the order in which $f(x)$ tends to infinity. This division introduces an "error" which is an additive polynomial in p, so that $\int_{-\infty}^{+\infty}f(x)e^{ipx}dx$ is finally defined modulo polynomials.

Schwartz's Fourier transforms of "tempered distributions" (which include functions behaving like polynomials at infinity) are defined by means of Parseval's formula. In contrast we define the transform by splitting the integration from $-\infty$ to $+\infty$ into two parts, from $-\infty$ to 0 and from 0 to $+\infty$, which gives us a pair of holomorphic functions in the upper and lower half-planes. This definition is equivalent with the usual one in the sense that we obtain the "analytic continuation" (as defined previously) of the ordinary Fourier transform (Compare also Carleman[5]). (We prove this for square integrable functions. The question of equivalence with Schwartz's notion for distributions will be discussed in a further paper.)

Space of Rapidly Decreasing (C^∞) Functions

Let $(\mathcal{S})$ be the space of all (C^∞) functions that vanish faster than any power of x for $|x|\to\infty$ ("rapidly decreasing functions"). Convergence in $(\mathcal{S})$ is defined as follows: A sequence of functions $\varphi_j\in(\mathcal{S})$ is said to converge to 0 in $(\mathcal{S})$ if, and only if ,the following is true: Let P be an arbitrary polynomial, and let Q be an

[21] S. Bochner, *Vorlesungen über Fouriersche Integrale* (Leipzig, 1932).

arbitrary "polynomial of derivation," that is

$$Q=\sum_{\nu=0}^{N} a_\nu \delta^{(\nu)}.$$

Then $P(Q*\varphi_j)$ converges to zero uniformly on the whole real axis.

Tempered Distributions

The elements in the dual space $(\mathcal{S}')$ are called "tempered distributions." Schwartz[1] has shown: T is a tempered distribution exactly if $T\in(\mathcal{E}')$, and $T(x)$ behaves like a finite power $|x|^k$ for $|z|\to\infty$.

Theorem 9. *Let $T\in(\mathcal{S}')$. Let the support of T be contained in some "half axis" $a<x<+\infty$, $a\neq-\infty$. Then $T\cdot e^{ipx}$ exists and is a holomorphic function of p for Im $p>0$.*

Proof. Let $\alpha(x)$ be a (C^∞) function equal to 1 for $a\leq x$, and $\alpha(x)\equiv 0$ for $x\leq a-\epsilon$, $\epsilon>0$. Let $p=p'+ip''$. Then, for $p''>0$, $e^{ipx}=e^{ip'x}e^{-p''x}$ vanishes exponentially for $x\to\infty$, consequently $\alpha(x)e^{ipx}\in(\mathcal{S})$ for $\mathrm{Ipm}>0$. Consequently $T\cdot\alpha(x)e^{ipx}=T\cdot e^{ipx}$ exists. That this function is holomorphic for $\mathrm{Im}p>0$ follows as in the proof of theorem 1 by differentiating with respect to p. Analogously one obtains:

If $T\in(\mathcal{S}')$ and has support in $-\infty<x<b$, $b\neq 0$, then the Fourier transform is holomorphic for $Imp<0$.

For the general case of noncompact support, there is usually no domain in the complex p plane for which the whole integral converges. We therefore split the integral into two parts. This may seem to be a somewhat arbitrary procedure. The deeper justification for it, as we will see, lies in the fact that the Fourier transform of the kernel of the Cauchy integral involves a step function.

Definition 10. *Let $T\in(\mathcal{S}')$, and let T have at the origin at most a pole like singularity. Then we define:*

$$\tilde{T}(p)=\int_{-\infty}^{+\infty} T(x)\Theta(x)e^{ipx}dx+\tfrac{1}{2}Q_x\cdot e^{ipx} \quad \textit{for } Imp>0,$$

$$\tilde{T}(p)=\int_{-\infty}^{+\infty} T(x)\Theta(-x)e^{ipx}dx+\tfrac{1}{2}Q_x\cdot e^{ipx} \quad \textit{for } Imp<0,$$

where Q is the distribution with support at the origin defined in Sec. VII.

For $\int_{-\infty}^{+\infty}T(x)\Theta(x)e^{ipx}dx$, we will also write $\int_0^\infty T(x)e^{ipx}dx$, and correspondingly for $\int_{-\infty}^{+\infty}T(x)\times\Theta(-x)e^{ipx}dx$, we write $\int_{-\infty}^{0}T(x)e^{ipx}dx$.

From the preceding theorem follows that $\tilde{T}(p)$ is holomorphic in the complex p plane except on the real axis. ($Q_x\cdot e^{ipx}$ is a polynomial in p.)

We will call $\tilde{T}(p)$ the Fourier transform of T, and we will also write: $\tilde{T}(p)=\mathfrak{F}(T,p)$. Thus the Fourier transform of a distribution is again given by a pair of holomorphic functions, which may or may not correspond to a distribution.

Theorem 10. *(Connection between our definition and the ordinary definition of Fourier transforms): Let $T(x)$ be a square integrable function. Let the Fourier transform of T in the ordinary sense be denoted by $F(T,p)$. Then the analytic continuation of $F(T,p)$ (as defined in section VII) equals $\tilde{T}(p)$:*

$$\mathbf{F}^1(T,p)\equiv\tilde{T}(p) \quad \textit{for all } p \textit{ with } Imp=0.$$

For the proof we need the Fourier transform of the kernel of the Cauchy integral, and we therefore first prove the following lemma:

Lemma 6. *(Fourier transform of the kernel of the Cauchy integral.) The following formulas hold:*

$$\textit{for } Imp'>0: \quad \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{1}{p-p'}e^{ipx}dp=e^{ip'x}\Theta(x)$$

$$\textit{for } Imp'<0: \quad \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{1}{p-p'}e^{ipx}dp=-e^{ip'x}\Theta(-x).$$

Proof. Let $\mathrm{Im}p'>0$, then $e^{ip'x}\Theta(x)$ is a square integrable function, hence, by Plancherel's theorem, its Fourier transform exists and is square integrable, and we have

$$\frac{1}{2\pi}\mathfrak{F}\mathfrak{F}^{\mathrm{inv}}[e^{ip'x}\Theta(x)]=e^{ip'x}\Theta(x).$$

Now

$$\mathfrak{F}^{\mathrm{inv}}[e^{ip'x}\Theta(x)]=\int_{-\infty}^{+\infty}e^{-i(p-p')x}\Theta(x)dx$$

$$=\lim_{R\to\infty}\frac{e^{-i(p-p')R}-1}{-i(p-p')}=\frac{1}{i(p-p')}.$$

Hence

$$\mathfrak{F}\left[\frac{1}{2\pi i(p-p')},\,x\right]=e^{ip'x}\Theta(x).$$

Analogously one proves the second formula.

Proof of the theorem. $F(T,p)$ exists and is a square integrable function by Plancherel's theorem. Being square integrable, F has an analytic continuation $\mathbf{F}^1$. We have by definition

$$\mathbf{F}^0(T,p')=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{1}{p-p'}F(T,p)dp$$

$$=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\frac{1}{p-p'}\left\{\int_{-\infty}^{+\infty}T(x)e^{ipx}dx\right\}dp.$$

Since all the functions involved are square integrable, and since $\int_{-R}^{R}T(x)e^{ipx}dx$ converges for $R\to\infty$ to $F(T,p)$ in L^2 norm and the same is true for $\int_{-R}^{R}[1/(p-p')]e^{ipx}dp$, we can exchange the order of integration

and write

$$\mathbf{F}^1(T,p')$$

$$=\frac{1}{2\pi i}\int_{-\infty}^{+\infty}T(x)\left\{\int_{-\infty}^{+\infty}\frac{1}{p-p'}e^{ipx}dp\right\}dx$$

$$=\int_{-\infty}^{+\infty}T(x)\Theta(x)e^{ip'x}dx=\int_0^{\infty}T(x)e^{ip'x}dx$$

for $\mathrm{Im}p'>0$,

$$=-\int_{-\infty}^{+\infty}T(x)\Theta(-x)e^{ip'x}dx=-\int_{-\infty}^{0}T(x)e^{ip'x}dx$$

for $\mathrm{Im}p'<0$.

