ABC of instantons

A. I. Vaĭnshteĭn, V. I. Zakharov, V. A. Novikov, and M. A. Shifman

Institute of Theoretical and Experimental Physics, Moscow
and Institute of Nuclear Physics, Siberian Branch, USSR Academy of Sciences, Novosibirsk
Usp. Fiz. Nauk **136**, 553–591 (April 1982)

An attempt is made to present an instanton "calculus" in a relatively simple form. The physical meaning of instantons is explained by the example of the quantum-mechanical problem of energy levels in a two-humped potential. The nonstandard solution to this problem based on instantons is analyzed, and the reader is acquainted with the main technical elements used in this approach. Instantons in quantum chromodynamics are then considered. The Euclidean formulation of the theory is described. Classical solutions of the field equations (the Belavin-Polyakov-Shvarts-Tyupkin instantons) are obtained explicitly and their properties are studied. The calculation of the instanton density is described and the complete result is given for an arbitrary number of colors. The effects associated with fermion fields are briefly described.

PACS numbers: 03.70. + k, 11.10. – z, 12.35.Cn

CONTENTS

It appears that all fundamental interactions in nature are of the gauge type. The modern theory of hadrons—quantum chromodynamics (QCD)—is no exception. It is based on local gauge invariance with respect to the color group SU(3), which is realized by an octuplet of massless gluons. The idea of gauge invariance, however, is much older and derives from quantum electrodynamics, which was historically the first field model in which successful predictions were obtained. By the end of the forties, theoreticians had already learned how to calculate all observable quantities in electrodynamics in the form of series in $\alpha=1/137$. The first steps in QCD at the end of the seventies were also made in the framework of perturbation theory. However, it gradually became clear that, in contrast to electrodynamics, quark-gluon physics is not exhausted by perturbation theory. The most interesting phenomena—the confinement of colored objects and the formation of the hadron spectrum—are associated with nonperturbative (i.e., not describable in the framework of perturbation theory) effects, or rather, with the complicated structure of the QCD vacuum, which is filled with fluctuations of the gluon field.

It is now clear that the construction of the complete "wave function" of the vacuum is a very difficult problem. It still remains unsolved, despite numerous attacks by theoreticians. Nevertheless, quite a lot is already known. Study of the "old," traditional hadrons gives information about the fundamental properties of the vacuum. In turn, having obtained this information, we can make a number of nontrivial predictions about gluonium and other such poorly investigated aspects of hadron phenomenology.

The corresponding approach has been developed by the authors over a number of years, but it will not be discussed here. We note only that the main element is the introduction of several vacuum expectation values. For example, the intensity of gluon fields in vacuum is obviously measured by the quantity

$$\langle 0 \mid G^a_{\mu\nu} G^a_{\mu\nu} \mid 0 \rangle,$$

where $G^a_{\mu\nu}$ is the tensor of the intensity of the gluon field ($a=1,\ldots,8$ is the color index). Similarly, the quark condensate expectation value $\langle 0|\bar{q}q|0\rangle$ serves as a measure of the quark fields.

In a "final theory," if such is constructed, it will be possible to calculate all phenomenological matrix elements on the basis of the Lagrangian of QCD. It can

 0038-5670/82/040195-21$01.80

already be said that this will require knowledge of nonperturbative fluctuations in the physical vacuum. Here, phenomenology makes contact with the purely theoretical development, which as yet has not had great applications, though it has made it possible to reexamine a number of problems.

In 1975, there was discovered one of the most beautiful phenomena in quantum chromodynamics. We are referring to instantons—classical solutions of the field equations with nontrivial topology. The beauty of the theoretical constructions has attracted the interest of many physicists and mathematicians, and it is difficult to overestimate the popularity of instantons. The importance of instantons as the first example of fluctuations of the gluon field not encompassed by perturbation theory is in no doubt. Therefore, although one can hardly speak of any practical fruits, it appears appropriate to explain the physical essence of the phenomenon and derive the basic formulas to enable the reader to find his (or her) way about the literature.

One of the main conclusions which we shall attempt to establish is that the original Belavin–Polyakov–Shvarts–Tyupkin solution[1] (BPST instanton) is not the fluctuation which is dominant in the vacuum wave function. However, there is no danger of its beauty being wasted. In one form or another, it will certainly play a part in the future theory of strong interactions.

We begin with a simple quantum-mechanical problem that illustrates the role of nonperturbative fluctuations. This example was analyzed in detail by Polyakov,[2] who made a major contribution to the development of the entire subject. Having studied the main technical elements, we then turn to a more complicated case—quantum chromodynamics. At the very end, we discuss the question of the importance of the BPST instanton in the real world.

1. QUANTUM MECHANICS, IMAGINARY TIME, PATH INTEGRALS

In this section, we consider the problem of the one-dimensional motion of a spinless particle in a potential $V(x)$. This problem is usually treated in all textbooks on quantum mechanics, but we shall use a somewhat unusual method to solve it. The reader may find it inconvenient, just as sum rules are "inconvenient" for finding the eigenvalues of a Schrödinger equation. But—and this is the most important property—the method can be directly generalized to field theory.

If we take the mass of the particle equal to unity, $m=1$, then the Lagrangian of the system has the simple form

$$\mathscr{L}=\frac{1}{2}\left(\frac{dx}{dt}\right)^2-V(x). \tag{1}$$

Suppose that the particle at the initial time $(-t_0/2)$ is at the point x_i and at the final time $(+t_0/2)$ at the point x_f. An elegant method of expressing the amplitude of such a process was invented by Feynman (see the book of Ref. 3). The prescription is that the amplitude is equal to the sum over *all* paths joining the world points $(-t_0/2, x_i)$ and $(t_0/2, x_f)$ taken with weight

$$e^{i(\text{action})}.$$

The action, which we shall in what follows denote by the letter S, is related to the Lagrangian by

$$S=\int_{-t_0/2}^{t_0/2} dt\,\mathscr{L}(x,\dot{x}). \tag{2}$$

Thus, the transition amplitude is

$$\langle x_f \mid e^{-iHt_0} \mid x_i\rangle = N\int [Dx]\, e^{iS[x(t)]}, \tag{3}$$

where H is the Hamiltonian and $\exp(-iHt_0)$ is the ordinary evolution operator of the system. The factor N on the right-hand side is a normalization factor, to the discussion of which we shall return below. $[Dx]$ denotes integration over all functions $x(t)$ with boundary conditions $x(-t_0/2)=x_i$ and $x(t_0/2)=x_f$.

Before we consider dynamical questions, we examine the left-hand side. If we go over from states with a definite coordinate to states with a definite energy,

$$H\mid n\rangle = E_n \mid n\rangle,$$

then, obviously,

$$\langle x_f \mid e^{-iHt_0} \mid x_i\rangle = \sum_n e^{-iE_nt_0}\langle x_f \mid n\rangle\langle n \mid x_i\rangle, \tag{4}$$

and we obtain a sum of oscillating exponentials. If we are interested in the ground state (and in field theory we are always interested in the lowest state—the vacuum), it is much more convenient to transform the oscillating exponentials into decreasing exponentials. To this end, we make the substitution $t\to -i\tau$. Then in the limit $\tau_0\to\infty$ only a single term survives in the sum (4), and this directly tells us what are the energy E_0 and the wave function $\psi_0(x)$ of the lowest level $e^{-E_0\tau_0}\psi_0 \times(x_f)\psi_0^*(x_i)$.

In the literature, the transition to an imaginary time is frequently called the Wick rotation, and the corresponding variant of the theory of Euclidean variant. Below, we shall see that the substitution $t\to -i\tau$ is in a certain sense not only a matter of convenience, since it gives a new language for describing a very important aspect of the theory.

We now turn to the right-hand side of Eq. (3). In the Euclidean formulation, the action takes the form

$$iS[x(t)]\to \int_{-\tau_0/2}^{\tau_0/2}\left[-\frac{1}{2}\left(\frac{dx}{d\tau}\right)^2-V(x)\right]d\tau, \tag{5}$$

where we assume the boundary condition $x(-t_0/2)=x_i$, $x(\tau_0/2)=x_f$, and the origin of the energy is chosen such that $\min V(x)=0$.

We call

$$S_E=\int_{-\tau_0/2}^{\tau_0/2}\left[\frac{1}{2}\left(\frac{dx}{d\tau}\right)^2+V(x)\right]d\tau \tag{6}$$

the Euclidean action. Since $S_E\geqslant 0$, we have acquired an exponentially decreasing weight on the right-hand side of Eq. (3). In the present review, we shall remain in the Euclidean space and shall not return to the Minkowski space (i.e., to a real time); therefore, in all that follows we shall omit the subscript E.

The Euclidean variant of (3) is

$$\langle x_f | e^{-H\tau_0} | x_i \rangle = N \int [Dx]\, e^{-S}. \tag{7}$$

It is now time to make the next important step and explain what integration over all paths actually means. Let $X(\tau)$ be some function satisfying the boundary conditions. Then an *arbitrary* function with the same boundary conditions can be represented in the form

$$x(\tau) = X(\tau) + \sum_n c_n x_n(\tau), \tag{8}$$

where $x_n(\tau)$ is a complete set of orthonormal functions that vanish at the boundary:

$$\int_{-\tau_0/2}^{\tau_0/2} d\tau x_n(\tau) x_m(\tau) = \delta_{nm}, \quad x_n\left(\pm \frac{\tau_0}{2}\right) = 0.$$

The measure $[Dx]$ can be chosen in the form

$$[Dx] = \prod_n \frac{dc_n}{\sqrt{2\pi}}. \tag{9}$$

The coefficient of proportionality in this relation does not in general have in itself a particular meaning until the normalization factor N has been fixed.

Now suppose that in the problem under consideration the characteristic value of the action is large for certain reasons. Well known is the situation when the quasiclassical approximation, or, which is the same thing, the method of steepest descent (the latter, "mathematical" term may be more readily understood by some of the readers), "works." In other words, the entire integral in (7) is accumulated from regions near the extremum (minimum) of S. The path corresponding to the least action, which we denote by $X(\tau)$, is known in the literature as an extremal path, an extremal, or a stationary point. If there is one extremal and $S[X(\tau)] = S_0$, then

$$N \int [Dx]\, e^{-S} \sim e^{-S_0}. \tag{10}$$

Thus, to find the principal, *exponential factor* in the result, it is sufficient to put in information about a single, extremal path. (If there are several stationary points, we have in general the sum of the contributions of all the stationary points.)

There exists a standard procedure which enables us to take the next step and fix the pre-exponential factor. This operation is already somewhat more laborious. Suppose for simplicity that there is a single stationary point, $X(\tau)$. The following formula expressed in mathematical language the fact that $X(\tau)$ realizes a minimum of the action:

$$\delta S = S[X(\tau) + \delta x(\tau)] - S[X(\tau)] = \int_{-\tau_0/2}^{\tau_0/2} d\tau \delta x(\tau) \left[-\frac{d^2 X}{d\tau^2} + V'(X) \right] = 0,$$

where $V' = dV/dx$. The equation

$$\frac{d^2 X}{d\tau^2} = V'(X), \tag{11}$$

is of course well known to the reader from school days (we recall that "the mass multiplied by the acceleration is equal to the force"). It is the *classical* equation of motion of a particle in the potential *minus* $V(x)$.[1)]

[1)]The minus sign is due to the fact that the Euclidean formulation is considered [see (6)].

We shall shortly return to this circumstance, but first recall how the pre-exponential factor in (10) is calculated. It is determined by an entire "pencil" of paths near the extremal path, i.e., by paths with action that differs little from S_0. In other words, we take into account only the quadratic deviation:

$$S[X(\tau) + \delta x(\tau)] = S_0 + \int_{-\tau_0/2}^{\tau_0/2} d\tau \delta x \left[-\frac{1}{2} \frac{d^2}{d\tau^2} \delta x + \frac{1}{2} V''(X)\, \delta x \right] \tag{12}$$

(as the reader will recall, there is no term linear in the deviation).

Suppose we know a complete set of eigenfunctions and eigenvalues of the equation

$$-\frac{d^2}{d\tau^2} x_n(\tau) + V''(X)\, x_n(\tau) = \varepsilon_n x_n(\tau). \tag{13}$$

Then we can choose these functions as the orthonormalized system which occurs in (8), and the action (12) is transformed to the simple *diagonal* form

$$S = S_0 + \frac{1}{2} \sum_n \varepsilon_n c_n^2.$$

Recalling the definition (9) and the rule of Gaussian integration

$$\int_{-\infty}^{+\infty} dc \exp\left(-\frac{1}{2} \varepsilon c^2 \right) = \frac{\sqrt{2\pi}}{\sqrt{\varepsilon}}$$

(it is important that after the diagonalization each such integration can be performed independently of the others), we obtain

$$\langle x_f | e^{-H\tau_0} | x_i \rangle = e^{-S_0} N \prod_n \varepsilon_n^{-1/2}. \tag{14}$$

Sometimes, instead of the product of eigenvalues one uses the notation

$$\prod_n \varepsilon_n^{-1/2} = \left[\det \left(-\frac{d^2}{d\tau^2} + V''(X(\tau)) \right) \right]^{-1/2}, \tag{15}$$

which, of course, derives from the theory of ordinary finite-dimensional matrices. In fact, the relation (15) can be regarded as the definition of the determinant of a differential operator. It is here appropriate to make three comments. First, the result (14) does not depend on the explicit form of the eigenfunctions but only on the eigenvalues. Second, we have assumed that all the ε_n are positive. In the majority of cases, this is so, but in the instanton example, which is the final aim of the present review, several eigenvalues vanish. The resulting infinity has a simple physical meaning. The problem of how it should be handled is the subject of the next section. The third and final comment is the following. The normalization factor N has still not yet been fixed. We shall not even attempt to give a general prescription but consider a simple example, which will serve us in the future too. Suppose the original particle with mass $m = 1$ is placed in the potential $V(x)$ shown in Fig. 1. We do not need the actual form of this potential, but to achieve "normalization" to the harmonic oscillator (in which the potential is usually taken to be $m\omega^2 x^2/2$), we set $V''(x=0) = \omega^2$. As the initial and final points of the motion we choose $x_i = x_f = 0$.

