SOVIET PHYSICS JETP VOLUME 36(9), NUMBER 5 NOVEMBER, 1959

ON CAUSALITY IN A THEORY WITH AN INDEFINITE METRIC

D. A. SLAVNOV and A. D. SUKHANOV

Moscow State University

Submitted to JETP editor November 17, 1958

J. Exptl. Theoret. Phys. (U.S.S.R.) **36**, 1472-1479 (May, 1959)

The possibility of constructing a unitary and macroscopically causal scattering matrix in a theory with an indefinite metric is examined. The construction is carried out in the framework of perturbation theory by means of the interaction Lagrangians of the complete (physical plus sum of nonphysical) fields. By a special choice of the spectrum of the nonphysical fields it is possible to satisfy the requirements of unitarity and macroscopic causality in the second and third orders. These requirements cannot, however, be fulfilled together in the fourth order; thus within the framework of our assumptions it is not possible to construct a unitary and macroscopically causal scattering matrix in a theory with indefinite metric.

1. Recently much attention has been given to the ideas of Heisenberg, who has proposed the introduction of an indefinite metric in the space of the state amplitudes, for the purpose of eliminating divergences. This program leads to the appearance of a number of difficulties associated with the necessity of introducing "nonphysical" fields that have negative norms, and with violation of the condition that the scattering matrix be unitary. Several schemes have been proposed, however, that make it possible to eliminate the "nonphysical" states from the asymptotic expressions for observable quantities and to restore the unitary character of the scattering matrix.

The papers in question do not deal with the problems of causality, which certainly are of great interest. In the present paper we examine the question of the possibility of constructing a macroscopically causal theory with indefinite metric, on the basis of rather general postulates.

Following Heisenberg, we shall assume that the complete space H of state amplitudes decomposes into two orthogonal subspaces: H_1, the subspace of physical state amplitudes, and H_2, the subspace of "nonphysical" state amplitudes. As is done in reference 2, we introduce the operator P that projects the complete state amplitude onto the subspace H_1 of physical state amplitudes and is such that $P = P^+ = P^2$. We represent the complete field $\chi(x)$ in the form

$$\chi(x) = \varphi_0(x) + \sum_n c_n \varphi_{m_n}(x), \tag{1}$$

where $\varphi_0(x)$ is the physical field of mass m and $\varphi_{m_n}(x)$ is a "nonphysical" field of mass m_n, and impose the following commutation conditions:

$$\{\varphi_0(x),\ \varphi_0(y)\} = \mathcal{D}(x-y),$$

$$\{\varphi_{m_n}(x),\ \varphi_{m_n}(y)\} = \varepsilon_n \mathcal{D}_{m_n}(x-y),$$

where $\epsilon_n = \pm 1$. In accordance with this the bracket for the complete field $\chi(x)$ is

$$\{\chi(x),\ \chi(y)\} = D(x-y) = \mathcal{D}(x-y) + \Delta(x-y), \tag{2}$$

where

$$\Delta(x-y) = \sum_n \varepsilon_n c_n^2 \mathcal{D}_{m_n}(x-y).$$

In what follows we shall use the term "nonphysical bracket" to denote $\Delta(x-y)$. By a suitable choice of ϵ_n, c_n, and m_n we can always secure regularity of the function $D(x-y)$ on the light cone.

In references 1 — 3 the theory with indefinite metric has been constructed by the use of the concept of the scattering matrix in the complete space H. As has been shown in reference 4, within the framework of perturbation theory it is impossible to have any theory of such a sort that satisfies the condition of macroscopic causality. In the present paper we give up the idea of the scattering matrix in the complete space. Assuming, however, that the interaction occurs between the complete fields, we introduce the concept of the interaction Lagrangian of the complete fields.

2. The problem is to construct a scattering matrix $\tilde{S}$ connecting the asymptotic expressions for the physical state amplitudes of the subspace H_1, and which should be expansible in a series in the interaction constant

$$\widetilde{S} = 1 + \sum_n \widetilde{S}_n.$$

We shall use the hypothesis of adiabatic turning on and off of the interaction. Then $\widetilde{S}$ is a functional of the "turning-on intensity" $g(x)$.[5]

We impose the following conditions on $\widetilde{S}$:

1) Relativistic covariance.

2) Unitarity: $\widetilde{S}\widetilde{S}^+ = 1$.

3) Weak causality. We formulate this condition in the following way: suppose there are two non-overlapping space-time regions G_1 and G_2 in which the interaction is turned on with the respective intensities $g_1(x)$ and $g_2(y)$, and such that $x \lesssim y$. Then the difference

$$R_{12} = \widetilde{S}(g_1 + g_2) - \widetilde{S}(g_2)\widetilde{S}(g_1)$$

must go to zero with sufficient rapidity as the "distance" between the regions increases; the necessary degree of rapidity of this approach to zero will be discussed later.

