Construction of a Complete Set of Independent Observables in the General Theory of Relativity*

ARTHUR KOMAR
Department of Physics, Syracuse University, Syracuse, New York
(Received April 7, 1958)

The construction of a complete set of quantities in general relativity, whose functional form is invariant under coordinate transformations, is indicated. The set obtained is highly redundant. The Cauchy problem for obtaining an independent complete set of such quantities ("observables") is therefore discussed. It is also pointed out that the observables obtained may alternatively be viewed as the metric tensor in a special "gauge" (i.e., with a special coordinate condition). This latter viewpoint may facilitate the quantization of general relativity.

1. INTRODUCTION

THE well-known and most conspicuous property of the general theory of relativity, the general covariance of its laws under arbitrary nonsingular coordinate transformations, has as a necessary consequence the unobservability of the coordinates by which we identify points in space-time. For clearly, by an appropriate coordinate transformation, a given world point can be given arbitrary coordinates relative to any other set of world points. A further, and perhaps more serious, consequence of the general covariance is that the metric tensor, the potentials of the gravitational field, as a function of these coordinates, is also not observable. This circumstance can be reflected, for example, in two alternative ways: (1) Given two metric tensor fields, one cannot readily tell whether they represent two distinct physical situations or whether they represent the same physical situation but in two different coordinate systems (the so-called equivalence problem). (2) Given the metric tensor and its first normal derivative on a space-like hypersurface, its continuation in time is not uniquely determined by the field equations of the theory, for we are evidently free to perform coordinate transformations which do not alter the coordinates and Cauchy data on the initial hypersurface.

Observables in classical gravitation theory are analogous to gauge-invariant quantities in classical electromagnetic theory. They are uniquely determined by the physical situation under consideration. Furthermore, if one has a *complete* set of such observables, they will uniquely determine or characterize the physical situation, totally removing the ambiguity engendered by general covariance. Apart from the obvious interest that such observables possess for the classical theory of gravitation, the efforts toward quantization of general relativity indicate the necessity for having only such quantities correspond to operators in the Hilbert space of physical states.[1]

The purpose of this paper is to indicate a general procedure for the construction of a complete set of observables in gravitation theory. We shall construct those observables in Sec. 2 and shall indicate the limits of validity of the procedure. In the third section of this paper we shall digress slightly to present a somewhat unconventional but straightforward version of the Cauchy problem in general relativity. We shall employ this approach in Sec. 4 to attack the problem of finding a complete set of *independent* observables, that is, those observables whose values or functional forms can be arbitrarily assigned, so that each distinct assignment will lead uniquely to a physically distinct solution of the field equations of the theory.

2. CONSTRUCTION OF TRUE OBSERVABLES

The coordinates of world points are evidently not observables, but a world point may be identified invariantly by different methods. Clearly, the scalars which one is able to construct by combining the metric tensor, the Riemann tensor, and its covariant derivative, will in general have different values at different world points. Since scalars do not alter their values under coordinate transformations, these numerical values can serve to identify world points in an intrinsic fashion, thus giving a first means for the construction of observables. In particular, for spaces which do not admit a symmetry group of isometries we know that some set of four of these curvature scalars may be found which is functionally independent.[2] Such a set of scalars can then serve to identify world points uniquely. For the remainder of this paper we shall limit our attention to the consideration of such nonsymmetric spaces. The nonsymmetric spaces clearly comprise the vast majority of those spaces which satisfy the gravitational field equations. In addition, a lack of symmetry corresponds to our most immediate and direct observations and impressions of the universe. Unfortunately, in obtaining all our known solutions of the gravitational field equations, for simplicity some form of symmetry had to be assumed. These solutions

* Supported in part by the National Science Foundation and the Air Force Office of Scientific Research.

[1] For further and more thorough discussions of these questions see P. G. Bergmann, Nuovo cimento 3, 1177 (1956); E. Newmann and P. G. Bergmann, Revs. Modern Phys. 29, 443 (1957) and the references found in this latter paper. See also P. A. M. Dirac, *Quantum Mechanics* (Clarendon Press, Oxford, 1947), third edition, p. 286.

[2] A. Komar, Proc. Natl. Acad. Sci. U. S. 41, 758 (1955).

are thereby included in the "set of zero measure" we are choosing not to consider.