The rich physical intuition that we each have for potential mechanical motion enables us to find the extremal from Eq. (11) without knowing the explicit form of

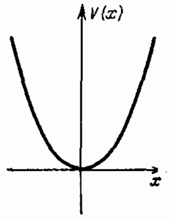


FIG. 1.

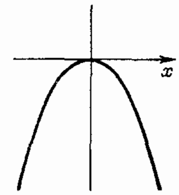


FIG. 2.

$V(x)$. Indeed, this equation describes the motion of a ball on the profile shown in Fig. 2. At the time $-\tau_0/2$, the ball is displaced from the upper point, to which it returns at the time $+\tau_0/2$. It is entirely clear that there exists only one path with such properties: $X(\tau)\equiv 0$. Any other path corresponds to an infinite motion with the ball going away to plus or minus ∞. It is also clear that the action on the path $X(\tau)=0$ vanishes.

Thus, in the given particular problem the general formula (14) becomes

$$\langle x_f=0\,|\,e^{-H\tau_0}\,|\,x_1=0\rangle = N\Big[\det\Big(-\frac{d^2}{d\tau^2}+\omega^2\Big)\Big]^{-1/2}(1+\text{following terms}),$$

and all the eigenvalues ε_n are immediately fixed by the boundary conditions $x_n(\pm\tau_0/2)=0$:

$$\varepsilon_n=\frac{\pi^2n^2}{\tau_0^2}+\omega^2, \qquad n=1,\,2,\,\ldots\,.$$

We have now arrived at the point at which it is possible to advance further without saying what is the value of N. We split the determinant into two brackets:

$$N\Big[\det\Big(-\frac{d^2}{d\tau^2}+\omega^2\Big)\Big]^{-1/2}=\Big[N\Big(\prod_{n=1}^{\infty}\frac{\pi^2n^2}{\tau_0^2}\Big)^{-1/2}\Big]\Big[\prod_{n=1}^{\infty}\Big(1+\frac{\omega^2\tau_0^2}{\pi^2n^2}\Big)\Big]^{-1/2}. \tag{16}$$

Obviously, the first square brackets corresponds to *free* motion of the particle, and therefore, it must, of course, reproduce the free result:

$$N\Big(\prod_{n=1}^{\infty}\frac{\pi^2n^2}{\tau_0^2}\Big)^{-1/2}=\langle x_f=0\,|\,e^{-\hat{p}^2\tau_0/2}\,|\,x_1=0\rangle=\sum_n|\langle p_n\,|\,x=0\rangle|^2e^{-p_n^2\tau_0/2}$$
$$=\int_{-\infty}^{+\infty}\frac{dp}{2\pi}e^{-p^2\tau_0/2}=\frac{1}{\sqrt{2\pi\tau_0}}. \tag{17}$$

Of course, Eq. (17) is somewhat symbolic, but it can be regarded as the definition of the normalization factor N. We now consider the second, less trivial brackets. For the infinite product which occurs in it we have the well-known formula [see, for example, formula (1.431.2) in Ref. 4][2)]

$$\pi y\prod_{n=1}^{\infty}\Big(1+\frac{y^2}{n^2}\Big)=\operatorname{sh}\pi y,$$

where in our case $y=\omega\tau_0/\pi$.

We now collect together all the factors, take into account (16) and (17), and write down the final result:

$$\langle x_f=0\,|\,e^{-H\tau_0}\,|\,x_1=0\rangle=N\Big[\det\Big(-\frac{d^2}{d\tau^2}+\omega^2\Big)\Big]^{-1/2}$$
$$=\frac{1}{\sqrt{2\pi\tau_0}}\Big(\frac{\operatorname{sh}\omega\tau_0}{\omega\tau_0}\Big)^{-1/2}=\Big(\frac{\omega}{\pi}\Big)^{1/2}(2\operatorname{sh}\omega\tau_0)^{-1/2}. \tag{18}$$

Going to the limit $\tau_0\to\infty$, we find

$$\langle x_f=0\,|\,e^{-H\tau_0}\,|\,x_1=0\rangle\xrightarrow[\tau_0\to\infty]{}\Big(\frac{\omega}{\pi}\Big)^{1/2}e^{-\omega\tau_0/2}\Big(1+\frac{1}{2}e^{-2\omega\tau_0}+\ldots\Big),$$

from which it follows that for the lowest state $E_0=\omega/2$, $[\psi_0(0)]^2=(\omega/\pi)^{1/2}$. The next term in the expansion corresponds to the level of an oscillator with $n=2$ [the odd n do not contribute, since for them $\psi_n(0)=0$]. The results are exact for the harmonic oscillator and serve as a zeroth approximation for a potential with small anharmonicity, say $(\omega^2/2)x^2+\lambda x^4$.

2. TWO-HUMPED POTENTIAL. TUNNELING

In the previous section, we reformulated in the language of Euclidean space and path integrals one of the most fundamental problems—an oscillator system near the equilibrium position. This problem provides the basis of all field theory. In fact, we have taken into account small vibrations—small deviations from the equilibrium position—and have made the first step to ordinary perturbation theory. For more than 20 years, right up to the middle of the seventies, all field-theoretical models (apart from the small exception of exactly solvable two-dimensional models) were developed in this and only this direction. The field variables were regarded as a system of an infinitely large number of oscillators coupled to each other and each possessing zero-point vibrations; one then considered small deviations, with respect to which perturbation theory was constructed successively. In this sense, the "infant" period of quantum chromodynamics, when quark–gluon perturbation theory was created, did not introduce anything fundamentally new. It was only the discovery of instantons which showed that it contains effects which cannot be described if one does not go beyond the framework of small deviations from the equilibrium position. It is in principle impossible to describe these effects by expansions in the coupling constant. Here, we again turn to a simple quantum-mechanical analogy, in which, however, all the main features are already present.

Thus, we again consider the one-dimensional potential motion of a spinless particle with unit mass. The potential

$$V(x)=\lambda(x^2-\eta^2)^2 \tag{19}$$

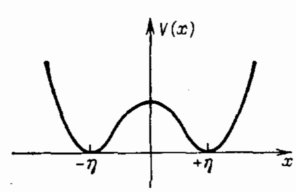


FIG. 3.

---

2) *Translator's Note.* The Russian notation for the trigonometric, inverse trigonometric, hyperbolic trigonometric functions, etc., is retained here and throughout the article in the displayed equations.

is shown in Fig. 3. We fix the parameters λ and η in such a way that $8\lambda\eta^2=\omega^2$, where ω is the frequency introduced in the previous section. Then near each of the minima, which are indicated by the symbols $\pm\eta$, the curve is identical to the potential of the previous section. If $\lambda \ll \omega^3$, then the wall separating the two minima is high. Its height is $\omega^4/64\lambda$. Suppose for a moment that it is actually equal to infinity. Then the lower state of the system has a twofold degeneracy—the particle may be in the right-hand well or in the identical left-hand well, i.e., it executes small vibrations near the point $+\eta$ or $-\eta$. At first glance, the solution to our problem should be constructed in exactly the same way. The expectation value of the coordinate in the ground state should be

$$\langle x\rangle_0 = +\eta\,(1 + \text{corrections}) \quad \text{or} \quad \langle x\rangle_0 = -\eta\,(1 + \text{corrections}),$$

the original symmetry of the system with respect to the substitution $x \to -x$ is broken, $E_0=(\omega/2)(1+\text{corrections})$ in both cases, and at small λ the corrections are small. In fact, it is known from courses of quantum mechanics that this picture is *qualitatively* incorrect. The symmetry is *not* broken, the expectation value of x for the ground level is *exactly* zero, and there is *no* degeneracy:

$$E_0=\frac{\omega}{2}-\sqrt{\frac{2\omega^3}{\pi\lambda}}\,e^{-\omega^3/12\lambda}\,\frac{\omega}{2},$$
$$E_1=\frac{\omega}{2}+\sqrt{\frac{2\omega^3}{\pi\lambda}}\,e^{-\omega^3/12\lambda}\,\frac{\omega}{2}. \tag{20}$$

We note the fact that $E_1-E_0\sim\exp(-\omega^3/12\lambda)$ and this quantity cannot be expanded in a series in λ. [It is assumed that $\omega^3/\lambda \gg 1$. In reality, Eqs. (20) begin to "work" when $\omega^3/12\lambda \gtrsim 6$.]

Thus, we have gone wrong and failed to take into account an important element that leads to qualitative changes. What is this element? Everyone knows the standard answer given in courses of quantum mechanics. If at the initial time the particle is concentrated in, say, the left-hand minimum, it nevertheless feels the existence of the right-hand well despite the fact that the latter is inaccessible according to the classical laws of motion. Quantum-mechanical tunneling transfers the wave function from one well to the other and, in Polyakov's terminology, "mixes" the ground states. The correct wave function of the ground state is an even superposition of the wave functions in each of the wells.

We now consider how this phenomenon appears in the imaginary time and how the technique presented in the previous section is changed. It turns out—and this is a great good fortune—that all the fundamental technical elements remain unchanged. It is only necessary to take into account the fact that the classical equations of motion in the imaginary time have not only the trivial solutions $X(\tau)=\text{const}$ considered earlier but also additional nontrivial topological solutions which extend far from both the minima. These solutions connect the points $\pm\eta$, and they are entirely responsible for the phenomenon under discussion. We emphasize that in real time there are no additional classical solutions, since the transition from the one minimum to the other occurs below the barrier and is classically forbidden.

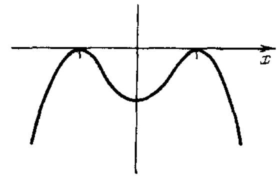


FIG. 4.

The solutions arise only after the Euclidean rotation.

We consider the calculation of the amplitudes

$$\langle\eta\,|\,e^{-H\tau_0}\,|-\eta\rangle \text{ and } \langle\eta\,|\,e^{-H\tau_0}\,|\,\eta\rangle.$$

The first step, as the reader may still recall, consists of solving Eq. (11). The "mechanical profile" for this equation is shown in Fig. 4. We are interested in solutions of Eq. (11) that have finite action in the limit $\tau_0 \to \infty$, since it is such solutions that are important in the quasiclassical approximation we are discussing. Most of the paths correspond to either vibrational motion or to $x\to\infty$ as $\tau\to\infty$, and they have infinite action.

A finite action in the limit $\tau_0\to\infty$ is obviously obtained when the particle stays at the top of a hump, i.e., $X(\tau)=\eta$ and $X(\tau)=-\eta$. The contribution of these trajectories was considered above. Another interesting motion leading to a finite action as $\tau_0\to\infty$ corresponds to the particle sliding from one hump and stopping on the other. Thus, we are interested in a path which begins at $-\tau_0/2$ at the point $-\eta$ and ends at the point η at the time $\tau_0/2$.[3)] Physical intuition suggests that such trajectories exist, though their explicit form for finite τ_0 is complicated. We are always interested in only the lowest state, and therefore we can directly assume that $\tau_0\to\infty$. In this limit, the solution is very simple:

$$X(\tau)=\eta\,\text{th}\,\frac{\omega(\tau-\tau_c)}{2} \tag{21}$$

[it corresponds to mechanical motion with zero energy, $E=(1/2)\dot{x}^2-V(x)=0$, so that the equations can be readily integrated].

Such a solution is called an instanton (Polyakov proposed the name "pseudoparticle," which can also be found in the literature); the arbitrary parameter τ_c indicates its center. Of course, there also exist anti-instantons, which begin at $+\eta$ and end at $-\eta$. They are obtained from (21) by the substitution $\tau\to-\tau$.

Since all the integrals can be calculated, it is easy to obtain a closed expression for the action of the instanton [we recall that for the instanton $\frac{1}{2}\dot{x}^2=V(x)$]:

$$S_0=S\,[X(\tau)]_{\text{inst}}=\int_{-\infty}^{+\infty}d\tau\dot{X}^2=\frac{\omega^3}{12\lambda}. \tag{22}$$

We recall that the principal exponential factor in the amplitude is $e^{-\text{action}}$ [see Eq. (10)]. The exponential which occurs in (20) has emerged. Of course, we still have a long way to go before we can reproduce the complete answer.

3)Here we have allowed a slight inaccuracy. If τ_0 is large but not infinite, the path begins just to the right of $-\eta$ and ends just to the left of $+\eta$. It is only in the limit $\tau_0\to\infty$ that the end points coincide with $\pm\eta$.

We draw attention to one further property of an instanton, which has far reaching consequences. The center of the solution may be at any point, and the action of the instanton does not depend on the position of the center. This circumstance obviously reflects the symmetry of the original problem. Namely, the Lagrangian of the system is invariant with respect to shifts in time, and the time origin can be chosen arbitrarily. Each concrete solution (21) has a definite position with respect to the origin, and thus there exists an infinite family of solutions distributed arbitrarily with respect to the origin. Intuitively, it is clear that the instanton must occur in any physical quantity in the form of an integral over the position of its center. How does this integral arise formally and what weight is then obtained? Answers to these questions are given in the following section.