Hence

$$\mathbf{F}^1(T,p')=\mathrm{sgn}(\mathrm{Im}p')\mathbf{F}^0(T,p')=\tilde{T}(p'),$$

which proves our theorem.

Examples.

1. $T(x)\equiv 1$

$$\int_0^{\infty}e^{ipx}dx=\lim_{R\to\infty}\frac{e^{ipR}-1}{ip}=-\frac{1}{ip},\quad \mathrm{Im}p>0$$

$$\int_{-\infty}^{0}e^{ipx}dx=\lim_{R\to\infty}\frac{1-e^{-ipR}}{ip}=\frac{1}{ip},\quad \mathrm{Im}p<0.$$

Since $T(x)$ is continuous at the origin, the distribution Q vanishes. Thus we obtain

$$\tilde{T}(p)=\mp 1/ip \quad \text{for } \mathrm{Im}p\gtrless 0.$$

The right-hand side is equal to $2\pi\delta^1(p)$, the analytic continuation of $2\pi\delta(p)$. This result is consistent with the usual theory, in which one has

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{ipx}dx=\delta(p).$$

2. $T(x)=x^n$, n integer, $n\geq 0$.

Since $T(x)$ is continuous, Q vanishes. We obtain at once

$$\int_0^{\infty}x^ne^{ipx}dx=\lim_{R\to\infty}\left(x^n\frac{e^{ipx}}{ip}-nx^{n-1}\frac{e^{ipx}}{(ip)^2}+\cdots+n!(-1)\frac{^{n-1}e^{ipx}}{(ip)^{n+1}}\right)\Bigg|_0^R=\frac{n!}{(-ip)^{n+1}},\quad \text{for } \mathrm{Im}p>0,$$

and analogously

$$\int_{-\infty}^{0}x^ne^{ipx}dx=-\frac{n!}{(-ip)^{n+1}},\quad \text{for } \mathrm{Im}p<0.$$

Thus

$$\tilde{T}(p)=\mathrm{sgn}(\mathrm{Im}p)\frac{(-1)^{n+1}n!}{(ip)^{n+1}}.$$

On the other hand, we have

$$\delta^{1(n)}(p)=\mathrm{sgn}(\mathrm{Im}p)\frac{(-1)^{n+1}n!}{2\pi ip^{n+1}}.$$

Thus

$$\tilde{T}(p)=[2\pi/(i)^n]\delta^{1(n)}(p).$$

We can combine the factor $(i)^n$ together with x^n to obtain

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}(ix)^ne^{ipx}dx=\delta^{1(n)}(p).$$

Theorem 11. *Let T be as in definition* 10. *Then*

$$\frac{d}{dp}\tilde{T}(p)=\int_0^{\infty}(ix)T(x)e^{ipx}dx+\tfrac{1}{2}Q_x\cdot(ix)e^{ipx}\quad \textit{for } Imp>0$$

and

$$\frac{d}{dp}\tilde{T}(p)=\int_{-\infty}^{0}(ix)T(x)e^{ipx}dx+\tfrac{1}{2}Q_x\cdot(ix)e^{ipx}\quad \textit{for } Imp<0.$$

This is shown as in theorem 1.

Inverse Fourier Transform

Let $f(x)$ be a square integrable function; then according to Plancherel's theorem

$$\mathfrak{F}(f,p)=\int_{-\infty}^{+\infty}f(x)e^{ipx}dx$$

exists, and $\mathfrak{F}(f,p)$ is square integrable. Therefore

$$\mathfrak{F}^{\mathrm{inv}}[\mathfrak{F}(f,p),x]=\int_{-\infty}^{+\infty}\mathfrak{F}(f,p)e^{-ipx}dx$$

exists, and we have

$$(1/2\pi)\,\mathfrak{F}^{\mathrm{inv}}[\mathfrak{F}(f,p),x]=f(x).$$

As we have seen, the Fourier transform of $\delta^{(n)}(x)$ is $(-ip)^n$. While this is not a square integrable function, and the inverse Fourier transform does not exist in the ordinary sense, our extended definition of a Fourier transform gives the correct inverse in the sense of giving the correct analytic continuation of $\delta^{(n)}$ in the complex x plane:

$$\mathfrak{F}(\delta^{(n)},p)=\delta^{(n)}\cdot e^{ipx}=(-ip)^n,$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}(-ip)^ne^{-ipx}dx\to\tilde{T}^{\mathrm{inv}}[(-ip)^n,x]=\delta^{1(n)}(x).$$

We will investigate this situation in more generaltiy.

Theorem 12. *Let T be a distribution with compact support. Let $\mathfrak{F}(T,p)\equiv\tilde{T}(p)$ be bounded by a polynomial for $p\to\mp\infty$, p real. Then the inverse Fourier transform*

$$\tilde{T}^{\text{inv}}(z)=\int_0^\infty \tilde{T}(p)e^{-ipz}dp \quad \textit{for } Imz<0,$$

$$=\int_{-\infty}^0 \tilde{T}(p)e^{-ipz}dp \quad \textit{for } Imz>0,$$

exists, and

$$\frac{1}{2\pi}\tilde{T}^{\text{inv}}(z)=\mathbf{T}^1(z),$$

where $\mathbf{T}^1(z)$ is the analytic continuation of T.

Proof. Since T has compact support, $\tilde{T}(p)$ exists and is holomorphic for all p, $|p|<\infty$, and has in particular no singularities at the origin. Consequently, no distribution corresponding to Q has to be included in the definition of $\tilde{T}^{\text{inv}}(z)$. The existence of $\tilde{T}^{\text{inv}}(p)$ follows immediately from the assumption about the behavior of $\tilde{T}(p)$.

For $\text{Im}z<0$, we have

$$\tilde{T}^{\text{inv}}(z)=\int_0^\infty\left\{\int_{-\infty}^\infty T(x)e^{ipx}dx\right\}e^{-ipz}dp$$

$$=\int_0^\infty\left\{\int_{C_0}\mathbf{T}^0(\xi)e^{ip\xi}d\xi\right\}e^{-ipz}dp,$$

where the closed contour C_0 around the support of T is chosen such that $|\text{Im}\xi|<|\text{Im}z|$ for all ξ on C_0. Then $\text{Im}(\xi-z)>0$ for all ξ on C_0. We have

$$\tilde{T}^{\text{inv}}(z)=\lim_{R\to\infty}\int_0^R\left\{\int_{C_0}\mathbf{T}^0(\xi)e^{ip\xi}d\xi\right\}e^{-ipz}dp$$

$$=\lim_{R\to\infty}\int_{C_0}\mathbf{T}^0(\xi)\left\{\int_0^R e^{ip(\xi-z)}dp\right\}d\xi.$$

Now

$$\int_0^R e^{ip(\xi-z)}dp=\frac{e^{iR(\xi-z)}-1}{i(\xi-z)}$$

converges uniformly to $-[1/i(\xi-z)]$ for $\xi\in C_0$. Hence we can interchange limit and intergration and we obtain

$$\tilde{T}^{\text{inv}}(z)=i\int_{C_0}\mathbf{T}^0(\xi)\frac{1}{\xi-z}d\xi=i\int_{-\infty}^{+\infty}T(x)\frac{1}{x-z}dx$$

$$=-2\pi\mathbf{T}^0(z) \quad \text{for } \text{Im}z<0.$$

In an analogous manner one obtains

$$\tilde{T}^{\text{inv}}(z)=2\pi\mathbf{T}^0(z) \quad \text{for } \text{Im}z>0.$$

Hence $\tilde{T}^{\text{inv}}(z)=2\pi\mathbf{T}^1(z)$. This proves our theorem.