Although it will not be essential for the considerations of this article, there is good reason, for the case of empty space-times, to impose an additional limitation. We can require that the four functionally independent scalars, whose existence we are assuming, be the four independent real scalars which are determined by the eigenvalue problem[3]

$$(R_{ijmn}-\rho g_{ijmn})V^{mn}=0,$$

where R_{ijkl} is the Riemann Tensor, $g_{ijkl}=g_{ik}g_{jl}-g_{il}g_{jk}$, and V^{kl} is an antisymmetric tensor. The requirement that these four scalars be functionally independent is a much stronger condition, which is satisfied only in the most asymmetric of spaces and which corresponds to the exclusion of Pirani's type II and III spaces of pure radiation, in addition to excluding symmetric type I spaces. At this stage of our exposition, the principal motivating factors in imposing this additional limitation are twofold: (1) apart from scalars constructed trivially by taking algebraic functions of these four scalars, they are the only nontrivial scalars which are of least possible order in derivatives of the metric, thus making them the simplest and most natural choice; (2) since they stem from an eigenvalue problem, their structure will be much simpler to analyze than other possible choices of scalars.

Throughout this paper we shall adhere to the convention that Latin indices run from 1 to 4, and Greek indices run from 1 to 3. Let A^i be the four functionally independent curvature scalars (i.e., four specific and distinguishable scalar functions constructed from the metric tensor and its derivatives). To emphasize that these four functions uniquely and intrinsically identify world points, let us go to the new coordinate system determined by the A^i:

$$\bar{x}^i=A^i(x). \tag{2.1}$$

If we inquire into what the metric tensor looks like in this new coordinate system we find the usual expression:

$$\bar{g}^{ij}=\frac{\partial A^i}{\partial x^m}\frac{\partial A^j}{\partial x^n}g^{mn}. \tag{2.2}$$

However, we now note that since A^i is a scalar, the $\partial A^i/\partial x^m$ is a covariant vector and therefore $\bar{g}^{ij}$ is component by component a well defined scalar constructed from the metric tensor and its derivatives. If we consider two metric tensor fields and ask whether they represent the same physical situation, differing perhaps by being viewed in different coordinate systems, we now have a ready criterion for determining the answer. Clearly, at corresponding points in any identification of the two spaces, the values of all scalars must agree if the spaces are to be equivalent. We are therefore compelled to identify points in the two spaces which have the same "intrinsic" coordinates defined by Eq. (2.1). Furthermore at these corresponding points it is necessary that the ten scalars $\bar{g}^{ij}$ defined by Eq. (2.2) have the same values in the two spaces. However, since in this coordinate system the 10 scalars $\bar{g}^{ij}$ are also the components of the metric tensor, we see immediately that these conditions are not only necessary but also sufficient for the solution of the equivalence problem. Thus we find that the functional form of the 10 scalars $\bar{g}^{ij}$ as functions of the four scalars A^i (a) is uniquely determined by the metric space independently of any choice of coordinate system, and furthermore (b) uniquely characterizes the space. Part (a) of this statement implies that the $\bar{g}^{ij}$ as functions of A^i are "invariants" or "observables" of the space; and part (b) implies that this set of observables is a complete set. However, we have clearly a highly redundant set of observables.

We will consider in the remainder of this paper those spaces which are solutions of the Einstein field equations

$$R_{ij}=0. \tag{2.3}$$

Since the ten scalars of Eq. (2.2) are also components of the metric tensor in the coordinate system determined by Eq. (2.1), they must therefore satisfy the ten Eqs. (2.3) as functions of these coordinates. In addition, the A^i may be considered as four (nonlinear) differential operators acting on the metric tensor. These operators, via Eq. (2.1), must yield back the coordinates, to insure that we are in fact in the intrinsic coordinate system. We thus obtain four additional conditions

$$A^i(g_{mn})=x^i. \tag{2.4}$$

We should mention in this connection that although the observables are defined as ten scalar functions of four other scalars, heuristically it is much easier to consider the quantities as the metric tensor with a particular coordinate condition given by Eq. (2.4)—even though this viewpoint obscures the observable character of the quantities to some extent.