3. DETERMINANT AND ZERO-FREQUENCY MODES

In this section, we find the one-instanton contribution to $\langle -\eta | e^{-H\tau_0} | \eta \rangle$. We shall not, of course, be concerned with the exponential factor, which has actually already been found, but rather the pre-exponential factor, whose calculation presents a more laborious problem. It is true that in the case under consideration one can employ various devices that significantly simplify the problem and are sometimes discussed in the literature.[5] However, we shall proceed in a "head on" manner, which is closest to the method used by 't Hooft[6] to calculate the instanton determinant in QCD. We hope that this will subsequently enable the reader to reproduce for himself all details of 't Hooft's work, which is central for the entire instanton problem.

The original formula (14) is conveniently rewritten as

$$\langle -\eta | e^{-H\tau_0} | \eta \rangle = N\left[\det\left(-\frac{d^2}{d\tau^2}+\omega^2\right)\right]^{-1/2}\left\{\frac{\det[-(d^2/d\tau^2)+V''(X)]}{\det[-(d^2/d\tau^2)+\omega^2]}\right\}^{-1/2} e^{-S_0}(1+\text{corrections}).$$

We have multiplied and divided by a known number—the determinant for the harmonic oscillator [see (18)]. The harmonic oscillator will serve as a "point of reference" for manipulations with the more complicated determinant in the numerator. Substituting the explicit expression $X(\tau)=\eta\tanh(\omega\tau/2)$ in $V''(X)$, we arrive at the eigenvalue equation

$$-\frac{d^2}{d\tau^2}x_n(\tau)+\left(\omega^2-\frac{3}{2}\omega^2\frac{1}{\mathrm{ch}^2(\omega\tau/2)}\right)x_n(\tau)=\varepsilon_n x_n(\tau). \tag{23}$$

It can be regarded as a certain Schrödinger equation, which, fortunately, is very well studied. Indeed, Eq. (23) is described in detail in, for example, the textbook of Landau and Lifshitz (Ref. 7, pp. 97 and 105), and we shall use this source. We recall that the boundary conditions are $x_n(\pm\tau_0/2)=0$ and $\tau_0\to\infty$. These conditions are automatically satisfied with exponential accuracy for bound levels, i.e., for the truly discrete spectrum.[4)] There are two such levels in Eq. (23). One of them corresponds to the eigenvalue $\varepsilon_1=(3/4)\omega^2$, and the other to

$$\varepsilon_0=0.$$

The wave function of the latter, normalized to unity, is

$$x_0(\tau)=\sqrt{\frac{3\omega}{8}}\frac{1}{\mathrm{ch}^2(\omega\tau/2)}. \tag{24}$$

The vanishing of the eigenvalue may discourage the reader, since the answer contains $\varepsilon_n^{-1/2}$! However, this result, $\varepsilon_0=0$, cannot be regarded as a surprise. Indeed, Eq. (23) actually describes the response of the dynamical system under consideration to small perturbations imposed on $X(\tau)$. Since $X(\tau)$ is a solution which realizes a "local" minimum of the action, a perturbation of $X(\tau)$ increases the action. Accordingly, the ε_n are positive. However, we already know that there is one direction in the function space along which the solution can be perturbed without changing the action. We have in mind a shift of the center. By virtue of the translational invariance,

$$S[X(\tau,\tau_c)]-S[X(\tau,\tau_c+\delta\tau_c)]=0.$$

The so-called zero-frequency mode (i.e., the mode with $\varepsilon=0$) is obviously proportional to $X(\tau,\tau_c)=X(\tau,\tau_c+\delta\tau_c)$. The correctly normalized zero-frequency mode has the form

$$x_0(\tau)=S_0^{-1/2}\left(-\frac{d}{d\tau_c}\right)X(\tau,\tau_c),$$

or, which is the same thing,

$$x_0(\tau)=S_0^{-1/2}\frac{d}{d\tau}X(\tau). \tag{25}$$

The correctness of the normalization follows from the expression (22). It is readily seen that (25) is identical to (24), and we now see that this agreement is not fortuitous but a consequence of the translational invariance.

Thus, integration with respect to the coefficient c_0 corresponding to the zero-frequency mode [see (8) and (9)] is non-Gaussian, and the integral between infinite limits does not exist at all. The way out of the dilemma is simple. We shall not calculate this integral explicitly. It is clear that the integration over dc_0 is the same as integration over $d\tau_c$ apart from a coefficient of proportionality. We have here the same integral over the position of the center of the instanton whose appearance our intuition required. In the literature, this trick is sometimes called the introduction of a collective coordinate.

We determine the coefficient of proportionality. If c_0 changes by Δc_0, then $x(\tau)$ changes by

$$\Delta x(\tau)=x_0(\tau)\Delta c_0$$

[see (8)]. On the other hand, the change $\Delta x(\tau)$ on a shift $\Delta\tau_c$ of the center is

$$\Delta x(\tau)=\Delta X(\tau)=\frac{dx}{d\tau_c}\Delta\tau_c=-\sqrt{S_0}\,x_0(\tau)\Delta\tau_c.$$

Equating the two increments, we obtain

$$dc_0=\sqrt{S_0}\,d\tau_c. \tag{26}$$

[In Eq. (26), we have not inserted the minus sign to en-

4) Without boundary conditions, the complete spectrum is in fact discrete. The genuine discrete levels can however be readily distinguished from the quasidiscrete levels formed from the continuum after the system has been enclosed in the "box" $x(\pm\tau_0/2)=0$. The former are separated by intervals of order ω^2, while the latter are at a distance of order $1/\tau_0^2$ from their neighbors.

sure that as c_0 varies from $-\infty$ to $+\infty$ the parameter τ_c changes in the same interval.] This is not yet everything, since we agreed to normalize the result to the ordinary oscillator (we recall that we are interested in the ratio of determinants). In the oscillator problem, the minimal eigenvalue is $\omega^2+\pi^2/\tau_0^2 \to \omega^2$ in the limit $\tau_0\to\infty$. Finally,

$$\left\{\frac{\det[-(d^2/d\tau^2)+V''(X)]}{\det[-(d^2/d\tau^2)+\omega^2]}\right\}^{-1/2}=\sqrt{\frac{S_0}{2\pi}}\,\omega\,d\tau_c\left\{\frac{\det'[-(d^2/d\tau^2)+V''(X)]}{\omega^{-2}\det[-(d^2/d\tau^2)+\omega^2]}\right\}^{-1/2}, \tag{27}$$

where det′ denotes the reduced determinant with the zero-frequency mode removed.

We emphasize that although we have analyzed only a single specific example with the simplest instanton $\eta\tanh(\omega\tau/2)$, the method of dealing with zero-frequency modes is in fact general. Thus, in the BPST instanton any invariance will generate a zero-frequency mode, and the integration with respect to the corresponding coefficient must be replaced by integration with respect to some collective variable. We have already learned how to find the Jacobian of the transformation.

We now consider positive-frequency modes. It is easiest to deal with the second discrete level, whose eigenvalue is $(3/4)\omega^2$. If we denote by Φ the ratio

$$\Phi=\frac{\det'[-(d^2/d\tau^2)+V''(X)]}{\omega^{-2}\det[-(d^2/d\tau^2)+\omega^2]}, \tag{28}$$

then the contribution of this level to Φ as $\tau_0\to\infty$ is obviously

$$\frac{3}{4}. \tag{29}$$

We now turn to other modes with $\varepsilon>\omega^2$. If we did not have the boundary condition $x(\pm\tau_0/2)=0$, Eq. (23) in this region would have a continuous spectrum. Let us forget the boundary conditions for a moment. The general solution of (23) is given in the book of Landau and Lifshitz; however, we do not require its explicit form. It is sufficient to know the following. First, the solutions with $\varepsilon>\omega^2$ are labeled by a continuous index p. This index is related to the eigenvalue ε by $p=\sqrt{\varepsilon_p-\omega^2}$ and ranges over the entire inverval $(0,\infty)$. Second, for the values of the parameters that occur in (23) there is no reflection. In other words, choosing one of the linearly independent solutions in such a way that

$$x_p(\tau)=e^{ip\tau}\quad \text{as}\quad \tau\to+\infty,$$

we have in the other asymptotic region the same exponential:

$$x_p(\tau)=e^{ip\tau+i\delta_p}\quad \text{as}\quad \tau\to-\infty.$$

The second exponential, $e^{-ip\tau}$, which should in principle arise, is absent, and the entire dynamical effect has been reduced to the phase

$$e^{i\delta_p}=\frac{1+(ip/\omega)}{1-(ip/\omega)}\,\frac{1+(2ip/\omega)}{1-(2ip/\omega)} \tag{30}$$

(we have used here the formula from the textbook of Ref. 7 on p. 106). The second linearly independent solution can be chosen in the form $x_p(-\tau)$. The general solution is $Ax_p(\tau)+Bx_p(-\tau)$, where A and B are arbitrary constants.

This information is already sufficient to find the spectrum if we recall the boundary condition $x(\pm\tau/2)=0$. The equations for A and B,

$$Ax_p\left(\frac{\tau_0}{2}\right)+Bx_p\left(-\frac{\tau_0}{2}\right)=0,\quad Ax_p\left(-\frac{\tau_0}{2}\right)+Bx_p\left(\frac{\tau_0}{2}\right)=0,$$

have nontrivial solutions if and only if

$$\frac{x_p(\tau_0/2)}{x_p(-\tau_0/2)}=\pm 1.$$

This gives an equation for p:

$$e^{ip\tau_0-i\delta_p}=\pm 1,$$

or, which is the same thing,

$$p\tau_0-\delta_p=\pi n,\qquad n=0,1,\ldots. \tag{31}$$

We denote the nth solution by $\tilde p_n$. In the case of $\det[-(d^2/d\tau^2)+\omega^2]$, by which we normalize, the equation is $p\tau_0=\pi n$ and the nth solution $p_n=\pi n/\tau_0$. We need to calculate the product[5)]

$$\prod_{n=1}^{\infty}\frac{\omega^2+\tilde p_n^2}{\omega^2+p_n^2}.$$

For any preassigned n, the ratio $(\omega^2+\tilde p_n^2)/(\omega^2+p_n^2)$ is arbitrarily close to unity as $\tau_0\to\infty$. Only the multiplication of a very large number of factors with $n\sim\omega\tau_0$, each of them differing from 1 by an amount of order $1/\omega\tau_0$, gives an effect. (For $n\gg\omega\tau_0$, the difference between $\omega^2+\tilde p_n^2$ and $\omega^2+p_n^2$ again becomes unimportant, in complete agreement with our physical intuition.) Under these conditions, we can write

$$\prod\frac{\omega^2+\tilde p_n^2}{\omega^2+p_n^2}=\exp\left(\sum_n\ln\frac{\omega^2+\tilde p_n^2}{\omega^2+p_n^2}\right)\approx\exp\left[\sum_n\frac{2p_n(\tilde p_n-p_n)}{\omega^2+p_n^2}\right],$$

where we have made an expansion with respect to the small difference $\tilde p_n-p_n$. Going over from summation over n to integration over p_n and using (31) for $\tilde p_n-p_n$, we obtain on the right-hand side

$$\exp\left[+\frac{1}{\pi}\int_0^\infty\frac{\delta_p\cdot 2p\,dp}{p^2+\omega^2}\right]=\exp\left[-\frac{1}{\pi}\int_0^\infty\frac{d\delta_p}{dp}\ln\left(1+\frac{p^2}{\omega^2}\right)dp\right].$$

Differentiating the phase by means of (30) and introducing the dimensionless variable $y=p/\omega$, we transform this expression identically to

$$\exp\left[-\frac{2}{\pi^2}\int_0^\infty dy\left(\frac{1}{1+y^2}+\frac{2}{1+4y^2}\right)\ln(1+y^2)\right]=\frac{1}{9}. \tag{32}$$

Finally, combining (32) and (29), we find that

$$\Phi=\frac{1}{12}. \tag{33}$$

We have now made all the necessary preparations, namely, we have derived formulas (33), (28), (27), (22), and (18), and we write down the result for the one-instanton contribution:

$$\langle-\eta|e^{-H\tau_0}|\eta\rangle_{\text{one-inst}}=\left(\sqrt{\frac{\omega}{\pi}}e^{-\omega\tau_0/2}\right)\left(\sqrt{\frac{6}{\pi}}\sqrt{S_0}\,e^{-S_0}\right)\omega\,d\tau_c. \tag{34}$$

---

5) The reader may recall that we have already "taken up" in the denominator two eigenvalues, $\omega^2+\pi^2/\tau_0^2$ and $\omega^2+4\pi^2/\tau_0^2$, in calculating the contribution of the discrete modes with $\varepsilon=0$ and $\varepsilon=3\omega^2/4$. Therefore, it would be more correct in the denominator to write $\omega^2+p_{n+2}^2$. However, as we shall see very shortly, it is the region of very large n, of order $\omega\tau_0$, that is important, so that the difference between p_{n+2} and p_n is immaterial.

This result can be trusted as long as

$$\sqrt{S_0}\, e^{-S_0}\omega\tau_0 \ll 1.$$

At large τ_0, when this condition is violated, it is necessary to take into account paths constructed from many instantons and anti-instantons, and this will be done in the following section.

It is here appropriate to make some comments. The factor in the first square brackets corresponds to a simple harmonic oscillator. By separating it, we have been able to normalize, or regularize, the instanton calculations. A similar device for regularization is used in quantum chromodynamics. The factor in the second square brackets can naturally be called the instanton density. Besides the exponential factor e^{-S_0}, the density contains the pre-exponential $\sqrt{S_0}$, which is associated with the existence of the zero-frequency mode. This circumstance is also of a general nature. In quantum chromodynamics too, each zero-frequency mode is associated with $\sqrt{S_0}$. Finally, the existence of the zero-frequency mode leads to the appearance of a regularization frequency and of integration over the collective coordinate $\omega d\tau_c$.