Fourier Transforms of $\Theta(x)$, $\delta_+(x)$, $\varepsilon(x)$, and $P(x^{-n})$.

1. $$\int_{-\infty}^{+\infty}\Theta(x)e^{ipx}dx=-(1/ip) \quad \text{for } \text{Im}p>0,$$
$$=0 \quad \text{for } \text{Im}p<0.$$

The right-hand side equals $\pi\delta_+{}^0(p)$ (compare Sec. V). Hence

$$\mathfrak{F}[\Theta(x),p]=\pi\delta_+(p).$$

2. $$\mathfrak{F}[\delta_+(x),p]=-\frac{1}{i\pi}\lim_{\epsilon\to 0+}\int_{-\infty}^{+\infty}\frac{1}{x+i\epsilon}e^{ipx}dx.$$

According to lemma 6,

$$-\frac{1}{i\pi}\int_{-\infty}^{+\infty}\frac{1}{x+i\epsilon}e^{ipx}dx=2e^{\epsilon p}\Theta(-p).$$

The integral converges uniformly in any compact interval of the real p axis to $2\Theta(-p)$. Hence

$$\mathfrak{F}[\delta_+(x),p]=2\Theta(-p),$$

and

$$\mathfrak{F}^{\text{inv}}[\delta_+(x),p]=2\Theta(p).$$

3. According to examples 3 and 4 of Sec. V, we have

$$P(x^{-1})=i\pi[\delta(x)-\delta_+(x)].$$

Hence

$$\mathfrak{F}[P(x^{-1}),p]=i\pi[1-2\Theta(-p)]=i\pi\epsilon(p),$$

where

$$\epsilon(p)=\begin{cases}1 & \text{for } p>0\\ -1 & \text{for } p<0.\end{cases}$$

4. Analogously one computes:

$$\mathfrak{F}[\epsilon(p),x]=2iP(x^{-1}).$$

5. $$P(x^{-n})=\frac{(-1)^{n-1}}{(n-1)!}\frac{d^{n-1}}{dx^{n-1}}P(x^{-1}).$$

Hence

$$\mathfrak{F}[P(x^{-n}),p]=\frac{(-1)^{n-1}}{(n-1)!}\mathfrak{F}\left(\frac{d^{n-1}}{dx^{n-1}}P(x^{-1}),p\right)$$

$$=\frac{(ip)^{n-1}}{(n-1)!}\mathfrak{F}[P(x^{-1}),p].$$

Hence

$$\mathfrak{F}[P(x^{-n}),p]=\frac{i\pi\epsilon(p)(ip)^{n-1}}{(n-1)!}.$$

Fourier Transforms and Convolutions

Let f, g be square integrable functions; then

$$(1/2\pi)\mathfrak{F}^{\text{inv}}[\mathfrak{F}(f)\cdot\mathfrak{F}(g)]=f*g,$$

where $f*g$ is the convolution of f and g. As this is one of the most important formulas in the theory of Fourier transformations, we would like to extend it to distributions.

Theorem 13. *Let $S\in(\mathcal{E}')$ and $T\in(\mathcal{E}')$, and let $\mathfrak{F}(S,p)$ and $\mathfrak{F}(T,p)$ be bounded by polynomials for $p\to\pm\infty$.* Then:

(a)

$$(S*^1T)=\frac{1}{2\pi}\mathfrak{F}^{\text{inv}}[\mathfrak{F}(S,p)\cdot\mathfrak{F}(T,p)]$$

$$=\frac{1}{2\pi}\int_0^\infty \mathfrak{F}(S,p)\mathfrak{F}(T,p)e^{-ipz}dp \quad \textit{for } Imz<0,$$

$$=\frac{1}{2\pi}\int_{-\infty}^0 \mathfrak{F}(S,p)\mathfrak{F}(T,p)e^{-ipz}dp \quad \textit{for } Imz>0.$$

Proof. T has compact support. Therefore, $(S*^1T)$ exists,

$$(S*^1T)(z)=\int_{-\infty}^{+\infty}\mathbf{S}^1(z-x)T(x)dx=\int_a^b\mathbf{S}^1(z-x)T(x)dx$$

for some a, b finite (because T has compact support). From theorem 12,

$$\mathbf{S}^1(z-x)=(2\pi)^{-1}\mathfrak{F}^{\text{inv}}[\mathfrak{F}(S,p),z-x]$$

$$=\begin{cases}\dfrac{1}{2\pi}\displaystyle\int_0^\infty \mathfrak{F}(S,p)e^{-ip(z-x)}dp, & \operatorname{Im}(z-x)<0\\[2ex] \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^0 \mathfrak{F}(S,p)e^{-ip(z-x)}dp, & \operatorname{Im}(z-x)>0.\end{cases}$$

We take x real, $\operatorname{Im}z<0$. But $\int_0^R\mathfrak{F}(S,p)e^{-ip(z-x)}dp$ converges uniformly in x for $R\to\infty$, x in any compact interval of the x axis, z fixed with $\operatorname{Im}z<0$. Therefore, when substituting this expression for $\mathbf{S}^1$, we may interchange the order of integration and obtain

$$(S*^1T)(z)=\int_a^b\mathbf{S}^1(z-x)T(x)dx$$

$$=\frac{1}{2\pi}\int_a^b\left\{\int_0^\infty\mathfrak{F}(S,p)e^{-ip(z-x)}dp\right\}T(x)dx$$

$$=\frac{1}{2\pi}\int_0^\infty\mathfrak{F}(S,p)e^{-ipz}\left\{\int_{-\infty}^\infty e^{ipx}T(x)dx\right\}dp$$

$$=\frac{1}{2\pi}\int_0^\infty\mathfrak{F}(S,p)\mathfrak{F}(T,p)e^{-ipz}dp, \quad \operatorname{Im}z<0.$$

A similar argument for $\operatorname{Im}z>0$ completes the proof.

(b) *We establish by analogous methods under the same conditions*

$$\mathfrak{F}(S*T,\,p)=\mathfrak{F}(S,p)\mathfrak{F}(T,p).$$

Theorem 13 can be generalized to cases with less stringent conditions on the distributions S and T.

Example.

$$S\in(\mathcal{E}'), \quad T(x)=\sum_{\nu=0}^N a_\nu\delta^\nu(x),$$

then

$$\mathfrak{F}(S*T,\,p)=\mathfrak{F}(S,p)\Big(\sum_{\nu=0}^N a_\nu(ip)^\nu\Big).$$

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Dr. J. R. Oppenheimer and the Institute for Advanced Study for their hospitality and support during the course of this work. One of us (L.D.) would like also to thank the National Science Foundation for a Postdoctoral Fellowship.

On the Capacity of the Icosahedron

JAMES CONLAN, J. B. DIAZ,* AND W. E. PARR
U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland

An application of Dirichlet's principle, using a simple trial function suggested by the symmetries of the problem, is shown to furnish fairly close, readily computable, upper bounds for the capacity of any regular solid.