4) Furthermore we require that $\widetilde{S}_n$ be a polynomial function of the complete-field Lagrangians and operators P; we emphasize that in the present paper we are using perturbation theory, and therefore we require that all the above conditions be fulfilled in each order independently.

It is easy to show that an $\widetilde{S}_n$ satisfying the first and fourth conditions is a polynomial function of operators P and S_i, where

$$1 + S_1 + S_2 + \ldots = T \exp\left\{i \int L(x) g(x) dx\right\} = S(g).$$

The matrix $S(g)$ has the form of the usual scattering matrix in the complete space. Without assigning to it any meaning, we shall hereafter use only its formal properties of unitarity, $SS^+ = 1$, and strict causality[5]

$$S(g_1 + g_2) = S(g_2) S(g_1).$$

3. We now proceed to the direct construction of a matrix $\widetilde{S}$ that satisfies the conditions listed above.

We shall first examine S correct to the second order. The most general form for $(\widetilde{S})_2$ is

$$(\widetilde{S})_2 = P(1 + S_1 + aS_2 + a_1 S_1 S_1 + a_2 S_1 P S_1) P. \qquad (3)$$

The coefficients of the first two terms are chosen from considerations of correspondence with classical theory.

From the unitarity condition we have

$$a = a^*, \qquad a_1^* + a_1 + 1 = 0, \qquad a_2^* + a_2 - 1 = 0. \qquad (4)$$

In order to make use of the causality condition we consider the expression

$$(R_{12})_2 = P[(a + a_1) S_1(g_2) S_1(g_1) + (a_2 - 1) S_1(g_2) P S_1(g_1) + a_1 S_1(g_1) S_1(g_2) + a_2 S_1(g_1) P S_1(g_2)] P. \qquad (5)$$

We must require that this quantity go to zero with increase of the "distance" between the regions G_1 and G_2. Let us expand this expression by Wick's theorem and examine the terms not containing contractions. These terms do not depend on the "distance" between the regions, and therefore their sum must be identically zero. From this consideration we get

$$a + 2a_1 + 2a_2 - 1 = 0. \qquad (6)$$

Then from Eqs. (4) and (6) we have

$$a = 1, \qquad a_1 = -\tfrac{1}{2} + i\alpha, \qquad a_2 = -(-\tfrac{1}{2} + i\alpha),$$

where α is an arbitrary real number. Therefore the right member of Eq. (5) can be rewritten in the form

$$P[(-\tfrac{1}{2} + i\alpha) S_1(g_2)(1 - P) S_1(g_1) + (\tfrac{1}{2} + i\alpha) S_1(g_1)(1 - P) S_1(g_2)] P. \qquad (7)$$

This expression can be handled conveniently by means of the following lemma, which will be proved in the Appendix.

Suppose that in the expression

$$P\Pi(P, \chi_1(x_{k_1}) \ldots \chi_n(x_{k_n}), \chi_1(y_{l_1}) \ldots \chi_m(y_{l_m})) P \qquad (*)$$

Π is a polynomial function of the operators P and χ, the x_{k_i} being points of the region G_1 and the y_{l_j} points of the region G_2. Then the necessary and sufficient condition for the expression (*) to be equal to a sum of terms each proportional to a "nonphysical contraction" depending on $(x_{k_i} - y_{l_j})$ is that it be possible to put (*) in the form of a sum of terms of the type

$$P\Pi_1(P, \ldots \chi_i(x_{k_i}) \ldots \chi_j(y_{l_j}) \ldots) \times P\Pi_2(P, \ldots \chi_i(x_{k_i}) \ldots)(1 - P)\; P\Pi_3(P, \ldots \chi_j(y_{l_j}) \ldots) \times P\Pi_4(P, \ldots \chi_i(x_{k_i}) \ldots \chi_j(y_{l_j}) \ldots) P. \qquad (8)$$

Returning to the expression (7), we see that according to the lemma it is equal to a sum of terms each proportional to a "nonphysical contraction" of fields from the regions G_1 and G_2. Thus a sufficient condition for causality in second order is that "nonphysical contractions" approach zero rapidly with increase of the "distance" between the regions.

On the other hand, if in the expansion of the expression (7) by Wick's theorem we take the terms that depend linearly on contractions, which by the lemma are necessarily "nonphysical," we can note that they are of different operator structures. And since only the contractions depend on the "distance" between the regions, the expression (7) will go to zero with increase of the "distance" only if the "nonphysical contractions" go to zero.

Thus the necessary and sufficient condition for the fulfillment of the causality condition for S in second order is that the "nonphysical contractions" of fields from the regions G_1 and G_2 go to zero with increase of the distance between these regions.