3. THE USUAL CAUCHY PROBLEM

In order to determine a nonredundant complete set of observables, we are led to consider the initial-value problem on a space-like hypersurface for the combined set of Eqs. (2.3) and (2.4). Customarily one selects the coordinate system in such a way that this space-like hypersurface has the equation $x^4=$const. However in view of Eq. (2.4) we no longer have this freedom. This section is therefore devoted to indicating briefly the form of the initial-value problem in an arbitrary coordinate system.

The initial space-like hypersurface may be specified in two equivalent ways. We can specify that it have the

[3] F. A. E. Pirani, Phys. Rev. **105**, 1089 (1957). See also J. Geheniau and R. Debever, Helv. Phys. Acta. Suppl. IV, 101 (1956).

equation

$$S(x^i)=0, \tag{3.1}$$

or, alternatively, we can give the equation of the surface in parametric form:

$$x^i=X^i(U^\alpha). \tag{3.2}$$

The parameters U^α may be considered as a coordinate system which specifies points on the hypersurface. Using these parameters, we determine the initial-value problem for the field equations (2.3) by specifying

$$g_{ij}=g_{ij}(U^\alpha), \tag{3.3}$$

and

$$\xi^m\frac{\partial g_{ij}}{\partial x^m}=h_{ij}(U^\alpha), \tag{3.4}$$

where ξ^i is the unit normal to the initial surface. We shall call these 20 functions of 3 parameters the Cauchy data, and shall indicate how they determine a solution of the field equations (2.3).

Since ξ^i is the unit normal to S, it satisfies the equations

$$g_{mn}X^m{}_{,\alpha}\xi^n=0 \quad \left(X^m{}_{,\alpha}\equiv\frac{\partial X^m}{\partial U^\alpha}\right), \tag{3.5}$$

and

$$g_{mn}\xi^m\xi^n=-1. \tag{3.6}$$

Thus from Eqs. (3.2) and (3.3) we can determine ξ^i as a function of the U^α. We shall indicate this by writing

$$\xi^i=f(U) \tag{3.7}$$

where throughout the remainder of this section $f(U)$ shall denote a *generic* function [or set of functions, as in Eq. (3.7)] completely determined by a knowledge of the Cauchy data and/or the equation of the surface S.

Differentiating Eq. (3.3) with respect to U^α we obtain

$$X^m{}_{,\alpha}g_{ij,m}=f(U). \tag{3.8}$$

Since the rank of the matrix

$$M\equiv\begin{pmatrix} X^m{}_{,\alpha} \\ \xi^m \end{pmatrix}=f(U) \tag{3.9}$$

is four, we deduce from Eqs. (3.4) and (3.8) that

$$g_{ij,k}=f(U). \tag{3.10}$$

It is easy to see that the metric of Eq. (3.3) uniquely determines the metric of the hypersurface, $\gamma_{\alpha\beta}$, via

$$\gamma_{\alpha\beta}=X^m{}_{,\alpha}X^n{}_{,\beta}g_{mn}=f(U). \tag{3.11}$$

We shall use this metric to raise and lower Greek indices in the usual fashion. Note that under transformations in x-space the $\gamma_{\alpha\beta}$ behave as a set of 6 scalars, while under transformations in U space it transforms like a symmetric tensor. Similarly $X^m{}_{,\alpha}$ behaves as a contravariant vector under x-space transformations, and as a covariant vector under U-space transformations. Such behavior will be true in general, namely, Latin indices will behave as tensor indices under x-space transformations and as scalar indices (or enumerators) under U-space transformations, whereas we have the opposite behavior for Greek indices. With this in mind it becomes trivial to deduce the important relationship

$$g^{ij}=X^i{}_{,\alpha}X^{j,\alpha}-\xi^i\xi^j. \tag{3.12}$$

For we may first make an x-space transformation in such a manner that $g^{i4}=-\delta_4{}^i$ and $\xi^i=\delta_4{}^i$. This is obtained by taking the surfaces geodesically parallel to S to have the equation $x^4=\text{const.}$, where the constant is the geodesic distance of the surface from S. We then perform a U-space transformation so that $U^\alpha=x^\alpha$. Now, in view of Eq. (3.11), Eq. (3.12) becomes a trivial identity. However, since it is a covariant relationship it remains true in arbitrary coordinate systems both in x-space and U-space.