1. INTRODUCTION

THE main purpose of the present note is to show how an application of Dirichlet's principle (see, e.g., Pólya and Szegö,[1] p. 43) using a simple trial function [see Eq. (5)] furnishes fairly close, readily computable, upper bounds for the capacity of a regular solid. While this function seems to be naturally dictated by the symmetry of the domain in question, and may be used *a priori* independently of the variational considerations of Sec. 2, the Euler-Lagrange argument given there suggests that (in a certain sense) it is the "best" trial function of such a simple nature. A similar trial function has been used in the estimation of the torsional rigidity by Diaz and Weinstein.[2] In Sec. 3, an attractive attempt at improving the bounds for the capacity is analyzed. This attempt was developed in seeking to improve the numerical bounds obtained in Sec. 2. It is shown that a certain minimization process [which leads to the inequalities (13)] actually furnishes worse upper and lower bounds than the simple algebraic process leading to the inequality (16). This byproduct of the computation of numerical bounds for the capacity in Sec. 2 is believed to be of interest in itself, since the remark made in Sec. 3 applies equally well to many quadratic functionals in mathematical physics, and not only to the particular one under consideration here.

2. UPPER BOUNDS FOR THE CAPACITY

For any two sufficiently well-behaved functions $f(x,y,z)$ and $g(x,y,z)$, define the inner product (f,g) by

$$(f,g)=\iiint_D \text{grad}f\cdot\text{grad}g\,dV, \tag{1}$$

where D is the region exterior to a given closed bounded smooth surface S. An upper bound for the capacity C of S is given by

$$(1/4\pi)(w,w)\geq C, \tag{2}$$

where $w(x,y,z)$ is a sufficiently smooth function such that $w(x,y,z)=1$ if (x,y,z) is a point of S, and $w=O(r^{-1})$ as $r=(x^2+y^2+z^2)^{\frac{1}{2}}$ approaches infinity.

Let P be a regular polyhedron of n faces, each face being a regular polygon of r edges. Suppose that P is circumscribed about the unit sphere with center at the origin. Let P have one face F contained in the plane $x=1$, with the x axis passing through the center of F; and let one edge, call it E, of F be parallel to the y axis. This serves to fix the position of P in space. Now consider the triangle T, with E as one edge and the point (1,0,0) as opposite vertex. If F is a regular polygon of r edges, then F will consist of r triangles congruent to T. Let the vertices of T be (1,0,0), $(1,a,b)$, and $(1,-a,b)$, and let $\alpha=a/b$. Then

$$(w,w)=\iiint_D |\text{grad}w|^2dV$$
$$=nr\int_1^\infty dx\int_0^{bx}dz\int_{-\alpha z}^{\alpha z}|\text{grad}w|^2dy, \tag{3}$$

where we assume w to be symmetric with respect to the polyhedron P, and also to be symmetric with respect to each of the r triangles in the face F.

If we let $w(x,y,z)=f(x)$ in that portion D' of D covered by the integration on the extreme right of (3), then

$$(w,w)=2nr\int_1^\infty dx\int_0^{bx}dz\int_0^{\alpha z}[f'(x)]^2dy$$
$$=nr\alpha b^2\int_1^\infty x^2[f'(x)]^2dx. \tag{4}$$

To minimize this expression for (w,w) (with respect to all suitable functions f) note that by the Euler-Lagrange equation

$$\frac{\partial}{\partial f}\{x^2[f'(x)]^2\}-\frac{d}{dx}\frac{\partial}{\partial f'}\{x^2[f'(x)]^2\}=0,$$

we obtain

$$2xf'(x)+x^2f''(x)=0;$$

and since $f(1)=1$ and $f(x)=O(x^{-1})$ for large x, we have

$$f(x)=1/x, \tag{5}$$

so that (5) gives the best possible choice of f.

An improved value for (w,w) can be obtained by letting

$$w(x,y,z)=(1/x)+(\lambda|y|+\mu z)(1/x^3-1/x^2) \tag{6}$$

* Also at the Institute for Fluid Dynamics and Applied Mathematics, University of Maryland.

[1] G. Pólya and G. Szegö, "Isoperimetric inequalities in mathematical physics," *Annals of Mathematics Studies*, No. 27 (Princeton University Press, Princeton, New Jersey, 1951).

[2] J. B. Diaz and A. Weinstein, Am. J. Math. **70**, 107 (1948).

in D' and minimizing (w,w) with respect to λ and μ. Note that the choice of w given by (6) still satisfies the boundary conditions on F and at ∞. In accordance with the well-known Rayleigh-Ritz method, the first term on the right-hand side of (6) is just the function arrived at in (5), while the second term is just a simple "coordinate function" (in the terminology of Walther Ritz) which satisfies the homogeneous boundary condition (i.e., it has value zero on the boundary). From (6) we have

$$
\begin{aligned}
(w,w) &= 2\pi r\int_1^\infty dx\int_0^{bx} dz\int_0^{\alpha z} |\operatorname{grad} w|^2 dy \\
&= 2\pi r\int_1^\infty dx\int_0^{bx} dz\int_0^{\alpha z}\left\{\frac{1}{x^4}+\lambda\left[-\frac{2y}{x^2}\left(-\frac{3}{x^4}+\frac{2}{x^3}\right)\right]\right. \\
&\quad +\mu\left[-\frac{2z}{x^2}\left(-\frac{3}{x^4}+\frac{2}{x^3}\right)\right]+\lambda^2\left[\left(-\frac{3}{x^4}+\frac{2}{x^3}\right)^2 y^2\right. \\
&\quad \left.+\left(\frac{1}{x^3}-\frac{1}{x^2}\right)^2\right]+\mu^2\left[z^2\left(-\frac{3}{x^4}+\frac{2}{x^3}\right)^2+\left(\frac{1}{x^3}-\frac{1}{x^2}\right)\right] \\
&\quad \left.+2\lambda\mu yz\left(-\frac{3}{x^4}+\frac{2}{x^3}\right)^2\right\} dy \\
&= 2\pi r\alpha b^2\left\{\frac{1}{2}-\lambda\frac{\alpha b}{6}-\mu\frac{b}{3}+\lambda^2\left(\frac{\alpha^2 b^2}{12}+\frac{1}{6}\right)\right. \\
&\quad \left.+\mu^2\left(\frac{b^2}{4}+\frac{1}{6}\right)+\lambda\mu\frac{\alpha b^2}{4}\right\}. \qquad (7)
\end{aligned}
$$

For (w,w) to be a relative minimum (as a function of λ and μ), it is necessary that

$$
\begin{aligned}
\frac{\partial(w,w)}{\partial\lambda} &= 2\pi r\alpha b^2\left\{-\frac{\alpha b}{6}+2\lambda\left(\frac{\alpha^2 b^2}{12}+\frac{1}{6}\right)+\mu\frac{\alpha b^2}{4}\right\}=0, \\
\frac{\partial(w,w)}{\partial\mu} &= 2\pi r\alpha b^2\left\{-\frac{b}{3}+2\mu\left(\frac{b^2}{4}+\frac{1}{6}\right)+\lambda\frac{\alpha b^2}{4}\right\}=0,
\end{aligned} \qquad (8)
$$

which is a system of linear equations to be solved for λ and μ.

On applying the foregoing to the icosahedron, we have

$$n=20,\quad r=3,\quad b=\sqrt{3}\{\tfrac{1}{2}(\sqrt{5}+1)\}^{-2},\quad \alpha=1/\sqrt{3}$$

(see Coxeter[3]), and this gives $C\leq(1/4\pi)(w,w)\leq 1.096$. Use of the volume radius (see Pólya-Szegö,[1] p. 63) gives the lower bound $1.064\leq C$.

For the case of the cube we have $n=6$, $r=4$, $b=1$, $\alpha=1$, and so $C\leq 1.6103$. Use of the volume radius gives the lower bound $1.240\leq C$.

[3] H. S. M. Coxeter, *Regular Polytopes* (Methuen and Company, Ltd., London, 1948).