We wish to emphasize that it is worth remembering the lessons we have learned, since they can be directly transferred to the BPST instanton. The only thing specific in the present case is the number $-\sqrt{6/\pi}$. If this number is not particularly important (and in QCD, as we shall see below, this is indeed the case), all the remaining result can be reconstructed almost at once, without calculations. We have taken so much bother with the relatively simple determinant for a pedagogical reason—to avoid greater boredom in the case of the BPST instanton.

4. INSTANTON GAS

It remains for us to make the final, small step to reproduce formula (20). The energy of the lower state is determined by the transition to the limit $\tau_0 \to \infty$. We cannot go to this limit directly in Eq. (34). At very large τ_0, paths constructed of many instantons and anti-instantons are important. If the distance between their centers is large, such a path is also a classical solution.

Suppose we have n instantons or anti-instantons with centers $\tau_1, \tau_2, \ldots, \tau_n$ (Fig. 5). The points τ_i satisfy the condition

$$-\frac{\tau_0}{2} < \tau_1 < \tau_2 < \ldots < \tau_n < \frac{\tau_0}{2},$$

and otherwise can be distributed arbitrarily. If the characteristic intervals satisfy $|\tau_i - \tau_j| \gg \omega^{-1}$ (we shall verify the condition *a posteriori*), then the action corresponding to such a configuration is nS_0, where S_0 is the action of one instanton. With regard to the determinant, it is obvious that if we did not have the n narrow transition regions (near $\tau_1, \tau_2, \ldots, \tau_n$) we should obtain the same result as in the case of the harmonic oscillator, $\sqrt{\omega/\pi}\, e^{-\omega\tau_0/2}$. The transition regions lead to a correction, and we now know in what way:

$$\sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2} \to \sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2} \left(\sqrt{\frac{6}{\pi}}\sqrt{S_0}\, e^{-S_0}\right)^n \prod_{1}^{n} (\omega\, d\tau_i).$$

Finally, the contribution of the n-instanton configuration can be written in the form

$$\sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2} d^n \int_{-\tau_0/2}^{\tau_0/2} \omega\, d\tau_n \int_{-\tau_0/2}^{\tau_n} \omega\, d\tau_{n-1} \ldots \int_{-\tau_0/2}^{\tau_2} \omega\, d\tau_1 = \sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2} d^n \frac{(\omega\tau_0)^n}{n!},$$

where we have denoted by d the instanton density,

$$d = \sqrt{\frac{6}{\pi}}\sqrt{S_0}\, e^{-S_0}. \tag{35}$$

The amplitudes $\langle -\eta | e^{-H\tau_0} | \eta\rangle$ and $\langle \eta | e^{-H\tau_0} | \eta\rangle$ are obtained by summation over n. In the first case, we start from $-\eta$ and arrive at $+\eta$ and therefore the number of pseudoparticles is odd. In the second case, conversely, only an even number of pseudoparticles works:

$$\langle -\eta | e^{-H\tau_0} | \eta\rangle = \sum_{n=1,3,\ldots} \sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2} \frac{(\omega\tau_0 d)^n}{n!} = \sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2}\, \mathrm{sh}\,(\omega\tau_0 d), \tag{36}$$

$$\langle \eta | e^{-H\tau_0} | \eta\rangle = \sum_{n=0,2,\ldots} \sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2} \frac{(\omega\tau_0 d)^n}{n!} = \sqrt{\frac{\omega}{\pi}}\, e^{-\omega\tau_0/2}\, \mathrm{ch}\,(\omega\tau_0 d).$$

Going to the limit $\tau_0 \to \infty$, we immediately reproduce formula (20) for the energy of the lowest state. Denoting the ground state of the system by $|0\rangle$, we see that $\langle \eta | 0\rangle = \langle -\eta | 0\rangle = (\omega/4\pi)^{1/4}$, i.e., the symmetry between the right- and left-hand well is indeed not broken.

We now return to the assumption that the characteristic distances between the centers of the instantons are large,

$$|\tau_i - \tau_j| \gg \omega^{-1},$$

and consider how well it works. It is clear that the sums in (36) converge well, and all terms with number $n \gg d\omega\tau_0$ are unimportant. Thus, $n_{\mathrm{char}} \sim d\omega\tau_0$ and $|\tau_i - \tau_j|_{\mathrm{char}} \sim d^{-1}\omega^{-1}$. Having at our disposal the free parameter λ, we can achieve an arbitrary smallness of d, since $d \to 0$ as $e^{-\omega^3/12\lambda}$ in the limit $\lambda \to 0$.

Thus, for $\lambda \ll 1$ we are fully justified in "stringing" instantons and anti-instantons on one another, forming thereby a chain of noninteracting pseudoparticles. Noninteracting in the sense that they are all far from one another, know nothing about the remaining partners, and the total weight function is obtained by multiplying the individual weight functions [d^n in formulas (36)].

Such an approximation is called a dilute instanton gas. In quantum chromodynamics, it has been exploited particularly by Callan, Dashen, and Gross.[8] Unfortunately, in QCD we do not dispose of free parameters like λ that can be kept small. Therefore, a dilute

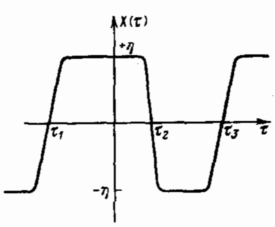


FIG. 5.

instanton gas is not suitable from the quantitative point of view in QCD, and the most we can extract from it are are heuristic indications.

To conclude the section, we note that a somewhat more extensive exposition of the instanton approach to the two-humped potential is contained in Coleman's lecture.[5] The reader interested in special questions, for example, situations not covered by the gas approximation, must consult Ref. 9.

5. EUCLIDEAN FORMULATION OF QCD

Thus, in the simple example of the two-humped potential we have seen that if there exist nontrivial solutions of the classical equations qualitatively new effects occur in the theory. Tunneling from one well to another makes the vacuum wave function quite different from the one obtained in perturbation theory. Our aim in this review is, of course, chromodynamics and not quantum mechanics. However, in chromodynamics too there is a similar phenomenon, which we shall discuss in this and all the following sections of the review.

As we said above, we are concerned with the solution of classical equations in *Euclidean* space. Therefore, we first formulate the Euclidean version of QCD. We give the formulas for the transition from Minkowski to Euclidean space. The spatial coordinates x_1, x_2, x_3 are not changed. For the time coordinate x_0, we make the substitution

$$x_0 = -ix_4. \tag{37}$$

Clearly, when x_0 is continued to imaginary values the zeroth component of the vector potential A_μ also becomes imaginary.

We define the Euclidean vector potential $\hat{A}_\mu$ as follows:

$$A_m = -\hat{A}_m \quad (m=1,2,3), \quad A_0 = i\hat{A}_4 \tag{38}$$

(in this section, we shall use the caret to denote all quantities defined in the Euclidean space). With this definition, the quantities $\hat{A}_\mu$ $(\mu = 1, \ldots, 4)$ form a Euclidean vector. The difference between formulas (38) and the corresponding relations for the vector x_μ [the difference is in the common sign of $\hat{A}_\mu$ $(\mu = 1, \ldots 4)$] is introduced for convenience in the expression of the following formulas.[6)]

Thus, for the operator of covariant differentiation

$$D_\mu = \partial_\mu - igA^a_\mu T^a, \tag{39}$$

where T^a are the matrices of the generators in the representation being considered, we obtain

$$\begin{aligned} &D_m = -\hat{D}_m, \quad D_0 = i\hat{D}_4, \\ &\hat{D}_\mu = \frac{\partial}{\partial \hat{x}_\mu} - ig\hat{A}^a_\mu T^a. \end{aligned} \tag{40}$$

We recall that the operator ∂_μ in Minkowski space has the form $\partial_\mu = (\partial/\partial x_0, -\partial/\partial x_m)$.

For the intensities $G_{\mu\nu}$ we obtain the formulas

$$G^a_{mn} = \hat{G}^a_{mn} \quad (m,\ n=1,\ 2,\ 3), \quad G^a_{0n} = -i\hat{G}^a_{4n}, \tag{41}$$

where the Euclidean intensities $\hat{G}^a_{\mu\nu}$,

$$\hat{G}^a_{\mu\nu} = \frac{\partial}{\partial \hat{x}_\mu}\hat{A}^a_\nu - \frac{\partial}{\partial \hat{x}_\nu}\hat{A}^a_\mu + gf^{abc}\hat{A}^b_\mu \hat{A}^c_\nu \qquad (\mu,\ \nu=1,\ \ldots,\ 4), \tag{42}$$

can be expressed in terms of $\hat{A}_\mu$ and $\partial/\partial\hat{x}_\mu$ in the same way as the Minkowskian $G^a_{\mu\nu}$.

To complete the transition to the Euclidean space, it remains to give the formulas for the Fermi fields. We begin with the definition of four Hermitian γ matrices $\hat{\gamma}_\mu$:

$$\begin{aligned} &\hat{\gamma}_4 = \gamma_0, \quad \hat{\gamma}_m = -i\gamma_m \qquad (m=1,\ 2,\ 3), \\ &\{\hat{\gamma}_\mu,\ \hat{\gamma}_\nu\} = 2\delta_{\mu\nu} \qquad (\mu,\ \nu=1,\ \ldots,\ 4), \end{aligned} \tag{43}$$

where γ_0 and γ_m are the ordinary Dirac matrices.

The fields ψ and $\bar{\psi}$ are regarded as independent anticommuting variables, with respect to which integration is performed in the functional integral. On the transition to the Euclidean space, it is convenient to define the variables $\hat{\psi}$ and $\hat{\bar{\psi}}$ by

$$\psi = \hat{\psi}, \quad \bar{\psi} = -i\hat{\bar{\psi}}. \tag{44}$$

Note that under rotations of the pseudo-Euclidean space, $\bar{\psi}$ transforms as $\psi^+\gamma_0$. In the Euclidean space, $\hat{\bar{\psi}}$ transforms as $\hat{\psi}^+$. Indeed, under infinitesimal rotations of the pseudo-Euclidean space characterized by the parameters $\omega_{\mu\nu}$ $(\mu, \nu = 0, 1, \ldots, 3)$ the spinor ψ acquires the addition

$$\delta\psi = -\frac{1}{4}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)\,\omega_{\mu\nu}\psi.$$

For the change in $\bar{\psi} = \psi^+\gamma_0$ we deduce from this

$$\delta(\psi^+\gamma_0) = -\frac{1}{4}\psi^+\gamma_0\gamma_0(\gamma^+_\nu\gamma^+_\mu - \gamma^+_\mu\gamma^+_\nu)\,\gamma_0\omega_{\mu\nu} = \frac{1}{4}(\psi^+\gamma_0)(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)\,\omega_{\mu\nu},$$

so that $\psi^+_1\gamma_0\psi_2$ is a scalar and $\psi^+_1\gamma_0\gamma_\mu\psi_2$ a vector.

On the transition to the Euclidean space, the parameters ω_{mn} $(m, n = 1, 2, 3)$ do not change, and $\omega_{0n} = i\omega_{4n}$ (because of the substitution $x_0 = -ix_4$). For the variations of $\hat{\psi}$ and $\hat{\psi}^+$ under rotations, we obtain

$$\delta\hat{\psi} = \frac{1}{4}(\hat{\gamma}_\mu\hat{\gamma}_\nu - \hat{\gamma}_\nu\hat{\gamma}_\mu)\,\hat{\omega}_{\mu\nu}\hat{\psi}, \quad \delta\hat{\psi}^+ = -\frac{1}{4}\psi^+(\hat{\gamma}_\mu\hat{\gamma}_\nu - \hat{\gamma}_\nu\hat{\gamma}_\mu)\,\hat{\omega}_{\mu\nu},$$

so that $\psi^+_1\psi_2$ and $\psi^+_1\gamma_\mu\psi_2$ are a scalar and vector, respectively.

Finally, we can write down an expression for the Euclidean action:

$$\begin{aligned} &iS = -\hat{S}, \\ &S = \int d^4x\left[-\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} + \bar{\psi}\,(i\gamma_\mu D_\mu - M)\,\psi\right], \\ &\hat{S} = \int d^4\hat{x}\left[\frac{1}{4}\hat{G}^a_{\mu\nu}\hat{G}^a_{\mu\nu} + \hat{\bar{\psi}}\,(-i\hat{\gamma}_\mu\hat{D}_\mu - iM)\,\hat{\psi}\right], \end{aligned} \tag{45}$$

where it is assumed that $\hat{\psi}$ is a column in the space of flavors (with color index), and M is a matrix in this space.

Below, we shall use the Euclidean space and omit the caret. The formulas given below make it possible to relate the quantities in the pseudo-Euclidean and Euclidean spaces.

To conclude the section, we note that if we are considering quantities such as the vacuum expectation val-

6)If we use the definition $\hat{A}_m = A_m$ $(m = 1, 2, 3)$, then in all the following connection formulas it is necessary to make the substitution $g \to -g$.

ues of the time-ordered products of currents for spacelike external momenta, i.e., when the sources do not produce real hadrons from the vacuum, the Euclidean formulation is not only merely possible but in fact is more adequate than the pseudo-Euclidean. The region of timelike momenta, where there are singularities, can be reached by means of analytic continuation. Such an approach is particularly necessary for quantum chromodynamics, for which the fundamental objects of the theory—the quarks and gluons—have meaning only in the Euclidean domain, and the real singularities corresponding to hadrons have to be obtained.