By a similar argument it is readily seen that on S we can replace the field Eqs. (2.3) by the equivalent set

$$G_i\equiv\xi_m(R_i{}^m-\tfrac{1}{2}\delta_i{}^mR)=0, \tag{3.13a}$$

$$P_{\alpha\beta}\equiv X^m{}_{,\alpha}X^n{}_{,\beta}R_{mn}=0. \tag{3.13b}$$

The reason for decomposing the field equations in this manner becomes evident by observing that Eq. (3.13a) depends purely on the Cauchy data, and in fact represents a set of four constraints which limit our freedom to assign the data arbitrarily. This is readily confirmed as follows:

$$G_k{}^i\equiv R_k{}^i-\tfrac{1}{2}\delta_k{}^iR=\tfrac{1}{2}g^{pq}[g^{im}(g_{mq,pk}+g_{pk,mq}-g_{pq,mk}-g_{mk,pq})-\delta_k{}^ig^{mn}(g_{mp,nq}-g_{mn,pq})]+f(U). \tag{3.14}$$

Employing (3.12) and (3.6), we find

$$G_i=-\tfrac{1}{2}X^m{}_{,\alpha}X^{n,\alpha}(\xi^pg_{mn,pi}+\xi_i\xi^p\xi^qg_{mn,pq})+f(U). \tag{3.15}$$

Therefore, again employing (3.6), we have

$$\xi^mG_m=f(U). \tag{3.16}$$

Furthermore if we observe that via Eq. (3.10)

$$X^m{}_{,\alpha}g_{ij,km}=\frac{\partial g_{ij,k}}{\partial U^\alpha}=f(U), \tag{3.17}$$

and that we have the orthogonality relationship (3.5), we readily obtain from (3.15)

$$X^m{}_{,\alpha}G_m=f(U). \tag{3.18}$$

Combining Eqs. (3.16) and (3.18), with due consideration of (3.9), we easily obtain

$$G_i=f(U). \tag{3.19}$$

Moreover, it is readily confirmed, by making use of (3.12) and the well known identity

$$G_i{}^m{}_{;m}=0, \tag{3.20}$$

that if (3.13) is satisfied on S, then the first derivative of (3.13a) also vanishes on S. Thus the field equations on S insure the propagation of the constraining relations (3.13a) away from S. We therefore turn our attention to Eq. (3.13b) to determine how the metric propagates from S. By direct computation, again making use of (3.12) and (3.17), we find

$$P_{\alpha\beta}=X^m{}_{,\alpha}X^n{}_{,\beta}\xi^p\xi^q g_{mn,pq}+f(U). \qquad (3.21)$$

By repeated use of (3.17) and (3.9) it is then easy to deduce from (3.13b) on S that

$$X^m{}_{,\alpha}X^n{}_{,\beta}g_{mn,ik}=f(U). \qquad (3.22)$$

However at this stage this is as far as we can go by relying just on the field equations. But what we are after is to show that we can obtain $g_{ij,kl}$ as a function of the Cauchy data. The difficulty in accomplishing this is clearly a reflection of the fact that the solutions of the field equations are not uniquely determined by the Cauchy data. For we are free to perform coordinate transformations which alter neither the equations of the initial surface S nor the Cauchy data on the surface. Such a transformation has the form

$$\bar{x}^i=x^i+S^3(x)\phi^i(x), \qquad (3.23)$$

where $S(x)$ is given by the equation of S, (3.1), and $\phi^i(x)$ is an arbitrary nonsingular function. It is easy to verify that the $\phi^i(x)$ may be chosen so that $\xi^p\xi^q\xi^r g_{ip,qr}$ can have any preassigned behavior on S. In particular we may take ϕ^i so that

$$\xi^p\xi^q\xi^r g_{ip,qr}=0. \qquad (3.24)$$

In fact, by appropriately choosing ϕ^i it can be arranged that

$$\xi^p\xi^q\xi^r\cdots\xi^s g_{ip,qr\cdots s}=0. \qquad (3.25)$$

One may also check directly that the coordinate transformation (3.23) does not alter Eq. (3.22), as we should expect by the fact that it does not alter the Cauchy data on S. If we now consider simultaneously the set of equations (3.22) and (3.24), by repeated application of (3.9) and (3.17) we find