Let us examine the behavior of the contractions $\mathcal{D}(x-y)$ with increase of the "distance" between x and y. For simplicity we confine ourselves to scalar fields. Three cases can arise: 1) $\lambda < 0$, 2) $\lambda > 0$, 3) $\lambda = 0$, where $\lambda = (x^0 - y^0)^2 - (\mathbf{x}-\mathbf{y})^2$. From the explicit form of the contractions[5] we have in the first case the following asymptotic expression:

$$\mathcal{D}(x-y) \sim \sqrt{m}(-\lambda)^{3/4}\exp\{-m\sqrt{-\lambda}\}.$$

It is seen that for spacelike intervals $\mathcal{D}(x-y)$ falls off exponentially. In the second case the asymptotic expression has the form

$$\mathcal{D}(x-y) \sim \sqrt{m}\,\lambda^{-3/4}\exp\{\pm im\sqrt{\lambda}\},$$

i.e., for timelike intervals $\mathcal{D}(x-y)$ oscillates (the gradual falling off does not prevent the propagation of particles through macroscopic times).

As was first shown by Fierz[6] in a discussion of the first type of regular theory proposed by Heisenberg,[1] precisely this fact leads to the violation of macroscopic causality in second order, owing to the propagation of "nonphysical" particles through macroscopic times.

We can, however, try to change the asymptotic behavior of the "nonphysical contractions" and thus remove this objection. In fact, if instead of the discrete spectrum (1) we introduce a spectrum in which each "nonphysical" field is averaged over a suitably small range of masses with some weight factor, for example a Gaussian distribution, the contraction of a "nonphysical" field takes the form

$$\mathcal{D}_{m_n} = \int \frac{1}{M}\exp\left\{-\frac{(m-m_n)^2}{M^2}\right\}\mathcal{D}_m(x-y)\,dm,$$

and the asymptotic expression for the "nonphysical contraction" gets an exponentially decreasing factor because of the averaging of rapidly oscillating functions:

$$\Delta(x-y) \sim \sum_n \lambda^{-3/4}\sqrt{m_n}$$
$$\times\exp\{\pm im_n\sqrt{\lambda}\}\exp\left\{-\lambda\frac{M^2}{4}\right\}. \quad (9)$$

At the same time the regular behavior of the function $D(x-y)$ on the light cone is not destroyed. Thus by the use of a "smeared-out" spectrum of the "nonphysical" fields the causality condition can be fulfilled in second order also for timelike intervals.*

As has been shown by B. V. Medvedev (private communication), the violation of causality in second order in the case considered by Fierz is easy to understand if we go back to the connection of the theory with indefinite metric with nonlocal theory. In fact, in the theory with indefinite metric one is actually using a Pauli-Villars regularization with finite masses

$$\frac{1}{m^2-p^2-i\varepsilon} - \frac{1}{m_1^2-p^2-i\varepsilon} = \frac{1}{m^2-p^2-i\varepsilon}\,\frac{m_1^2-m^2}{m_1^2-p^2-i\varepsilon}$$

(for simplicity we consider one "nonphysical" field). The cut-off factor $(m_1^2-m^2)/(m_1^2-p^2-i\epsilon)$ can be referred not to the propagation function but to the vertex. Such a situation corresponds to a nonlocal theory with a factorizable form factor. The violation of causality in the case of a discrete spectrum of "nonphysical" fields is due to the fact that for the resulting form factor the Chretien-Peierls conditions[7] are not fulfilled, because such a form factor has a pole on the real axis of the squared momentum. Our way of introducing the "nonphysical" fields also receives a natural interpretation, since we in fact achieve the removal of the singularity by integrating a generalized function in a class of sufficiently smooth functions,[5] i.e., we achieve the fulfillment of the Chretien-Peierls conditions.

Thus by choice of the coefficients a, a_1, a_2 and of a special form of the spectrum of the "nonphysical" fields one can satisfy the conditions of unitarity and causality of $\tilde{S}$ in second order. We have finally for $(\tilde{S})_2$ the expression

$$(\tilde{S})_2 = P\,[1 + S_1 + S_2 + (-{}^1\!/_2 + i\alpha)\,S_1(1-P)\,S_1]\,P, \quad (10)$$

which contains one arbitrary real parameter α.

4. Before going on to further orders, we note that:

a) the contractions $\mathcal{D}(x-y)$ of the physical fields fall off slowly for timelike intervals (we cannot require fast falling off, since then there would be no propagation of physical particles through macroscopic times),

*We do not examine here the fulfillment of the causality condition for regions lying along the light cone. It can, however, be shown, by use of the Chretien-Peierls method,[7] that the causality condition can be fulfilled in this case also, if we introduce the "nonphysical" fields in the way that has been described. This result is not needed in our further argument.

b) the "nonphysical contractions" fall off exponentially both for spacelike and also for timelike intervals,

c) the lemma formulated above is valid. Therefore a necessary and sufficient condition for the fulfillment of the causality condition in the higher orders is that the differences $(R_{12})_n$ be sums of terms of the form (8).