6. BPST INSTANTONS. GENERAL PROPERTIES

a) Finiteness of the action and the topological charge

It was already clear in the quantum-mechanical example discussed above what an important part is played by solutions that give a minimum of the Euclidean action in the limit $\tau_0\to\infty$. In general, the action increases unboundedly in the limit $\tau_0\to\infty$, and the condition that it be finite imposes strong restrictions on the paths.

Thus, in the one-dimensional example we have analyzed, the finite-action condition means that the function $x(\tau)$ as $\tau\to\pm\infty$ must have the limits $\pm\eta$. In this way there arises naturally a topological classification of functions giving a finite action on the basis of their limiting values. Formally, a topological charge can be introduced as follows:

$$Q=\frac{1}{2\eta}\int_{-\infty}^{+\infty}\mathrm{d}t\,\dot{x}(t)=\frac{x(+\infty)-x(-\infty)}{2\eta}.$$

It is obvious that Q can take on the values $0, +1, -1$. Functions with different Q cannot be carried into one another by a continuous deformation that leaves the action finite. Therefore, in each of the classes $Q=0, +1, -1$ there exists a corresponding minimum of the action and corresponding functions that realize it. The instanton and anti-instanton realize minima for $Q=\pm1$.

We now turn to "gluodynamics"—the theory of a non-Abelian vector field—and consider first the case of the group SU(2). We pose the same question: What must be the behavior of the vector fields A_μ^a as $x\to\infty$ if the action is to be finite? (We have in mind the Euclidean action $\hat{S}$; see (45).] It is clear that the intensities $G_{\mu\nu}^a$ must decrease more rapidly than $1/x^2$. But this by no means implies that the fields A_μ^a must decrease faster than $1/x$. Indeed, suppose A_μ^a in the limit $x\to\infty$ has the form

$$A_\mu=\frac{g\tau^a}{2}A_\mu^a\xrightarrow[x\to\infty]{}iS\,\partial_\mu S^+, \tag{46}$$

where we have introduced matrix notation: S is a unitary unimodular matrix that depends on the angles in the Euclidean space. Although the angular components of A_μ are proportional to $1/x$, it is clear that in the region in which the expression (46) holds the intensities $G_{\mu\nu}^a$ vanish, since A_μ^a has a purely gauge form.

Thus, the behavior of A_μ^a at large x is determined by the matrix S, which depends on the angles. Under a gauge transformation of A_μ defined by the matrix $U(x)$:

$$A_\mu\to U^+A_\mu U+iU^+\partial_\mu U,$$

the matrix S is replaced by $U^+(x\to\infty)S$. It would appear that one can always choose $U(x)$ such that $U(x\to\infty)=S$ and thus remove the terms $1/x$ from A_μ. However, this argument is correct only if the matrix $U(x)$ does not have singularities at any value of x. Otherwise, the problem of the behavior of $A_\mu(x)$ is transferred from the point at infinity to the position of the singularity of $U(x)$.

As a result, the problem of classifying the fields A_μ^a which give finite action reduces to the topological classification of the matrices S. We shall not present this classification, which was obtained in the pioneering paper of Ref. 1, but rather give examples of nontrivial (not reducible to the unit matrix) matrices S. For example, we have the matrix

$$S_1=\frac{x_4+i\mathbf{x}\boldsymbol{\tau}}{\sqrt{x^2}}. \tag{47}$$

It corresponds to unit topological charge (there is a one-to-one correspondence between the space of unitary unimodular matrices and the points of the hypersphere in Euclidean space). To topological charge n there corresponds a matrix of the form

$$S_n=(S_1)^n,\qquad n=0,\ \pm1,\ \pm2,\ \ldots\ . \tag{48}$$

Of course, one could choose a different form of the matrix S corresponding to the charge n, but the difference between it and S_n reduces to a gauge transformation.

For n, there exists the gauge-invariant integral representation

$$n=\frac{g^2}{32\pi^2}\int\mathrm{d}^4x\,G_{\mu\nu}^a\tilde{G}_{\mu\nu}^a, \tag{49}$$

where

$$\tilde{G}_{\mu\nu}^a=\frac{1}{2}\varepsilon_{\mu\nu\gamma\delta}G_{\gamma\delta}^a,\quad \varepsilon_{1234}=1. \tag{50}$$

The validity of Eq. (49) can be verified by using the fact that $G_{\mu}^a\tilde{G}_{\mu\nu}^a$ can be represented in the form of a total derivative,

$$G_{\mu\nu}\tilde{G}_{\mu\nu}=\partial_\mu K_\mu,$$
$$K_\mu=2\varepsilon_{\mu\nu\gamma\delta}\left(A_\nu^a\partial_\gamma A_\delta^a+\frac{1}{3}g\varepsilon^{abc}A_\nu^aA_\gamma^bA_\delta^c\right),$$

so that the volume integral (9) can be transformed into an integral over a distant surface, where A_μ^a has the form (46).

b) The distinguished role of the group SU(2)

Hitherto, we have discussed the group SU(2). For groups different from SU(2), the construction of instanton solutions with $n=1$ reduces to the case of SU(2) by means of separation of SU(2) subgroups. Why is the group SU(2) distinguished? We shall attempt to explain this without using topological terminology.

The possibility of deformation of the matrices S is determined by the gauge invariance discussed above. We attempt to fix the gauge, for which we represent an arbitrary field A_μ in the form

$$A_\mu(x)=S(x)\,\tilde{A}_\mu(x)\,S^+(x)+iS(x)\,\partial_\mu S^+(x), \tag{51}$$

where the field $\tilde{A}_\mu$ satisfies definite gauge conditions [for example, $\tilde{A}_0=0$ or $\partial_m\tilde{A}_m=0$ $(m=1,2,3)$]. This fixing does not completely determine the transition to the new fields $\tilde{A}_\mu(x)$ and $S(x)$, since A_μ is invariant under global transformations of the form

$$S(x)\to S(x)U_2^+,\quad \tilde{A}_\mu(x)\to U_2\tilde{A}_\mu U_2^+ \tag{52}$$

with matrix U_2 that does not depend on x.

In addition, even after the fixing of the gauge the theory is still invariant with respect to global isotopic rotations for A_μ, which in terms of the new fields $\tilde{A}_\mu(x)$ and $S(x)$ is equivalent to the transformations

$$S(x)\to U_1S,\quad \tilde{A}_\mu(x)\to\tilde{A}_\mu(x). \tag{53}$$

Thus, the isotopic SU(2) invariance of the theory together with the gauge invariance reduce to the set of *global* transformations (52) and (53), which obviously form the group SU(2)×SU(2). The field $S(x)$ transforms in accordance with the representation $(1/2,1/2)$, and $\tilde{A}_\mu(x)$ in accordance with the representation $(1,0)$.

On the other hand, the group of rotations of four-dimensional Euclidean space is again, as is well known, SU(2)×SU(2), and the generators of the SU(2) subgroups have the form

$$I_1^a=\frac{1}{4}\eta_{a\mu\nu}M_{\mu\nu},\quad I_2^a=\frac{1}{4}\bar{\eta}_{a\mu\nu}M_{\mu\nu}\quad \begin{pmatrix}a=1,\ 2,\ 3\\ \mu,\ \nu=1,\ \ldots,\ 4\end{pmatrix}, \tag{54}$$

where $M_{\mu\nu}=-ix_\mu\partial/\partial x_\nu+ix_\nu\partial/\partial x_\mu$ + spin part are the operators of infinitesimal rotations in the (μ,ν) plane, and $\eta_{a\mu\nu}$ are the numerical symbols

$$\eta_{a\mu\nu}=\begin{cases}\varepsilon_{a\mu\nu}, & \mu,\ \nu=1,\ 2,\ 3,\\ -\delta_{a\nu}, & \mu=4,\\ \delta_{a\mu}, & \nu=4,\\ 0, & \mu=\nu=4.\end{cases} \tag{55}$$

(The symbols $\bar{\eta}_{a\mu\nu}$ differ from η by a change in the sign of δ.) The coordinate vector x_μ transforms in accordance with the representation $(1/2,1/2)$. This is conveniently seen by considering transformations of the matrix

$$x_4+i\mathbf{x}\boldsymbol{\tau}=i\tau_\mu^+x_\mu, \tag{56}$$

where we have introduced the notation

$$\tau_\mu^\pm=(\boldsymbol{\tau},\ \mp i). \tag{57}$$

For $\tau_\mu^\pm$, we have

$$\tau_\mu^+\tau_\nu^-=\delta_{\mu\nu}+i\eta_{a\mu\nu}\tau^a,\quad \tau_\mu^-\tau_\nu^+=\delta_{\mu\nu}+i\bar{\eta}_{a\mu\nu}\tau^a. \tag{57'}$$

It is not difficult to find the law of transformation of the matrix (56),

$$e^{i\varphi_1^aI_1^a+i\varphi_2^aI_2^a}i\tau_\mu^+x_\mu=e^{-i\varphi_1^a(\tau/2)^a}(i\tau_\mu^+x_\mu)e^{i\varphi_2^a(\tau^a/2)},$$

where φ_1^a and φ_2^a are parameters of the rotations, i.e., there is multiplication by unitary unimodular matrices from the left and the right.

The choice of S in the form $S_1=ix_\mu\tau_\mu^+/\sqrt{x^2}$ distinguishes certain directions in the isotopic and coordinate spaces. However, under rotation through the same angles in the spatial SU(2)×SU(2) group and in the SU(2)×SU(2) group given by the transformations (52) and (53), the matrix S_1 obviously does not change. In other words, if instead of I_1^a and I_2^a we call $I_1^a+T_1^a$ and $I_2^a+T_2^a$, where $T_{1,2}^a$ are the operators of the infinitesimal transformations (52) and (53), the angular momentum operators, the introduced object has spin zero.

Thus, we see that the group SU(2) is distinguished on account of the dimension of the coordinate space.

c) Value of the action for instanton solutions

Although we do not yet have the explicit form of the instanton solution, we can nevertheless calculate the value of the action for it. Indeed, for positive values of the topological charge n, the Euclidean action can be rewritten in the form

$$S=\int d^4x\,\frac{1}{4}G_{\mu\nu}^aG_{\mu\nu}^a=\int d^4x\left[\frac{1}{4}G_{\mu\nu}^a\tilde{G}_{\mu\nu}^a+\frac{1}{8}(G_{\mu\nu}^a-\tilde{G}_{\mu\nu}^a)^2\right]=n\frac{8\pi^2}{g^2}+\frac{1}{8}\int d^4x(G_{\mu\nu}^a-\tilde{G}_{\mu\nu}^a)^2. \tag{58}$$

It is clear from this formula that in the class of functions with given positive n the minimum of S is attained for $G_{\mu\nu}^a=\tilde{G}_{\mu\nu}^a$ and is equal to $(8\pi^2/g^2)$. We recall that specification of n does not signify that we seek a conditional extremum, since functions with different n cannot be related by a continuous deformation if the action is to remain finite.

The case of negative n is obtained from (58) by the reflection $x_{1,2,3}\to-x_{1,2,3}$, under which $G_{\mu\nu}\tilde{G}_{\mu\nu}\to-G_{\mu\nu}\tilde{G}_{\mu\nu}$ and accordingly $n\to-n$. Thus, the minimum of the action for negative n is $(8\pi^2/g^2)|n|$, and it is attained when $G_{\mu\nu}^a=-\tilde{G}_{\mu\nu}^a$.

As can be seen from this discussion, fulfillment of the self-duality and antiself-duality conditions $G_{\mu\nu}^a=\pm\tilde{G}_{\mu\nu}^a$ automatically leads to satisfaction of the equations of motion $D_\mu G_{\mu\nu}=0$. This can also be seen directly; indeed, for a self-dual field, say, we have

$$D_\mu G_{\mu\nu}^a=D_\mu\tilde{G}_{\mu\nu}^a=\frac{1}{2}\varepsilon_{\mu\nu\gamma\delta}D_\mu G_{\gamma\delta}^a=\frac{1}{6}\varepsilon_{\mu\nu\gamma\delta}(D_\mu G_{\gamma\delta}^a+D_\gamma G_{\delta\mu}^a+D_\delta G_{\mu\gamma}^a)=0,$$

where we have used the Bianchi identity:

$$D_\mu G_{\gamma\delta}+D_\delta G_{\mu\gamma}+D_\gamma G_{\delta\mu}=0.$$

7. EXPLICIT FORM OF THE BPST INSTANTON

a) Solution with $n=1$

As discussed in the previous section, the asymptotic behavior of A_μ^a for this solution is

$$g\frac{\tau^a}{2}A_\mu^a\xrightarrow[x\to\infty]{}iS_1\partial_\mu S_1^+,\quad S_1=\frac{i\tau_\mu^+x_\mu}{\sqrt{x^2}}, \tag{59}$$

where the matrices $\tau_\mu^\pm$ are defined in (57). We shall also use the symbols $\eta_{a\mu\nu}$ and $\bar{\eta}_{a\mu\nu}$ defined by Eqs. (55). These numerical coefficients are frequently called the 't Hooft symbols, and some useful relations for $\eta_{a\mu\nu}$ are given in subsection c) of this section.