Parr,[4] by using an extension of the results of Pólya-Szegö obtained the upper bound $C\leq 1.084$ for the icosahedron, and $C\leq 1.3359$ for the cube; Payne and Weinberger[5] obtained, by still another method, the upper bound $C\leq 1.336$ for the cube.

3. A MIRAGE

In this section we discuss an attempt to derive new bounds for the capacity of a surface S. We introduce the functions u, v, w (functions of x, y, and z) where

$$u=1,\quad \text{on } S,\quad \Delta u=0 \text{ in } D,\quad u=O(1/r) \text{ as } r \text{ approaches } \infty;$$

$$\Delta v=0 \text{ in } D,\quad v=O(1/r) \text{ as } r \text{ approaches } \infty;\quad v \text{ not identically zero};$$

$$w=1 \text{ on } S,\quad w=O(1/r) \text{ as } r \text{ approaches } \infty.$$

Here Δ denotes the Laplacian operator $(\partial^2/\partial x^2)+(\partial^2/\partial y^2)+(\partial^2/\partial z^2)$.

The function w satisfies the same boundary conditions as u, while the function v satisfies the same partial differential equation as u. It is well known (see, e.g., Diaz[6]) that an upper bound for the capacity C may be obtained in terms of w alone, as follows:

$$C\leq(1/4\pi)(w,w); \qquad (9)$$

while a lower bound for C may be obtained in terms of v alone, as follows:

$$\frac{1}{4\pi}\frac{(v,u)^2}{(v,v)}\leq C, \qquad (10)$$

where, by Green's identity

$$(v,u)=\int\!\!\int_S \frac{\partial v}{\partial n}dS, \qquad (11)$$

and n is the inner unit normal to S.

The following process, which employs the functions w and v together, rather than singly, would appear, at first glance, to furnish better upper and lower bounds [see Eq. (13)] for the capacity C than are given by w and v, individually, in Eqs. (9) and (10), respectively.

By Schwarz's inequality, we have

$$(w,\, u-v)^2\leq(w,w)(u-v,\, u-v)$$

[4] W. E. Parr, "Upper and lower bounds for the capacitance of the regular solids," Ph.D. thesis, University of Maryland, June, 1960.

[5] L. E. Payne and H. F. Weinberger, J. Math. and Phys. **33**, 291 (1955).

[6] J. B. Diaz, in *Boundary Value Problems in Differential Equations*, Proceedings of a Symposium conducted by the Mathematics Research Center of the U. S. Army at the University of Wisconsin, Madison, Wisconsin, April 20–22 (1959), pp. 47–83.

or

$$(w,u)^2-2(w,u)(w,v)+(w,v)^2 \leq (w,w)[(u,u)-2(u,v)+(v,v)]. \quad (12)$$

But

$$(w,u)=\iiint_D \text{grad}w\cdot\text{grad}u dV,$$

and by Green's identity this is just

$$-\iiint_D w\Delta u dV+\iint_S w\frac{\partial u}{\partial n}dS=\iint_S \frac{\partial u}{\partial n}dS=(u,u).$$

Thus

$$(w,u)=(u,u).$$

On using the fact that $\Delta v=0$ in D and $u-w=0$ on S, the preceding type of reasoning shows that

$$(v,\ u-w)=0,$$

and hence

$$(v,u)=(v,w).$$

Thus, replacing (w,u) by (u,u) and (v,u) by (v,w) in (12), and transposing, we obtain

$$(u,u)^2-(u,u)[(w,w)+2(w,v)] \leq -(w,v)^2-2(w,w)(v,w)+(w,w)(v,v);$$

or, completing the square on the left-hand side,

$$\{(u,u)-[\tfrac{1}{2}(w,w)+(w,v)]\}^2 \leq (w,w)\{(v,v)-(v,w)+\tfrac{1}{4}(w,w)\}.$$

Since λv satisfies the same conditions as v, where λ is any real number, we may replace v by λv throughout, If we replace v by λv and minimize the right-hand side with respect to λ (in order to minimize the "error"), we obtain

$$\left\{(u,u)-\frac{1}{2}\left[(w,w)+\frac{(v,w)^2}{(v,v)}\right]\right\} \leq \frac{(w,w)}{4}\left\{(w,w)-\frac{(v,w)^2}{(v,v)}\right\}. \quad (13)$$

Equation (13) furnishes the presumably better bounds for C obtainable by employing the functions w and v together. However, from (9) and (10), in a readily understandable notation (using C_- to denote a lower bound for C, for example),

$$4\pi C_-\equiv\frac{(v,w)^2}{(v,v)}=\frac{(v,u)^2}{(v,v)}\leq(u,u)=4\pi C\leq(w,w)\equiv 4\pi C^-.$$

Then (13) can be rewritten

$$\{C-\tfrac{1}{2}[C^-+C_-]\}^2\leq\tfrac{1}{4}C^-[C^--C_-], \quad (14)$$

where

$$0\leq C_-\leq C\leq C^-. \quad (15)$$

However, upon subtracting $\frac{1}{2}(C_-+C^-)$, the inequality (15) yields

$$\tfrac{1}{2}C_--\tfrac{1}{2}C^-\leq C-\tfrac{1}{2}(C_-+C^-)\leq\tfrac{1}{2}C^--\tfrac{1}{2}C_-;$$

that is,

$$-\tfrac{1}{2}(C^--C_-)\leq C-\tfrac{1}{2}(C_-+C^-)\leq\tfrac{1}{2}(C^--C_-)$$

or

$$\{C-\tfrac{1}{2}(C_-+C^-)\}^2\leq\tfrac{1}{4}(C^--C_-)^2. \quad (16)$$

Since $C^-\geq C^--C_-$, the inequality (16) is actually sharper than (14). Thus, the process of minimizing with respect to λ, which led to (13), actually furnishes worse bounds, (13), than the bounds (16), which were obtained purely algebraically, without any minimization whatever!

Lower Bounds and Isoperimetric Inequalities for Eigenvalues of the Schrödinger Equation*

JOSEPH B. KELLER
Institute of Mathematical Sciences, New York University, New York

(Received September 15, 1960)

The potential which minimizes the lowest eigenvalue of the one-dimensional Schrödinger equation is determined among all potentials V for which the integral of V^n has the prescribed value k. For each value of n and k this potential is found to be a special case of the Epstein-Eckart potentials which were originally introduced because the Schrödinger equation for them could be solved explicitly. The minimum eigenvalue is determined and it provides a lower bound on the lowest eigenvalue of any potential for which $\int V^n dx = k$. The expression of this fact as an inequality yields an isoperimetric inequality. For an arbitrary potential, each value of n provides one lower bound on the lowest eigenvalue, the largest of which is the best. This best bound is determined for the square well, the exponential, and the inverse power potentials. In the case of the square well, it is compared with the exact value. In the limiting case $n=1$ our result reduces to that previously obtained by Larry Spruch, who showed that the delta function has the minimum lowest eigenvalue among all potentials of given "area."

1. INTRODUCTION

UPPER bounds on the lowest eigenvalue of the Schrödinger equation can be obtained easily because this eigenvalue is the minimum of a certain variational expression. However, it is not so easy to obtain lower bounds, although various methods have been devised for obtaining them. Therefore, we have re-examined the problem of obtaining lower bounds from a different viewpoint, i.e., that of isoperimetric inequalities. We seek that potential which in a specified class of potentials, yields the minimum lowest eigenvalue. Once we find it, its lowest eigenvalue is a lower bound on the lowest eigenvalue of all the potentials in the specified class. The resulting inequality is called an isoperimetric inequality by analogy with the classical inequality $A \leqq L^2/4\pi$ relating the length of a curve to the area A it encloses. This classical isoperimetric inequality is a consequence of the fact that of all closed curves of length L, the circle encloses the greatest area $L^2/4\pi$.