$$g_{ij,kl}=f(U). \qquad (3.26)$$

In a similar fashion, by considering higher derivatives of the field equations, and taking into account the coordinate conditions (3.25), we can obtain a unique expression for all the higher derivatives of the metric as a function of the Cauchy data. Thus we have re-established the well known result[4] that if the Cauchy data are assigned on an initial space-like hypersurface, arbitrarily *modulo* four constraints (3.13a), then the solution of the Einstein field equations (2.3) is uniquely determined *modulo* an arbitrary coordinate transformation which does not alter both the equation of the initial surface and the Cauchy data on the surface. The purpose of repeating these well-known results in such detail was to present them in a form and notation which can now be directly applied to the Cauchy problem of the combined set of Eqs. (2.3) and (2.4).

[4] A. Lichnerowicz, *Théories Relativistes de la Gravitation et de l'Electromagnétisme* (Masson et Cie, Paris, 1955), Chap. II.

4. CAUCHY PROBLEM FOR INDEPENDENT OBSERVABLES

As we can observe from Eq. (3.22), the field Eqs. (2.3) only determine the propagation of the "transverse" components of the metric tensor from the initial surface. The coordinate conditions (2.4) will determine the propagation of the remaining four components of the metric. However, if we attempt to consider Eqs. (2.3) and (2.4) simultaneously, there will be involved conditions of integrability occurring between them. For this reason it is much easier to build upon the structure which we have established in the previous section.

If we have a solution of the field equations (2.3) with a sufficient lack of symmetry, we can always perform a coordinate transformation to obtain (2.4). We therefore pose the following problem: given a solution of the field equations (2.3) determined by Cauchy data on an initial space-like hypersurface and a particular choice of coordinate conditions, say that of Eqs. (3.24) and (3.25), what additional conditions [other than (3.13a)] must the Cauchy data satisfy in order to insure that it may be continued in an intrinsic coordinate system [i.e., one that satisfies Eq. (2.4)]?

When phrased in this fashion, the solution of the Cauchy problem becomes transparent. Clearly, the four scalars A^i must have such a form that when we perform the coordinate transformation, in our given solution, to the intrinsic frame, the Cauchy data remain unaltered; that is, A^i must have the form of Eqs. (3.23),

$$A^i=x^i+S^3\phi^i. \qquad (4.1)$$

Thus we readily obtain a set of necessary conditions on the Cauchy data on S:

$$A^i=X^i(U^\alpha), \qquad (4.2a)$$

$$\xi^m A^i{}_{,m}=\xi^i(U^\alpha), \qquad (4.2b)$$

$$\xi^m\xi^n A^i{}_{,mn}=0. \qquad (4.2c)$$

More accurately stated, Eqs. (4.2) become conditions on the Cauchy data when all derivatives of the metric higher than the first which occur in the expressions (4.2) are replaced by the Cauchy data which determine them through the field equations (2.3) and our assumed initial choice of coordinate conditions (3.24) and (3.25). Furthermore, we shall state without proof that it is easy to verify that Eqs. (4.2) are sufficient conditions for the scalars A^i to have the form (4.1). Thus Eqs. (4.2) are the sought-for additional requirements on the Cauchy data which insure that the coordinate transformation to the intrinsic coordinate system will not alter the Cauchy data on the initial hypersurface (3.1).

The conditions (4.2) are clearly independent of each other and of (3.13a), for any one of them may be violated in an arbitrary fashion purely by a choice of coordinate transformations. We are therefore in a position to state that if, on an initial hypersurface, the 20 functions of 3 coordinates, (3.3) and (3.4), which comprise the Cauchy data, are assigned in conformity with the 16 independent constraining relations (3.13a) and (4.2), then this is necessary and sufficient to insure that the metric may be continued off the initial surface in a unique fashion such that the combined set of Eqs. (2.3) and (2.4) are satisfied. The independent complete set of observables would then be the 4 functions of 3 coordinates (if such four functions in fact exist) which could be arbitrarily assigned such that the 16 Eqs. (3.13a) and (4.2) are automatically satisfied. To find such four functions requires a closer investigation of these constraining relations which is now being carried out. The character of the independent complete set of observables deduced above, namely that they consist of 4 functions of 3 coordinates, agrees precisely with what has been determined by general considerations of the canonical formalism of general relativity.[5] This agreement in numerology gives added encouragement that precisely four such functions may be found.