The expression for the asymptotic behavior of A_μ^a can be rewritten in terms of the 't Hooft symbols as follows:

$$A_\mu^a\xrightarrow[x\to\infty]{}\frac{2}{g}\eta_{a\mu\nu}\frac{x_\nu}{x^2}.$$

For an instanton with center at the point $x=0$, it is natural to assume the same angular dependence of the field for all x, i.e., to seek the solution in the form

$$A_\mu^a(x)=\frac{2}{g}\eta_{a\mu\nu}x_\nu\frac{f(x^2)}{x^2}, \tag{60}$$

where $f(x^2)\xrightarrow[x^2\to\infty]{}1$, $f(x^2)\xrightarrow[x^2\to 0]{}\mathrm{const}\cdot x^2$. The last condition corresponds to the absence of a singularity at the origin. A justification for the assumption (60) will be the construction of a self-dual expression for $G_{\mu\nu}^a$. From (60), we obtain for $G_{\mu\nu}^a$

$$G_{\mu\nu}^a=-\frac{4}{g}\left\{\eta_{a\mu\nu}\frac{f(1-f)}{x^2}+\frac{x_\mu\eta_{a\nu\gamma}x_\gamma-x_\nu\eta_{a\mu\gamma}x_\gamma}{x^4}[f(1-f)-x^2f']\right\}. \tag{61}$$

In deriving (61), we have used the relation for $\varepsilon^{abc}\times\eta_{b\mu\gamma}\eta_{c\nu\delta}$ from the list of formulas in subsection c) at the end of this section. Using the formula for $\varepsilon_{\mu\nu\gamma\delta}\eta_{a\delta\rho}$ from the same list, we obtain for $\tilde{G}_{\mu\nu}^a$ the expression

$$\tilde{G}_{\mu\nu}^a=-\frac{4}{g}\left\{\eta_{a\mu\nu}f'-\frac{1}{x^4}(x_\mu\eta_{a\nu\gamma}x_\gamma-x_\nu\eta_{a\mu\gamma}x_\gamma)[f(1-f)-x^2f']\right\}.$$

The condition of self-duality, $G_{\mu\nu}^a=\tilde{G}_{\mu\nu}^a$, requires fulfillment of the equation $f(1-f)-x^2f'=0$, which determines the function f:

$$f(x^2)=\frac{x^2}{x^2+\rho^2}, \tag{62}$$

where ρ^2 is a constant of integration; ρ is called the scale of the instanton. The translational invariance guarantees the obtaining of a solution with center at an arbitrary point x_0, for which it is necessary to replace x by $x-x_0$.

Thus, the final expression for the instanton with center at the point x_0 and scale ρ has the form

$$\begin{aligned}A_\mu^a&=\frac{2}{g}\eta_{a\mu\nu}\frac{(x-x_0)_\nu}{(x-x_0)^2+\rho^2},\\ G_{\mu\nu}^a&=-\frac{4}{g}\eta_{a\mu\nu}\frac{\rho^2}{[(x-x_0)^2+\rho^2]^2}.\end{aligned} \tag{63}$$

It can now be verified that the action for the instanton is $8\pi^2/g^2$, as was shown in general form. The anti-instanton is obtained by the substitution $\eta_{a\mu\nu}\to\bar{\eta}_{a\mu\nu}$.

b) Singular gauge. The 't Hooft ansatz

It is frequently convenient to use the expression for A_μ^a in the so-called singular gauge, when the "bad" behavior of A_μ^a is transferred from the point at infinity to the center of the instanton. As was discussed in the previous section, such a transfer can be realized by a gauge transformation with a matrix $U(x)$ which becomes identical with $S(x)$ as $x\to\infty$.[7)] We write down the formulas of the gauge transformation,

$$\begin{aligned}g\frac{\tau^a}{2}\bar{A}_\mu^a&=U^+g\frac{\tau^a}{2}A_\mu^aU+iU^+\partial_\mu U,\\ g\frac{\tau^a}{2}\overline{G_{\mu\nu}^a}&=U^+g\frac{\tau^a}{2}G_{\mu\nu}^aU,\end{aligned} \tag{64}$$

and for an instanton with center at x_0 take a matrix of the form

$$U=\frac{i\tau_\mu^+(x-x_0)_\mu}{\sqrt{(x-x_0)^2}}. \tag{64'}$$

Then for the potential $\bar{A}_\mu^a$ and the intensities $\bar{G}_{\mu\nu}^a$ in the singular gauge we obtain

$$\begin{aligned}\bar{A}_\mu^a&=\frac{2}{g}\bar{\eta}_{a\mu\nu}(x-x_0)_\nu\frac{\rho^2}{(x-x_0)^2[(x-x_0)^2+\rho^2]},\\ \bar{G}_{\mu\nu}^a&=-\frac{8}{g}\left[\frac{(x-x_0)_\mu(x-x_0)_\rho}{(x-x_0)^2}-\frac{1}{4}\delta_{\mu\rho}\right]\bar{\eta}_{a\nu\rho}\frac{\rho^2}{[(x-x_0)^2+\rho^2]^2}-(\mu\leftrightarrow\nu).\end{aligned} \tag{65}$$

It is obvious that the quantities $G_{\mu\nu}^aG_{\gamma\delta}^a$ are invariants of the gauge transformation (see, however, the last footnote). Note also the circumstance that (65) contains the symbols $\bar{\eta}_{a\mu\nu}$ but not $\eta_{a\mu\nu}$. This difference is due to the fact that in the singular gauge the topological charge (49) is accumulated in the neighborhood of $x=x_0$ and not at infinity.

The expression (65) for $\bar{A}_\mu^a$ can be rewritten in the form

$$\bar{A}_\mu^a=-\frac{1}{g}\bar{\eta}_{a\mu\nu}\partial_\nu\ln\left[1+\frac{\rho^2}{(x-x_0)^2}\right]. \tag{66}$$

As was noted by 't Hooft, this expression can be generalized to a topological charge n greater than unity. Indeed, if

$$A_\mu^a=-\frac{1}{g}\bar{\eta}_{a\mu\nu}\partial_\nu\ln W(x), \tag{67}$$

then for $G_{\mu\nu}^a-\tilde{G}_{\mu\nu}^a$ we obtain [see the properties of the η symbols in subsection c)]

$$G_{\mu\nu}^a-\tilde{G}_{\mu\nu}^a=\frac{1}{g}\eta_{a\mu\nu}\frac{\partial_\gamma\partial_\gamma W}{W}.$$

The self-duality of $G_{\mu\nu}^a$ requires fulfillment of the equation $\partial_\gamma\partial_\gamma W/W=0$. The solution with topological charge n has the form

$$W=1+\sum_{i=1}^{n}\frac{\rho_i^2}{(x-x_i)^2}, \tag{68}$$

i.e., it describes instantons with centers at the points x_i. The effective scale of an instanton with center at the point x_i is obviously

$$\rho_i^{\mathrm{eff}}=\rho_i\left[1+\sum_{k\neq i}\frac{\rho_k^2}{(x_k-x_i)^2}\right]^{-1/2}.$$

It should be noted that the choice of A_μ^a in the form (67) did not give the most general solution with charge n, since all n instantons have the same orientation in the isotopic space (for the construction of the general solution, see Ref. 10).

c) Relations for the η symbols

We give a list of relations for the symbols $\eta_{a\mu\nu}$ and $\bar{\eta}_{a\mu\nu}$ defined by Eqs. (55):

$$\eta_{a\mu\nu}=\frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}\eta_{a\alpha\beta},$$

$$\eta_{a\mu\nu}=-\eta_{a\nu\mu},\qquad \eta_{a\mu\nu}\eta_{b\mu\nu}=4\delta_{ab},$$

$$\eta_{a\mu\nu}\eta_{a\mu\lambda}=3\delta_{\nu\lambda},\qquad \eta_{a\mu\nu}\eta_{a\mu\nu}=12,$$

$$\eta_{a\mu\nu}\eta_{a\gamma\lambda}=\delta_{\mu\gamma}\delta_{\nu\lambda}-\delta_{\mu\lambda}\delta_{\nu\gamma}+\varepsilon_{\mu\nu\gamma\lambda},$$

$$\varepsilon_{\mu\nu\lambda\sigma}\eta_{a\gamma\sigma}=\delta_{\gamma\mu}\eta_{a\nu\lambda}-\delta_{\gamma\nu}\eta_{a\mu\lambda}+\delta_{\gamma\lambda}\eta_{a\mu\nu},$$

$$\eta_{a\mu\nu}\eta_{b\mu\lambda}=\delta_{ab}\delta_{\nu\lambda}+\varepsilon_{abc}\eta_{c\nu\lambda},$$

$$\varepsilon_{abc}\eta_{b\mu\nu}\eta_{c\gamma\lambda}=\delta_{\mu\gamma}\eta_{a\nu\lambda}-\delta_{\mu\lambda}\eta_{a\nu\gamma}-\delta_{\nu\gamma}\eta_{a\mu\lambda}+\delta_{\nu\lambda}\eta_{a\mu\gamma},$$

$$\eta_{a\mu\nu}\bar{\eta}_{b\mu\nu}=0,\quad \eta_{a\gamma\mu}\bar{\eta}_{b\gamma\lambda}=\eta_{a\gamma\lambda}\bar{\eta}_{b\gamma\mu}.$$

[7)]More precisely, this transformation should be called a quasigauge transformation, since at the point where $U(x)$ has a singularity (and there must be such a singularity) this transformation changes the gauge-invariant quantities, for example, $G_{\mu\nu}^aG_{\mu\nu}^a$. To use such transformations, it is necessary to consider a space with the neighborhoods of the singular points deleted. This we shall do, remembering that the physical quantities are nonsingular at the singular points.

To go over from the relations for $\eta_{a\mu\nu}$ to those for $\bar{\eta}_{a\mu\nu}$ it is necessary to make the substitution

$$\eta_{a\mu\nu} \to \bar{\eta}_{a\mu\nu}, \quad \varepsilon_{\mu\nu\gamma\delta} \to -\varepsilon_{\mu\nu\gamma\delta}.$$

8. CALCULATION OF THE PRE-EXPONENTIAL FACTOR FOR THE BPST INSTANTON

a) Expansion near a saddle point. Choice of the gauge and regularization

As in the quantum-mechanical example, to calculate the pre-exponential factor in the instanton contribution to the vacuum–vacuum transition, it is necessary to represent the field A_μ^a in the form

$$A_\mu^a = A_\mu^{a(\mathrm{ins})} + a_\mu^a \tag{69}$$

and expand the action $S(A)$ with respect to the deviation a_μ^a from the instanton field $A_\mu^{a(\mathrm{ins})}$:

$$\begin{aligned} S(A) &= S_0 + \frac{1}{2}\int d^4x a_\mu^a L_{\mu\nu}^{ab}(A^{\mathrm{ins}}) a_\nu^b \\ &= \frac{8\pi^2}{g^2} + \frac{1}{2}\int d^4x a_\mu^a \{D^2 a_\mu^a - D_\mu D_\nu a_\nu^a - 2g\varepsilon^{abc} G_{\mu\nu}^b a_\nu^c\}, \end{aligned} \tag{70}$$

where the instanton field is substituted in D_μ and $G_{\mu\nu}$. As in the one-dimensional case, the integration with respect to the deviations a_μ reduces to calculation of the determinant of the operator $L_{\mu\nu}^{ab}$. There are however two important differences from the one-dimensional case:

The operator L is degenerate due to the gauge invariance. Indeed, fields a_μ^a of the form $a_\mu^a = (D_\mu \lambda)^a$ with arbitrary function $\lambda^a(x)$ make the quadratic form (70) vanish. In order to have the possibility of working with a degenerate form of this kind, it is necessary to fix the gauge. This can be done conveniently by adding to the action the term

$$\Delta S = \frac{1}{2}\int d^4x (D_\mu a_\mu^a)^2 = \frac{1}{2}\int d^4x a_\mu^a (\Delta L)_{\mu\nu}^{ab} a_\nu^b, \tag{71}$$

which lifts the degeneracy. To avoid changing the content of the theory, we must, as is well known, simultaneously add Faddeev–Popov ghosts:

$$\Delta S_{\mathrm{gh}} = \int d^4x \bar{\Phi}^a D^2 \Phi^a = \int d^4x \bar{\Phi}^a L_{\mathrm{gh}}^{ab} \Phi^b, \tag{72}$$

where Φ^a is a complex anticommuting field. As a result, the instanton contribution can be written in the form

$$\langle 0 \mid 0_T \rangle_{\mathrm{ins}} = [\det (L+\Delta L)]^{-1/2} (\det L_{\mathrm{gh}})\, e^{-S_0}, \tag{73}$$

where $|0_T\rangle$ is the vacuum after time T, $|0_T\rangle = e^{-HT}0\rangle$, H is the Hamiltonian, $S_0 = 8\pi^2/g^2$, $(L + \Delta L)_{\mu\nu}^{ab}$ is the operator in the quadratic form of the fields a_μ^a, and L_{gh} acts on the ghost fields. The determinant of L_{gh} occurs in a positive power, since $\Phi^a, \bar{\Phi}^a$ are anticommuting fields.

A second difference from the one-dimensional case is the presence in the theory of ultraviolet divergences. By virtue of the renormalizability, all the divergences must be eliminated by a renormalization of the coupling constant, but it is first necessary to regularize the expressions under consideration. The regularization can be done as follows. Instead of the determinant of the operator $L + \Delta L$ we consider the ratio $\det(L + \Delta L)/\det(L + \Delta L + M^2)$, where the introduction of the cutoff parameter M can be interpreted as the addition to the theory of a Pauli–Villars vector field with mass M. The determinant of L_{gh} is regularized similarly. Thus, it is necessary to calculate

$$\langle 0 \mid 0_T \rangle_{\mathrm{ins}}^{\mathrm{Reg}} = \left[\frac{\det (L+\Delta L)}{\det (L+\Delta+M^2)}\right]^{-1/2} \frac{\det L_{\mathrm{gh}}}{\det (L_{\mathrm{gh}}+M^2)} e^{-S_0}, \tag{74}$$

or, more precisely, the ratio of $\langle 0|0_T\rangle_{\mathrm{ins}}^{\mathrm{Reg}}$ to the corresponding perturbation-theoretical quantity $\langle 0|0_T\rangle_{\mathrm{p.th}}$, which differs in having $A_\mu^a \equiv 0$ substituted instead of the instanton field. For $A_\mu^a = 0$, it is obvious that $S_0 = 0$, while for the instanton $S_0 = 8\Pi^2/g_0^2$, where the subscript in the coupling constant g_0 emphasizes that this is the unrenormalized coupling constant normalized by the cutoff parameter M, $g_0 = g(M)$.