Our analysis is confined to the one-dimensional case. We consider a two-parameter family of classes of potentials and, therefore, we obtain a one-parameter family of isoperimetric inequalities. Thus, we obtain a one parameter family of lower bounds on the lowest eigenvalue of a given potential. These bounds are explicit formulas, each merely involving an integral of some power of the potential. Naturally, the largest of the lower bounds is the best, but which is largest depends upon the potential. To illustrate the accuracy of the bounds, we determine the best one for a square well and compare it with the exact eigenvalue. We also obtain the best lower bound for exponential and inverse power potentials. In principle, our method applies to higher-dimensional cases, but it then leads to nonlinear differential equations which cannot be solved explicitly, whereas they can be solved explicitly in the one-dimensional case. These equations are given and some consequences of them are presented.

One interesting aspect of our results is that the potentials which yield the minimum lowest eigenvalues, in the classes we have considered, are special cases of the potentials introduced by Epstein. Their potentials were introduced because they led to Schrödinger equations which could be solved explicitly in terms of known functions.

The present investigation was undertaken to generalize the result, proved by Larry Spruch (unpublished), that the delta function has the smallest lowest eigenvalue of all potentials of given "area," i.e., of given integral of the magnitude of the potential. His result appears as a limiting case of our results. Our method of analysis is one which was devised previously to determine the shape of the strongest column of given length and volume.[2,3] In the course of the analysis, we also make use of a suggestion of H. F. Weinberger. In the final section we show by the same method that the usual upper bound for the lowest eigenvalue also results from an isoperimetric inequality.

2. ISOPERIMETRIC PROBLEM

The one-dimensional Schrödinger equation for the wave function $u(x)$ of a particle of energy λ in a potential $-V(x)$ is, in appropriate units,

$$u_{xx} + V(x)u + \lambda u = 0. \tag{1}$$

This equation has a quadratically integrable solution if and only if λ has one of a discrete set of values called eigenvalues, which depend upon $V(x)$. We seek the potentials $V(x)$ which make stationary some eigenvalue

* The research in this document has been sponsored by the Office of Naval Research, Air Force Cambridge Research Laboratories, Office of Ordnance Research.

[1] P. S. Epstein, Proc. Nat. Acad. Sci. **16**, 627 (1930).
[2] J. B. Keller, Archive Ratl. Mech. and Anal. **5**, 275 (1960).
[3] I. Tadjbaksh and J. B. Keller, *Strongest Columns and Isoperimetric Inequalities for Eigenvalues*, J. Appl. Mech. (to be published).

λ of Eq. (1) among all potentials satisfying the condition

$$\int_{-\infty}^{\infty} V^n(x)dx=k. \tag{2}$$

Here n and k are two real constants which characterize the class of potentials under consideration.

Let us suppose that $V_0(x)$ is a solution of this problem and that $u_0(x)$ and λ_0 are the corresponding eigenfunction and stationary eigenvalue. We introduce a family of potentials $V(x,\epsilon)$ depending smoothly upon a parameter ϵ, satisfying Eq. (2), and such that $V(x,0)=V_0(x)$. Then the corresponding eigenfunction $u(x,\epsilon)$ can be so normalized that it, as well as the eigenvalue $\lambda(\epsilon)$, depends smoothly upon ϵ. If we denote differentiation with respect to ϵ by a dot, it follows that $\dot{\lambda}(0)=0$. We now differentiate Eq. (1) and (2) with respect to ϵ and obtain

$$\dot{u}_{xx}+V\dot{u}+\lambda\dot{u}=-\dot{\lambda}u-\dot{V}u \tag{3}$$

$$\int_{-\infty}^{\infty} V^{n-1}\dot{V}dx=0. \tag{4}$$

The inhomogeneous equation (3) has a quadratically integrable solution only if the right-hand side is orthogonal to u, the solution of Eq. (1). But since $u(x,\epsilon)$ is a smooth function of ϵ which is quadratically integrable, it follows that $\dot{u}$ exists and is also quadratically integrable. Therefore, the orthogonality condition is satisfied and it yields, when $\epsilon=0$,

$$\int_{-\infty}^{\infty} u_0^2\dot{V}dx=0. \tag{5}$$

The choice of $V(x,\epsilon)$ is arbitrary except that $V(x,0)=V_0(x)$, that it be a smooth function of ϵ and satisfy Eq. (2). Therefore, $\dot{V}$ is arbitrary except that it must satisfy Eq. (4). Thus Eq. (5) expresses the fact that u_0^2 is orthogonal to every function $\dot{V}$ which, by Eq. (4), is orthogonal to $V_0{}^{n-1}$. This implies that u_0^2 is a constant multiple of $V_0{}^{n-1}$. We shall choose the multiplier to be unity, since u can be multiplied by a constant factor and remain a solution of Eq. (1). Thus we have

$$u_0^2=V_0{}^{n-1}. \tag{6}$$

Let us now eliminate V_0 from Eq. (1) by means of Eq. (6), and obtain the following nonlinear equation for u_0:

$$u_{0xx}+u_0{}^{1+(2/n-1)}+\lambda_0u_0=0. \tag{7}$$

To solve Eq. (7) we multiply it by u_{0x} and integrate, obtaining

$$u_{0x}{}^2+\frac{(n-1)}{n}u_0{}^{2+(2/n-1)}+\lambda_0u_0{}^2=0. \tag{8}$$

The integration constant has been set equal to zero in Eq. (8), since u_0 and u_{0x} must vanish when x becomes infinite, in order that u_0 be an eigenfunction. From Eq. (8) we find

$$u_{0x}=(-\lambda_0)^{\frac{1}{2}}u_0\left(1+\frac{n-1}{n\lambda}u_0{}^{2/n-1}\right)^{\frac{1}{2}}. \tag{9}$$

To evaluate the integral which occurs in solving Eq. (9), it is convenient to replace u_0 by V_0 by means of Eq. (6). Then Eq. (9) becomes

$$V_{0x}=\frac{2(-\lambda)^{\frac{1}{2}}}{n-1}V_0\left(1+\frac{n-1}{n\lambda_0}V_0\right)^{\frac{1}{2}}. \tag{10}$$

The various solutions of Eq. (10) differ only by translations. We shall select that solution for which $V_0(0)=n/(n-1)$. Then Eq. (10) yields

$$\frac{2(-\lambda_0)^{\frac{1}{2}}}{n-1}x=\int_{n/n-1}^{V_0} V^{-1}\left(1+\frac{n-1}{n\lambda_0}V\right)^{-\frac{1}{2}}dV$$

$$=-2\tanh^{-1}\left(1+\frac{n-1}{n\lambda_0}V_0\right)^{\frac{1}{2}}. \tag{11}$$

Upon solving Eq. (11) for V_0, we obtain

$$V_0(x)=\frac{-n\lambda_0}{n-1}\operatorname{sech}^2\left[\frac{(-\lambda_0)^{\frac{1}{2}}x}{n-1}\right]. \tag{12}$$

From Eq. (12) we see that V_0 is a periodic function of x if $\lambda_0>0$, while $V_0\equiv0$, if $\lambda_0=0$. Since the integral in Eq. (2) would not exist, if V_0 were periodic, and no eigenvalues would exist, if $V_0\equiv0$, we conclude that $\lambda_0<0$. Then Eq. (12) shows that V_0 vanishes as $|x|$ becomes infinite. Since u_0 must also vanish at infinity, we see from Eq. (6) that $n>1$.