We have indicated that the constraining relations (4.2) on the Cauchy data are necessary and [taken together with (3.13a)] sufficient relations to insure the determination of the problem of the independent observables. However, in deriving these relations explicit use was seemingly made of the initial coordinate conditions (3.24) and (3.25). One might naturally question whether the form of the relations (4.2) as functions of the Cauchy data would be drastically altered if other coordinate conditions were initially used to fix the propagation of the four "normal" components of the metric tensor. In point of fact this does not happen—the form of Eqs. (4.2) as functions of the Cauchy data is independent of any assumed initial coordinate conditions. For Eq. (4.2a) this statement is evident: the value of a scalar at any world point is independent of the choice of coordinate conditions. In general, the independence of the form of Eqs. (4.2) on the choice of initial coordinate conditions stems from the invariance of the form of Eq. (4.1) under the group of all transformations which leave the Cauchy data unaltered, namely all transformations of the form (3.23). The explicit demonstration of this fact is straightforward and need not be carried out here.

5. CONCLUSION

The basic method for constructing observables in general relativity is quite simple, and in fact appreciably more general than we have thus far indicated. Namely, if the space admits four functionally independent scalars, we may use these scalars to replace the coordinates occurring in any other scalar. The resulting functional form of the scalar function of scalars is an observable. If the space admits only three functionally independent scalars, it necessarily admits a Killing vector field[2] and therefore all scalars can be written as functions of three coordinates, which can then be replaced by the three functionally independent scalars. The resulting functional forms are again observables. Similarly, when a space admits less than three functionally independent scalars, observables can be constructed, but the construction becomes increasingly difficult since one must then take into specific consideration the structure constants of the group of isometries which the space admits.

The principal reason why we, in this paper, neglected to discuss the observables of spaces with groups of motions was not so much the difficulty in constructing observables, but rather the difficulty in finding a complete set of observables. In general, the particular scalars which one has to consider in order to obtain a complete set differs from space to space according to the group of motions which the space admits. That, in general, there exists a complete set of observables, is assured us by an existence theorem on the equivalence problem in terms of scalars.[6]

When we turn to the problem of the quantization of general relativity, we find a very important reason for selecting the four scalars determined by the eigenvalue equation at the beginning of Sec. 2 as the scalars employed in condition (2.4). Apart from trivial algebraic combinations, these four scalars are the only nontrivial scalars which can be constructed purely from the Riemann tensor[3] (i.e., without having to go to higher derivatives). Due to the well-known symmetries of the Riemann tensor, only the "transverse" components of the metric tensor occur differentiated twice in the "time" (i.e., normal) direction. However these are precisely the terms which are determined via the field equations (2.3) in terms of quantities which are of at most first order in time. The intrinsic coordinate conditions (2.4) can therefore be put into a form which are of first order in "time." Should it prove to be impossibly difficult to extract the independent observables from Eqs. (3.13a) and (4.2), we now have recourse to considering Eq. (2.4) as a conventional coordinate condition and employing the entire established machinery for the canonical quantization.[5] The usual argument, that employing coordinate conditions may destroy the general covariance of the resulting quantum theory, does not apply in this case. For considering (2.4) as a coordinate condition is only a heuristic device to make it easier to visualize how to manipulate the

[5] P. G. Bergmann, Helv. Phys. Acta Suppl. IV, 79 (1956) (particularly the middle of p. 91).

[6] A. Komar, Phys. Rev. **99**, 662(A) (1955).

quantities with which we are dealing. In point of fact, we know how to interpret these quantities as true observables. (An apt analog in electromagnetic theory may clarify this point of view. The transverse components of the vector potential may be considered as the gauge-invariant true observables; or they may be considered as the components of the vector potential in a particular gauge, namely the radiation gauge.) An investigation of the quantization of general relativity from the point of view of considering (2.4) as a conventional coordinate condition is now in progress.