We shall not go into a detailed exposition of 't Hooft's calculations for $\langle 0|0_T\rangle/\langle 0|0_T\rangle_{\mathrm{p.th}}$ but obtain the result up to a numerical factor. Study of the zero-frequency modes plays the main part in obtaining the result.

b) Zero-frequency modes

As was shown in the one-dimensional example, each zero-frequency modes leads in $[\det(L + \Delta L)]^{-1/2}$ to a factor proportional to $\sqrt{S_0}$ and an integral with respect to a corresponding collective coordinate. What are the collective coordinates in the case of the BPST instanton in the group SU(2)?

First, there are the four coordinates of the center x_0, then the scale ρ, and, finally, the three Eulerian angles θ, φ, ψ, which specify the orientation of the instanton in the isospace. The spatial rotations need not be counted, since they are equivalent to isorotations (see Sec. 6b).

As a result of the regularization, $[\det(L + \Delta L)]^{-1/2}$ is multiplied by $[\det(L + \Delta L + M^2)]^{1/2}$, i.e., each zero-frequency mode gives rise to a factor M. Thus, from all (since we have listed *all* collective coordinates) zero-frequency modes there arises in $\langle 0|0_T\rangle_{\mathrm{ins}}^{\mathrm{Reg}}$ the factor

$$\int d^4x_0\, d\rho \sin\theta\, d\theta\, d\varphi\, d\psi\, M^8 (\sqrt{S_0})^8 \rho^3. \tag{75}$$

The factor ρ^3 arises from the Jacobian of the transition to integration over θ, φ, ψ and is recovered on the basis of dimensional considerations.

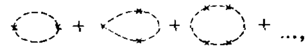
Using (75) we rewrite $\langle 0|0_T\rangle_{\mathrm{ins}}^{\mathrm{Reg}}/\langle 0|0_T\rangle_{\mathrm{p.th}}$ in the form

$$\frac{\langle 0 \mid 0_T\rangle_{\mathrm{ins}}^{\mathrm{Reg}}}{\langle 0 \mid 0_T\rangle_{\mathrm{p.th}}} = \mathrm{const} \int \frac{d^4x\, d\rho}{\rho^5} \left(\frac{8\pi^2}{g_0^2}\right)^4 \exp\left(-\frac{8\pi^2}{g^2} + 8\ln M\rho + \Phi_1\right), \tag{76}$$

where $\exp\Phi_1$ denotes the contribution of the positive-frequency modes.

c) Positive-frequency modes. Effective charge

The quantity Φ_1 depends on the dimensionless parameter $M\rho$ and in the limit $M\rho \gg 1$ can be readily found by means of ordinary perturbation theory. Indeed, calculation of the pre-exponential factory by retaining the terms quadratic in the deviation from the external field corresponds to calculation of the single-loop corrections in perturbation theory. We are here referring to diagrams of the form

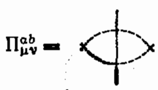
 (77)

where the cross denotes vertices of the interaction with the external field, and the broken lines correspond to the propagators of the fields a_μ^a (plus similar loops with the ghosts $\Phi^a, \bar{\Phi}^a$); the external field has the form $A_\mu^{a\,(\mathrm{ins})}$.

It is clear that complete calculation of the contribution of the zero-frequency modes requires summation of a complete chain of diagrams—the zero-frequency modes do not appear in any finite order. A manifestation of this nonanalyticity is the presence of the term $\ln(8\pi^2/g_0^2)$ in $\ln\langle 0|0_T\rangle_{\mathrm{ins}}$. It is also clear that there is no nonanalyticity of this kind for the positive-frequency modes.

In the limit in which we are interested, $M\rho \gg 1$, only the first of the diagrams (77) is important in the calculation, since all the following diagrams are convergent and do not give a dependence on the cutoff parameter M [they change the constant in (76)]. Moreover, in the second order in the external field it can be seen that the contribution of the positive-frequency modes is given by an unsubtracted dispersion relation for the polarization operator $\Pi_{\mu\nu}^{ab}$.

The imaginary part of $\Pi_{\mu\nu}^{ab}$ is obtained by cutting the first diagram (77) and is well defined. In its calculation, it is necessary to take into account only quanta with three-dimensionally transverse polarization states; the unphysical polarizations and ghosts are not necessary. Omitting the details of this simple calculation, we give the result for $\mathrm{Im}\,\Pi_{\mu\nu}$:

$\mathrm{Im}\,\Pi_{\mu\nu}^{ab} =$  $= \delta^{ab}(g_{\mu\nu}k^2 - k_\mu k_\nu)\frac{g^2}{16\pi}\cdot\frac{2}{3}.$

Writing down the unsubtracted dispersion representation for $\Pi_{\mu\nu}^{(1)}$ (the part of the polarization operator associated with the positive-frequency modes), we obtain

$$\Pi_{\mu\nu}^{ab(1)} = \delta^{ab}(g_{\mu\nu}k^2 - k_\mu k_\nu)\frac{1}{\pi}\int\frac{ds}{s-k^2}\cdot\frac{2}{3}\frac{g^2}{16\pi} = \delta^{ab}(g_{\mu\nu}k^2 - k_\mu k_\nu)\frac{2}{3}\frac{g^2}{16\pi^2}\ln\frac{M^2}{-k^2}, \tag{78}$$

where we have terminated the integration over s at M^2, since the regularization involves a subtraction of an analogous contribution with Pauli–Villars particles of mass M.

The result (78) for the contribution of the positive-frequency modes means that the action for the external field acquires from these quantum corrections the effective addition

$$\Delta S^{\mathrm{Mink}} = \frac{2}{3}\frac{g^2}{16\pi^2}\ln M^2\rho^2\int d^4x\left[-\frac{1}{4}(G_{\mu\nu}^a)^2\right], \tag{79}$$

where we use the notation of pseudo-Euclidean space and have replaced $1/(-k^2)$ by the square ρ^2 of the characteristic scale of the field (strictly speaking, we ought to write a differential operator, but for the calculation of the coefficient of $\ln M\rho$ this is not important). Going over to the Euclidean action and substituting the instanton $G_{\mu\nu}^a$, we obtain the result for Φ_1:

$$\Phi_1 = \frac{2}{3}\ln M\rho. \tag{80}$$

Thus, allowance for the zero-frequency and positive-frequency modes has the consequence that $8\pi^2/g_0^2$ in the argument of the exponential (76) is replaced by the effective charge $8\pi^2/g^2(\rho)$:

$$\frac{8\pi^2}{g^2(\rho)} = \frac{8\pi^2}{g_0^2} - 8\ln M\rho + \frac{22}{3}\ln M\rho = \frac{8\pi^2}{g_0^2} - \frac{22}{3}\ln M\rho. \tag{81}$$

Of course, this result is a direct consequence of the renormalizability, and we have wasted time on its derivation only to emphasize the very beautiful explanation of the antiscreening of the charge in a non-Abelian theory which arises when the zero-frequency modes are considered.

Indeed, both the sign and the magnitude of the coefficient of the "antiscreening" logarithm (76) are obvious consequences of the above—the coefficient is simply the number of zero-frequency modes.

In the framework of the perturbation-theoretical calculations, the "antiscreening" result can be most clearly explained in the framework of the ghostless Coulomb gauge, which was used in calculations by Khriplovich[11] as early as 1969. Besides the "dispersion" part, the calculation of which we have discussed above, the polarization operator in this gauge contains a contribution that does not have an imaginary part and arises when one of the virtual quanta has a three-dimensionally transverse polarization and the second is a Coulomb quantum. The opposite signs of the "nondispersion" and "dispersion" parts of $\Pi_{\mu\nu}$ correspond to the opposite signs of interactions due to the exchange of a Coulomb quantum and a transverse quantum (electric forces repel charges of the same sign, while magnetic forces attract currents of the same type).

The calculation of the "nondispersion" part in the Coulomb gauge requires care, since it is necessary to use the noncovariant Hamiltonian formalism, and the coefficient of the logarithm is not, of course, known *a priori*. As we have seen, none of these problems arise in the determination of the contribution of the zero-frequency modes. With this we conclude our panegyric to the zero-frequency modes.

d) Two-loop approximation

The above calculations led to replacement of the unrenormalized coupling constant g_0 in the classical action by the effective constant $g(\rho)$. However, the unrenormalized constant still remains in the factor $(8\pi^2/g_0^2)^4$ [see (76)], though it is clear that, because of the renormalizability, it should not occur in the result. The reason for this is that the accuracy obtained by using the single-loop approximation is inadequate to distinguish the factor $(8\pi^2/g_0^2)^4$ from $[8\pi^2/g^2(\rho)]^4$, and we require a two-loop calculation.

We show that from the two-loop calculation we actually require only the expression for the effective charge; such an expression is known from perturbation theory,[12]

$$\frac{8\pi^2}{g^2(\rho)} = \frac{8\pi^2}{g^2(\rho_0)} + N\left[\frac{11}{3}\ln\frac{\rho_0}{\rho} + \frac{17}{11}\ln\left(1 + \frac{11}{3}N\frac{g^2(\rho_0)}{8\pi^2}\ln\frac{\rho_0}{\rho}\right)\right], \tag{82}$$

where we have given the result for the group SU(N) (without the contribution of fermions). The unrenormalized constant is $g_0 = g(\rho_0 = 1/M)$. The instanton contribution to the vacuum–vacuum transition for the group SU(2) has the form

$$\frac{\langle 0 \mid 0_T \rangle_{\text{ins}}^{\text{Reg}}}{\langle 0 \mid 0_T \rangle_{\text{p. th}}} = \text{const} \cdot \left[\frac{8\pi^2}{g^2(\rho)} \right]^4 e^{-8\pi^2/g^2(\rho)} \left(1 + O\left(g^2(\rho)\right)\right), \tag{83}$$

where $g^2(\rho)$ is given by the expression (82) with $N = 2$. For the factor $[8\pi^2/g^2(\rho)]^4$, we can restrict ourselves to the single-loop expression for $g^2(\rho)$, the difference being of the order of the ignored terms which give relative corrections of order $g^2(\rho)$. Note that the complete two-loop calculation of the instanton contribution would determine these corrections.

The proof of the correctness of (83) is based on the renormalizability of the theory and the method of effective Lagrangians. In the functional integral, we integrate in the spirit of Wilson over fields of small scale (less than ρ_c), i.e., over configurations corresponding to instantons with small $\rho < \rho_c$. As a result, we obtain an effective Lagrangian of the fields with scales greater than ρ_c. In this Lagrangian, the small-scale fluctuations are taken into account in the coefficients of the expansion with respect to the operators.

The calculation of the contribution of the instantons to the vacuum–vacuum transition is equivalent to determination of their contribution to the coefficient of the identity operator. The calculation of the coefficients of the other operators will be considered in Sec. 10. A specific feature of the identity operator is the fact that its matrix elements are independent of the normalization point; ρ_c is the zero-frequency anomalous dimension. Therefore, the coefficient of it, expressed in terms of $g(\rho)$, cannot contain ρ_c (for operators with positive-frequency anomalous dimension the factor $[g^2/(\rho_c)/g^2(\rho)]^6$ arises).

It now only remains to express $g^2(\rho)$ in terms of $g^2(\rho_0)$ by means of the renormalization-group equations, and the retention of the two-loop correction in (82) is fully valid.

e) Density of instantons in the group SU(N)

How does the number of zero-frequency modes change on the transition to the group SU(N)? We have already said that the instanton field uses only a SU(2) subgroup of the complete group. Suppose this subgroup occupies the top left-hand corner in the $N \times N$ matrix of generators. It is clear that the five zero-frequency modes associated with shifts and dilatations remain the same as in the group SU(2), and only the modes associated with group rotations are changed. In SU(2) there were three, and in SU(N) they correspond to three generators in a 2×2 matrix at the top left (Fig. 6). Those of the remaining generators that occur in the $(N-2) \times (N-2)$ matrix in the bottom right obviously do not rotate the instanton field. Thus, to the three SU(2) rotations there are added a further $4(N-2)$ unitary rotations. The total number of zero-frequency modes is $5 + 3 + 4(N-2) = 4N$. Of course, this number $4N$ exactly corresponds to the coefficient of the "antiscreening" logarithm in the formula for $8\pi^2/g^2(\rho)$. Finally, we write down an expression for the reduced instanton density $d(\rho)$, which is defined as follows:

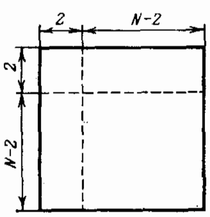


FIG. 6.

$$\frac{\langle 0 \mid 0_T \rangle_{\text{ins}}^{\text{Reg}}}{\langle 0 \mid 0_T \rangle_{\text{p. th}}} = \int \frac{d^4x\, d\rho}{\rho^5} d(\rho). \tag{84}$$

The function $d(\rho)$ is equal to

$$d(\rho) = \frac{C_1}{(N-1)!\,(N-2)!} \left[\frac{8\pi^2}{g^2(\rho)} \right]^{2N} e^{-[8\pi^2/g^2(\rho)] - C_2 N}, \tag{85}$$

where $g^2(\rho)$ is expressed in terms of $g_0^2 = g^2(\rho_0 = 1/M)$ by formula (82), and the constants C_1 and C_2 can be found by a certain modification of 't Hooft's calculations.[13] Concretely, it is necessary to take into account a further $4(N-2)$ vector fields with the above quantum numbers in both the zero-frequency and the positive-frequency modes. In addition, we require the embedding volume of SU(2) in SU(N); the factor $[(N-1)!\,(N-2)!]^{-1}$ is associated with it. This part of the modification proved to be the most complicated (see Ref. 13). The result for C_1 and C_2 has the form

$$C_1 = \frac{2e^{5/6}}{\pi^2} = 0.466,$$

$$C_2 = \frac{5}{3} \ln 2 - \frac{17}{36} + \frac{1}{3} (\ln 2\pi + \gamma) + \frac{2}{\pi^2} \sum_{s=1}^{\infty} \frac{\ln s}{s^2} = +1.679. \tag{86}$$

Note that the constant C_2 depends on the method of regularization, which actually provides the definition of the unrenormalized constant. Instead of Pauli–Villars regularization (PV scheme), so-called dimensional regularization is frequently used. Instead of logarithms of the cutoff parameter, poles with respect to the dimension of space arise in this method, $\ln M \to 1/(4-D)$. Use of the minimal scheme[14] (MS) for determining the coupling constant leads to an expression of the form (85) with the substitution

$$g(\rho) \to g_{\text{MS}}(\rho) \qquad C_2 \to C_{2\text{MS}},$$

$$C_{2\text{MS}} = C_2 - \frac{5}{36} - \frac{11}{6} (\ln 4\pi - \gamma) = C_2 - 3.721. \tag{87}$$

The numerical coefficient in $d(\rho)$ for the MS scheme is $e^{3.72N}$ times greater than in the PV scheme, which for SU(3) gives the factor $\sim 7 \cdot 10^4$.