Now Eqs. (12) and (2) yield a relation among λ_0, n, and k which is

$$\left(\frac{-n\lambda_0}{n-1}\right)^n\frac{2(n-1)}{(-\lambda_0)^{\frac{1}{2}}}\int_0^{\infty}\operatorname{sech}^{2n}ydy=k. \tag{13}$$

Since $n>1$ and $\lambda_0>0$, we see from Eq. (13) that $k>0$. The integral in Eq. (13) has the value[4]

$$\int_0^{\infty}\operatorname{sech}^{2n}ydy=\frac{\Gamma(n)\Gamma(\frac{1}{2})}{2\Gamma(n+\frac{1}{2})}. \tag{14}$$

By using Eq. (14) in Eq. (13) and solving for λ_0, we obtain

$$-\lambda_0=-F(n)k^{2/2n-1}. \tag{15}$$

Here $F(n)$ is given by

$$F(n)=\left[\frac{\Gamma(n+\frac{1}{2})}{(n-1)\pi^{\frac{1}{2}}\Gamma(n)}\left(\frac{n-1}{n}\right)^n\right]^{2/2n-1}. \tag{16}$$

[4] W. Grobner and N. Hofreiter, *Integraltafel* (Springer-Verlag, Berlin, Germany, 1949), Chap. II, p. 162, Eq. (12).

We have now found that for any $k>0$ and any $n>1$ there is exactly one potential $V_0(x)$ given by Eq. (12), with λ_0 given by Eq. (15), which renders stationary an eigenvalue of Eq. (1). There is no such potential, if $k\leqq 0$ or $n\leqq 1$. The corresponding eigenfunction u_0 is, from Eqs. (6) and (12),

$$u_0(x)=\left\{\frac{-\lambda_0 n}{n-1}\operatorname{sech}^2\left[\frac{(-\lambda_0)^{\frac{1}{2}}x}{n-1}\right]\right\}^{(n-1)/2} \tag{17}$$

Since $u_0(x)\neq 0$, λ_0 is the lowest eigenvalue of the potential $V_0(x)$.

3. ISOPERIMETRIC INEQUALITIES AND LOWER BOUNDS

Let us now assume that the stationary value λ_0 is actually the minimum value of the lowest eigenvalue of any potential satisfying Eq. (2), which we shall prove in the next section. Then, if λ is the lowest eigenvalue of some potential $V(x)$, we have $\lambda\geqq\lambda_0$ provided $V(x)$ satisfies Eq. (2). If we define k in terms of $V(x)$ by Eq. (2), and use Eq. (15) for λ_0 we then have the inequality

$$\lambda\geqq -F(n)\left[\int_{-\infty}^{\infty}V^n(x)dx\right]^{2/(2n-1)}. \tag{18}$$

Equality holds in Eq. (18) only if $V(x)=V_0(x)$. For each $n>1$ this inequality (18) is the isoperimetric inequality we sought. It provides lower bounds on the lowest eigenvalue λ. A graph of $F(n)$ is shown in Fig. 1.

In the limit $n=1$, Eq. (18) yields the following lower bound, obtained previously by Larry Spruch:

$$\lambda\geqq -\tfrac{1}{4}\left[\int_{-\infty}^{\infty}V(x)dx\right]^2. \tag{19}$$

As n tends to unity, the limiting form of the potential $V_0(x)$, given by Eq. (12), is the delta function for which equality holds in Eq. (19).

To illustrate the use of Eq. (18), we shall now apply it to a square well of depth V and width $2a$. The integral in Eq. (18) is then $2aV^n$ and Eq. (18) becomes

$$\lambda/V\geqq -F(n)(4a^2V)^{1/2n-1}. \tag{20}$$

A simple calculation shows that the right-hand side of Eq. (20) is largest when n satisfies the equation

$$4a^2V=\pi n(n-1)\frac{\Gamma^2(n)}{\Gamma^2(n+\frac{1}{2})}\times\exp\{(2n-1)[\Psi(n+\tfrac{1}{2})-\Psi(n)]\}. \tag{21}$$

Here $\Psi(n)=\Gamma'(n)/\Gamma(n)$. A graph of the lower bound Eq. (20) with n determined from Eq. (21) is shown in Fig. 2 as a function of a^2V. For comparison the exact value of λ is also shown.

Let us now apply Eq. (18) to the exponential potential of depth V and range a given by

$$V(x)=Ve^{-|x|/a}. \tag{22}$$

Upon inserting Eq. (22) into Eq. (18), we obtain

$$\lambda/V\geqq -F(n)(4a^2V/n^2)^{1/(2n-1)}. \tag{23}$$

The right-hand side of Eq. (23) is largest when n satisfies the equation

$$4a^2V=\pi n^3(n-1)\frac{\Gamma^2(n)}{\Gamma^2(n+\frac{1}{2})}\times\exp\{(2n-1)[\Psi(n+\tfrac{1}{2})-\Psi(n)-1/n]\}. \tag{24}$$

The lower bound on λ/V given by Eq. (23), with n determined by Eq. (24), is shown in Fig. 3 as a function of a^2V.

As another example of the use of Eq. (18), let us apply it to the inverse αth power potential of depth V and range a given by

$$V(x)=V/(1+a^{-1}|x|)^{\alpha}. \tag{25}$$

For this potential Eq. (18) yields

$$\lambda/V\geqq -F(n)[4a^2V/(n\alpha-1)^2]^{1/(2n-1)}. \tag{26}$$

The bound in Eq. (26) is largest when n satisfies the equation

$$4a^2V=\pi n(n-1)(n\alpha-1)^2[\Gamma^2(n)/\Gamma^2(n+1/2)]\times\exp\{(2n-1)[\Psi(n+\tfrac{1}{2})-\Psi(n)-\alpha/(n-1)]\}. \tag{27}$$

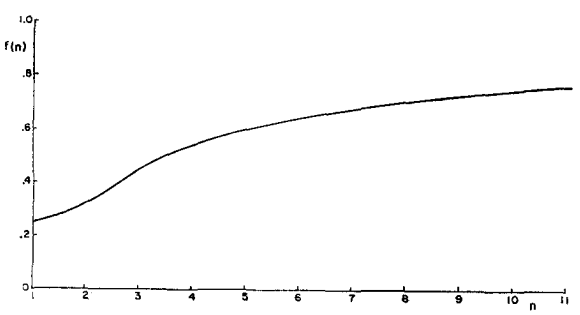


FIG. 1. The function $F(n)$, given by Eq. (16), as a function of n. This function occurs in the isoperimetric inequality (18). As n becomes infinite, $F(n)$ approaches unity.

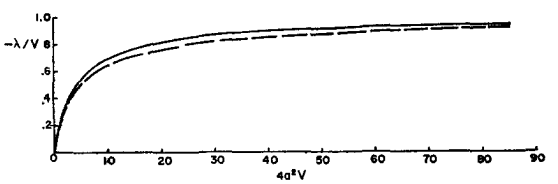


FIG. 2. The best lower bound on the lowest eigenvalue of a square well potential of depth V and width $2a$ is shown as a function of a^2V (solid curve). The exact lowest eigenvalue is also shown for comparison (dashed curve). The ordinate is $-\lambda/V$, and the bound is computed from Eqs. (20) and (21).