ACKNOWLEDGMENTS

I would like to acknowledge helpful discussions with Dr. P. G. Bergmann and Dr. R. Sachs concerning the significance of the problem of constructing observables in general relativity in relation to the problem of quantizing the theory.

PHYSICAL REVIEW VOLUME 111, NUMBER 4 AUGUST 15, 1958

Spectral Representations in Perturbation Theory. I. Vertex Function*

ROBERT KARPLUS, CHARLES M. SOMMERFIELD,† AND EYVIND H. WICHMANN
Physics Department, University of California, Berkeley, California
(Received April 21, 1958)

The vertex operator is examined in lowest order perturbation theory. It is found that, as a function of the invariant momentum transfer $q^2=\mathbf{q}^2-q_0^2$, it is analytic in a cut plane with the branch point on the negative real axis. A spectral representation (dispersion relation) may therefore be inferred. The threshold of the spectrum depends on the masses of all fields involved unless certain inequalities hold between the masses of the incident and outgoing particles on one hand and the particles in intermediate states on the other; in that case the threshold depends only on the intermediate masses.

DISPERSION relations and spectral representations in terms of physically accessible intermediate states have recently supplemented perturbation expansions as a theoretical tool in the study of elementary-particle interactions.[1] It has been possible to derive them as general consequences of a causal, Lorentz-invariant field theory, but only with the imposition of rather curious restrictions on the masses of the interacting particles.[2] We have in mind particularly the nucleon electromagnetic form factor and the nucleon-nucleon scattering amplitude, both as determined by coupling to the pion field, for which dispersion relations hold only if

$$m_\pi > (\sqrt{2}-1)M_N, \tag{1}$$

where m_π and M_N are the pion and nucleon masses, respectively. Unfortunately, the observed masses do not satisfy this inequality; the question, whether the quantities mentioned have the desired spectral representations, therefore remains unanswered.

As a guide in this matter we may appeal to perturbation theory[3]; and it is easily established that at least the low-order contributions calculated from pseudoscalar meson theory indeed have the analyticity properties from which the dispersion relations can be obtained. This result has given rise to speculation that the exact solution to the field-theoretical problem may have properties quite different from those of the perturbation series taken term by term.[4]

We believe that the explanation is far simpler: the perturbation expansion contains implicitly information about the interacting system that has not been used in the derivation of the dispersion relations. While the general discussions are based on selection rules derivable from invariance principles and the stability criterion that the rest mass of all intermediate states to which a single particle is coupled must exceed the rest mass of that particle, the perturbation expansions contain explicit lower limits on the mass of each particle in each intermediate state. For example, in the calculation of the nucleon form factor, each intermediate state coupled to the nucleon contains at least one particle with a mass equal to or greater than the nucleon mass (nucleon conservation). This is the decisive element which results in the validity of the perturbation-theoretical dispersion relations for the form factor and nucleon-nucleon scattering. We may point out here that nucleon conservation does not have such strong

* This work was supported in part by a grant from the National Science Foundation.

† National Science Foundation Post-Doctoral Fellow.

[1] Some references on the use of dispersion relations are Goldberger, Federbush, and Treiman (to be published); M. L. Goldberger and S. B. Treiman, Phys. Rev. **110**, 1178 (1958); J. Bernstein and M. L. Goldberger, Revs. Modern Phys. **30**, 465 (1958); Chew, Karplus, Gasiorowicz, and Zachariasen, Phys. Rev. **110**, 265 (1958); and Chew, Low, Goldberger, and Nambu, Phys. Rev. **106**, 1337 (1957).

[2] Bogoliubov, Medvedev, and Polivanov, Uspekhi Mat. Nauk (to be published); Bremermann, Oehme, and Taylor, Phys. Rev. **109**, 2178 (1958); R. Jost and H. Lehmann, Nuovo cimento **5**, 1598 (1957); F. Dyson, Phys. Rev. **110**, 1460 (1958).

[3] Y. Nambu, Nuovo cimento **6**, 1064 (1957).

[4] See, for instance, *Proceedings of the Seventh Annual Rochester Conference on High-Energy Nuclear Physics, 1957* (Interscience Publishers, Inc., New York, 1957), Session IV.