The possibility of choosing only such terms leads in the third order of perturbation theory to the following expression for $(\tilde{S})_3$:

$$(\tilde{S})_3 = P\{(\tilde{S})_2 + S_3 + (-{}^1/_2 + i\alpha)[S_1(1-P)S_2 + S_2(1-P)S_1 - S_1(1-P)S_1(1-P)S_1]\}P. \quad (11)$$

It can be checked by direct verification that the expression (11) is unitary. We note that $(\tilde{S})_3$ depends on the parameter α introduced in the second order; i.e., in the third order the degree of arbitrariness has not increased.

5. Finally, let us examine $\tilde{S}$ in fourth order. Here, as in the third order, the causality condition demands that $(R_{12})_4$ be a sum of terms of the type (8); after simple but very cumbersome calculations this condition gives

$$\begin{aligned}(\tilde{S})_4 = P\{&(\tilde{S})_3 + S_4 + (-{}^1/_2 + i\alpha)[S_1(1-P)S_3\\ &+ S_2(1-P)S_2 + S_3(1-P)S_1 - S_1(1-P)S_1(1-P)S_2\\ &- S_1(1-P)S_2(1-P)S_1 - S_2(1-P)S_1(1-P)S_1\\ &+ S_1(1-P)S_1S_1(1-P)S_1]\\ &+ bS_1(1-P)S_1PS_1(1-P)S_1\}P,\end{aligned} \quad (12)$$

where in order to satisfy the requirements of causality we must have

$$b = {}^1/_4 - \alpha^2 - i\alpha. \quad (13)$$

For $(\tilde{S})_4$ the requirement of unitarity reduces to

$$b^* + b + \alpha^2 - {}^3/_4 = 0. \quad (14)$$

Comparing Eq. (13) with Eq. (14), we get $\alpha^2 + \frac{1}{4} = 0$; that is, the parameter α, which was introduced in the second order as a real quantity on the basis of the requirement of unitarity, must be purely imaginary.

Thus we arrive at an obvious contradiction: the conditions of unitarity and causality in fourth order are in conflict with the condition of unitarity in second order.

To get a clearer idea of the nature of the violation of causality, let us impose on the coefficient b only the condition (14) obtained from the unitarity condition for $(\tilde{S})_4$, but not the condition (13), and examine the structure of the terms that violate causality. These terms in $(S)_4$ are

$$\begin{aligned}P\{&({}^1/_2\alpha^2 + {}^1/_8 + i\alpha + i\beta)\, S_1(g_1)(1-P)\\ &\times S_1(g_1)PS_1(g_2)(1-P)S_1(g_2)\\ &+(-{}^1/_2\alpha^2 - {}^1/_8 + i\alpha + i\beta)S_1(g_2)(1-P)\\ &\times S_1(g_2)PS_1(g_1)(1-P)S_1(g_1)\}P,\end{aligned} \quad (15)$$

where β is an arbitrary real number.

From the structure of the expression (15) it is clear that it can be put in the form of a sum of terms each proportional to a "nonphysical contraction," but these contractions depend on the "distance" between points in the same region (G_1 or G_2). Since of course "distances" between points in the same region can be arbitrarily small, for the fulfillment of the condition of causality we have to require that the "nonphysical contractions" vanish for arbitrary "distances". In this case we practically arrive at the ordinary theory, since the contractions of the complete fields reduce to those of the physical fields, and as a result of the action of the operator P normal products of complete field operators are equal to normal products of the corresponding physical fields (see Appendix).

On the other hand, if we set $\alpha = -\beta$, then the expression (15) reduces to a sum of terms each proportional to a contraction of the physical fields, depending on the "distance" between points of the regions G_1 and G_2. By introducing a "smeared-out" spectrum of physical fields we can, of course, arrange matters so that their contractions fall off sufficiently rapidly with increase of the "distance" between the regions. But, as was noted above, this leads to the impossibility of the propagation of physical particles through macroscopic times.

6. Thus we have shown that even the weak causality condition is incompatible with the unitarity condition for the scattering matrix $\tilde{S}$ connecting the asymptotic state amplitudes of the physical subspace H_1, if it is constructed by means of the interaction Lagrangians of the complete fields.

This result is not unexpected, because also in nonlocal theory, with which the theory with indefinite metric is closely related, it has not been possible to reconcile the conditions of unitarity and causality.[8]

Naturally the question of the construction of a theory with indefinite metric without use of the idea of the complete-field Lagrangaian is at present still an open one.

In conclusion we emphasize once again that this discussion has all been within the framework of perturbation theory. Therefore we have not touched at all on the question of the possibility of compensation between violations of causality and unitarity in different orders. This extremely in-