4. PROOF THAT λ_0 IS A MINIMUM

To prove that λ_0 is the minimum value of the lowest eigenvalue of any potential satisfying Eq. (2), we begin with the variational characterization of the lowest eigenvalue of a potential $V(x)$. It is

$$\lambda=\min_{v}\left[\int_{-\infty}^{\infty}(v_x^2-v^2V)dx\Big/\int_{-\infty}^{\infty}v^2dx\right]. \tag{28}$$

Following a suggestion of H. F. Weinberger, we make use of the Holder inequality which holds for any $n>1$

$$\int_{-\infty}^{\infty}v^2Vdx\leqq\left(\int_{-\infty}^{\infty}V^ndx\right)^{1/n}\left(\int_{-\infty}^{\infty}v^{2n/n-1}dx\right)^{(n-1)/n}. \tag{29}$$

Equality obtains in Eq. (29) if and only if $V^{n-1}=v^2$, which is the same condition as Eq. (6). Upon inserting Eq. (29) into Eq. (28), and making use of Eq. (2), we obtain

$$\lambda\geqq\min_{v}\left[\int_{-\infty}^{\infty}v_x^2dx-k^{1/n}\left(\int_{-\infty}^{\infty}v^{2n/(n-1)}dx\right)^{(n-1)/n}\Big/\int_{-\infty}^{\infty}v^2dx\right]. \tag{30}$$

We must now show that the right-hand side of Eq. (30) is minimized when $v=v_0$. If it is, then the right side of Eq. (30) is just λ_0 since for v_0 and V_0 equality holds in Eq. (29), and, therefore, in Eq. (30). Then Eq. (30) yields the desired inequality

$$\lambda\geqq\lambda_0. \tag{31}$$

A necessary condition for v to minimize the expression on the right-hand side of Eq. (30) is obtained by requiring the first variation of that expression to vanish. This yields the condition

$$v_{xx}+k^{1/n}\left(\int_{-\infty}^{\infty}v^{2n/(n-1)}dx\right)^{-1/n}v^{1+(2/n-1)}+\lambda'v=0. \tag{32}$$

Here λ' denotes the minimum value of the expression on the right-hand side in Eq. (30). This equation becomes identical with Eq. (7), if we introduce $x'=cx$ and $\lambda''=c^{-2}\lambda'$ where

$$c^2=k^{1/n}\left(\int_{-\infty}^{\infty}v^{2n/(n-1)}dx\right)^{-1/n}. \tag{33}$$

Therefore, its solution is just $c^{1-n}u_0(x)$, as we see from Eq. (17), with λ' in place of λ_0. Then Eq. (33), which determines λ', becomes identical with Eq. (13) so $\lambda'=\lambda_0$. Therefore, if the minimum in Eq. (30) exists, its value is λ_0 and the minimizing function is $v_0(x)$.

The existence of the minimum in Eq. (30) can be proved by standard methods of the calculus of variations, although the proof is by no means trivial.

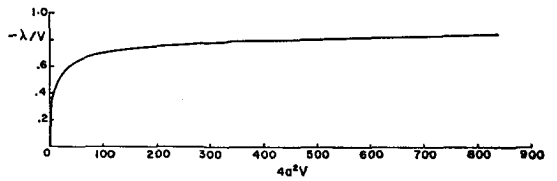


FIG. 3. The best lower bound on the lowest eigenvalue of the exponential potential $V(x)=V\exp(-|x|/a)$ of depth V and range a is shown as a function of a^2V. The ordinate is $-\lambda/V$ and the bound is computed from Eqs. (23) and (24).

5. AN UPPER BOUND

Let us now consider the isoperimetric problem obtained by replacing the class of potentials satisfying Eq. (2) by those satisfying the condition

$$\int_{-\infty}^{\infty}V(x)\rho^2(x)dx=k. \tag{2'}$$

Here $\rho(x)$ is a given function and k is a given constant. By proceeding as in Sec. 2 we obtain, instead of Eq. (4),

$$\int_{-\infty}^{\infty}\dot{V}\rho^2dx=0. \tag{4'}$$

Then instead of Eq. (6), we find from Eqs. (4') and (5) $u_0^2=\rho^0$ or equivalently

$$u_0(x)=\rho(x). \tag{6'}$$

Since u_0 is quadratically integrable, we see that $\rho(x)$ must also be so. Now Eqs. (1) and (6') yield

$$V_0(x)=-\lambda_0-\rho^{-1}\rho_{xx}. \tag{12'}$$

From Eq. (2') and (12') we find λ_0 which is given by

$$\lambda_0=\left(\int_{-\infty}^{\infty}\rho_x^2dx-k\right)\Big/\int_{\infty}^{\infty}\rho^2dx. \tag{15'}$$

If λ_0 is the maximal lowest eigenvalue of any potential satisfying Eq. (2'), then $\lambda\leqq\lambda_0$, or using Eqs. (15') and (2'),

$$\lambda\leqq\left(\int_{-\infty}^{\infty}\rho_x^2dx-\int_{-\infty}^{\infty}\rho^2Vdx\right)\Big/\int_{-\infty}^{\infty}\rho^2dx. \tag{16'}$$

But Eq. (16') is true, since the right-hand side is just the Rayleigh quotient evaluated for the trial function $\rho(x)$. Thus we have found that this Rayleigh quotient is the largest lowest eigenvalue of any potential satisfying Eq. (2') and it is attained for the potential in Eq. (12').

6. HIGHER DIMENSIONS

Some of the preceding considerations apply in any number of dimensions. Even many of our equations remain valid, if we interpret x as a vector and replace u_{xx} by ∇^2u. With these changes Eqs. (1)–(7) remain valid. Then Eq. (7) is the equation which must be

satisfied by the eigenfunction u_0 of the potential V_0 which makes λ stationary. If we seek a spherically symmetric solution $u_0(r)$, then Eq. (7) becomes, in three dimensions,

$$u_{0rr}+(2/r)u_{0r}+u_0^{1+(2/n-1)}+\lambda_0 u_0=0. \qquad (7'')$$

It has not been possible to solve this equation explicitly.

All of the equations of Sec. 4 remain valid, if we also replace v_x^2 by $(\nabla v)^2$. However, the proof of the existence of the minimum in Eq. (30) has not been carried out, nor has it been shown that the minimizing potential, if one exists, is spherically symmetric.

All the results of Sec. 5 hold in any number of dimensions.

Errata: Statistical Dynamics of Simple Cubic Lattices. Model for the Study of Brownian Motion

[J. Math. Phys. 1, 309 (1960)]

ROBERT J. RUBIN
National Bureau of Standards, Washington 25, D. C.

In Eqs. (16), (17), (19), (A1), and in the integral at the end of Sec. III (p. 312), replace V by $-V$ where it appears explicitly. In the fourth and fifth lines following Eq. (16), delete the expression in brackets.

In Eq. (7a), replace $(2N+1)^{-n/2}$ by $(2N+1)^{-n}$.

In Eq. (C2), replace $\ln(\frac{1}{2}\rho^2)$ by $\ln(1/2\rho^2)$.

satisfied by the eigenfunction u_0 of the potential V_0 which makes λ stationary. If we seek a spherically symmetric solution $u_0(r)$, then Eq. (7) becomes, in three dimensions,

$$u_{0rr}+(2/r)u_{0r}+u_0^{1+(2/n-1)}+\lambda_0 u_0=0. \qquad (7'')$$

It has not been possible to solve this equation explicitly.

All of the equations of Sec. 4 remain valid, if we also replace v_x^2 by $(\nabla v)^2$. However, the proof of the existence of the minimum in Eq. (30) has not been carried out, nor has it been shown that the minimizing potential, if one exists, is spherically symmetric.

All the results of Sec. 5 hold in any number of dimensions.

Errata: Statistical Dynamics of Simple Cubic Lattices. Model for the Study of Brownian Motion

[J. Math. Phys. 1, 309 (1960)]

ROBERT J. RUBIN
National Bureau of Standards, Washington 25, D. C.

In Eqs. (16), (17), (19), (A1), and in the integral at the end of Sec. III (p. 312), replace V by $-V$ where it appears explicitly. In the fourth and fifth lines following Eq. (16), delete the expression in brackets.

In Eq. (7a), replace $(2N+1)^{-n/2}$ by $(2N+1)^{-n}$.

In Eq. (C2), replace $\ln(\frac{1}{2}\rho^2)$ by $\ln(1/2\rho^2)$.