Of course, the relations between the observable amplitudes do not depend on the definition of g^2—the same conversion constants associated with the change of regularization occur, for example, in the corrections in g^2 to the cross section of e^+e^- annihilation into hadrons (though there, it is true, the dependence on them is not exponential). We note in this connection that in perturbation theory the $\overline{\text{MS}}$ scheme has proved helpful, since in it too large coefficients of the expansion in g^2 do not arise.[15] The difference between the $\overline{\text{MS}}$ scheme and the MS scheme reduces to the substitution

$$\begin{aligned} \frac{8\pi^2}{g^2_{\overline{MS}}} &= \frac{8\pi^2}{g^2_{MS}} - \frac{11}{6} N (\ln 4\pi - \gamma), \\ C_{2\overline{MS}} &= C_2 - \frac{5}{36} \approx 1.54. \end{aligned} \tag{88}$$

We give finally the explicit form of the dependence on ρ for the function $d(\rho)$:

$$d(\rho)=\frac{0.466}{(N-1)!(N-2)!}\left(\frac{\rho}{\rho_0}\right)^{11N/3}\left[1+\frac{11}{3}N\frac{g^2(\rho_0)}{8\pi^2}\ln\frac{\rho}{\rho_0}\right]^{5N/11} \times\left[\frac{8\pi^2}{g^2(\rho_0)}\right]^{2N}e^{-[8\pi^2/g^2(\rho_0)]-1.679N}. \tag{89}$$

9. INSTANTON GAS AND GENERAL THEOREMS

The calculated instanton contribution is proportional to $\int d^4x_0=V_4$, the volume of the considered region of the Euclidean space. As long as $V_4\rho^{-4}d(\rho)$, the probability of finding an instanton of scale ρ in the considered volume, is a small quantity, one can ignore fluctuations for which there are two or more instantons of scale ρ in this volume. But with increasing V_4, we naturally arrive at the need to consider an instanton gas.

As in the one-dimensional case, the vacuum–vacuum transition has the form

$$\langle 0\mid 0_T\rangle=\langle 0\mid \exp\left(-\int d^4x\mathscr{H}\right)\mid 0\rangle=e^{-\varepsilon V_4}, \tag{90}$$

where $\mathscr{H}$ is the Hamiltonian density, and ε can be called the vacuum energy density. Clearly, for the summation it is convenient to consider the logarithm of (90), i.e., the quantity ε:

$$\varepsilon=-\frac{1}{V_4}\ln\langle 0\mid 0_T\rangle=-\frac{1}{V_4}\ln\left[\langle 0\mid 0_T\rangle_{\text{p. th}}+\langle 0\mid 0_T\rangle_{\text{ins}}\right] \approx-\frac{1}{V_4}\left[\ln\langle 0\mid 0_T\rangle_{\text{p. th}}+\frac{\langle 0\mid 0_T\rangle_{\text{ins}}}{\langle 0\mid 0_T\rangle_{\text{p. th}}}\right]\approx\varepsilon_{\text{p. th}}-\int\frac{d\rho}{\rho^5}d(\rho). \tag{91}$$

Thus, the correction to the vacuum energy density in the gas approximation is negative and given by the integral $\int d\rho\,\rho^{-5}d(\rho)$.

Due to the power-law growth $d(\rho)\sim\rho^{11N/3}$, this integral is determined by large ρ, and the formal expression diverges as a power.

Unfortunately, $d(\rho)$ is known only in the region of fairly small ρ, which must be such as to guarantee that the ignored quantum corrections $\sim g^2(\rho)$ are small. In addition, ε contains a contribution of fluctuations with topological charge $|n|>1$, which, roughly speaking, is proportional to $[d(\rho)]^n$. Both these effects have the consequence that formula (91) does not hold at large ρ. Of particular interest is the possibility that fluctuations with large topological charge become important in the region of scales for which the corrections $\sim g^2(\rho)$ are still small. Such a situation appears all the more plausible because $d(\rho)$ increases with ρ much more rapidly than $g^2(\rho)$. The two-dimensional models analyzed in the interesting papers of Ref. 16 provide an example in which dense fluctuations with large topological charge are dominant in the vacuum wave function. In Ref. 16, this antigas situation was called melting of instantons.

The approximation of a dilute instanton gas was developed in Ref. 8. The approximation is based on the hypothesis that the phenomena associated with large ρ reduce effectively to the appearance of an upper limit ρ_m in the integral over ρ, and for all $\rho<\rho_m$ one can use the one-instanton formula (91) for $d(\rho)$.

In this subsection we shall demonstrate that the dilute gas hypothesis is not self-consistent by giving an example which violates a general relation. In the following section, we shall explicitly find the region of ρ in which the one-instanton expressions are valid on the basis of phenomenological information about the fields in the QCD vacuum. We shall see that the admissible ρ are too small to make a claim to a description of the vacuum structure in the region of the main scales even in order of magnitude.

The exact relation whose verification we have in mind is the connection between the vacuum energy density and the mean square intensity of the gluon field in the vacuum. For the derivation, we consider the vacuum expectation value of the energy–momentum tensor $\theta_{\mu\nu}(x)$. By virtue of relativistic invariance,

$$\langle 0\mid\theta_{\mu\nu}\mid 0\rangle=g_{\mu\nu}\varepsilon, \tag{92}$$

from which, after summation, we deduce an expression for ε in terms of the vacuum expectation value of the trace of the energy–momentum tensor:

$$\varepsilon=\frac{1}{4}\langle 0\mid\theta_{\mu\mu}\mid 0\rangle. \tag{93}$$

For $\theta_{\mu\mu}$ the gluodynamics with group SU(N) the following operator expression holds[17]:

$$\theta_{\mu\mu}=\frac{\beta(\alpha_s)}{4\alpha_s}G^a_{\mu\nu}G^a_{\mu\nu}\qquad(a=1,\ \dots,\ N^2-1), \tag{94}$$

where $\alpha_s=g^2/4\pi$, and $\beta(\alpha_s)$ is the Gell-Mann–Low function,

$$\beta(\alpha_s)=\mu\frac{d\alpha_s(\mu)}{d\mu}=-\frac{11}{3}N\frac{\alpha_s^2}{2\pi}+O(\alpha_s^3). \tag{95}$$

The expression (94) for $\theta_{\mu\mu}$ is called the trace anomaly of the energy–momentum tensor. The point is that for a classical massless vector field $\theta_{\mu\mu}=0$. The difference from zero appears at the single-loop level and is associated with the need to introduce a guage-invariant cutoff.

The appearance in $\theta_{\mu\mu}$ of the function $\beta(\alpha_s)$, which controls the charge renormalization, can be explained as follows. The stretching $x\to\lambda x$ of all scales is determined in the infinitesimal form of the transformation by the dilatation operator D:

$$D=\int d^3xD_0(x),\qquad D_\mu(x)=\theta_{\mu\nu}x_\nu.$$

It is readily seen that there is invariance with respect to dilatations only when the divergence of the dilatation current D_μ vanishes. This divergence is

$$\partial_\mu D_\mu=\theta_{\mu\mu},$$

i.e., the operator $\theta_{\mu\mu}$ determines the noninvariance under dilatations. The noninvariance indicates the existence of a certain distance scale. In a massless theory, the only possibility for a scale to appear is associated with the need to introduce the cutoff parameter M when considering the quantum effects. Under the simultaneous transformations $x\to\lambda x, M\to M/\lambda$ the theory is invariant, i.e., dilatations are equivalent to a change in M. It is for this reason that $\theta_{\mu\mu}$ is proportional to $\beta(\alpha_s)=M d\alpha_s/dM$.

Considering the action of dilatation transformations on transition amplitudes expressed in the form of path integrals, we can readily deduce the relation (94). We shall not give this derivation but restrict ourselves to two comments about it:

a) energy–momentum conservation, $\partial_\mu \theta_{\mu\mu}=0$, has the consequence that the right-hand side of the relation is independent of the normalization point μ of the operator $G^a_{\mu\nu}G^a_{\mu\nu}$, and the effective charge α_s must be taken at the same point, $\alpha_s=\alpha_s(\mu)$;

b) the quantum corrections also lead to a cutoff-dependent c-number part in $\theta_{\mu\mu}$. Therefore, the more accurate expression is

$$\theta_{\mu\mu}=(0\mid\theta_{\mu\mu}\mid 0)_{\text{p. th}}+\frac{\beta(\alpha_s)}{4\alpha_s}G^a_{\mu\nu}G^a_{\mu\nu}, \tag{96}$$

where the c-number part is separated by averaging over the perturbation theory vacuum $|0)$ (which differs from the exact physical vacuum $|0\rangle$).

Substituting the expression (96) for $\theta_{\mu\mu}$ in (93), we arrive at the desired connection between ε and the mean square of the field intensity in the vacuum:

$$\varepsilon=\varepsilon_{\text{p. th}}+\frac{\beta(\alpha_s)}{16\alpha_s}\langle 0\mid G^a_{\mu\nu}G^a_{\mu\nu}\mid 0\rangle. \tag{97}$$

It is clear from the derivation that in $\langle 0|G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle$ it is necessary to take into account only those fluctuations not given by perturbation theory.

Instantons are an example of such fluctuations. The one-instanton contribution to $\langle 0|G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle$ can be readily found, for which it is necessary to go over to the Euclidean space (above, we have used the notation of Minkowski space), replace the field $\hat{G}^a_{\mu\nu}$ by the instanton field, and add the factor $d^4x_0 d\rho\,\rho^{-5}d(\rho)$, the probability of finding an instanton with scale ρ with center at x_0. The result for the one-instanton contribution to $\varepsilon-\varepsilon_{\text{p. th}}$, integrated over x_0 [the integral is $\int d^4x_0 G^a_{\mu\nu}(x-x_0)\hat{G}^a_{\mu\nu}(x-x_0)=\int d^4x G^a_{\mu\nu}(x)\hat{G}^a_{\mu\nu}(x)=4\cdot 8\pi^2/g^2$], is

$$\varepsilon-\varepsilon_{\text{p. th}}=-\frac{11}{12}N\int\frac{d\rho}{\rho^5}d(\rho) \tag{98}$$

[for $\beta(\alpha_s)$ we have used the single-loop approximation]. On the other hand, the one-instanton contribution to $\varepsilon-\varepsilon_{\text{p. th}}$ is given by the expression (91) obtained earlier and differs from (98) by the absence of the factor $11N/12$.

What does this mean? Since the integral over ρ is determined by large ρ, the paradox is resolved by noting that the one-instanton approximation does not give the possibility of finding $\varepsilon-\varepsilon_{\text{p. th}}$ and $\langle 0|G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle$. Moreover, the attempt to take into account the effects other than the one-instanton effects by introducing a cutoff in ρ is inconsistent in that this cannot be done in a unified manner even for quantities associated with general relations—the cutoff in them is effectively different.

To conclude the section, we note that the inadequacy of the one-instanton approximation for quantities such as ε or $\langle 0|[\beta(\alpha_s)/\alpha_s]G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle$ can also be proved by a somewhat different argument. Physical quantities, of course, are independent of the normalization point. For such quantities having dimension m^4, the normalization point μ can occur only in the (μ-independent) combination

$$\left\{\mu\left[\frac{2\pi}{\alpha_s(\mu)}\right]^{b_1/b}e^{-(1/b)2\pi/\alpha_s}\right\}^4[1+O(\alpha_s(\mu))], \tag{99}$$

where b and b_1 are the first and second coefficients in the expansion of the Gell-Mann–Low function:

$$\beta(\alpha_s)=-b\frac{\alpha_s^2}{2\pi}-b_1\frac{\alpha_s^3}{(2\pi)^2}+O(\alpha_s^4). \tag{100}$$

In SU(N) gluodynamics [see (82)],

$$b=\frac{11}{3}N,\quad \frac{b_1}{b}=\frac{17}{11}N. \tag{100a}$$

On the other hand, the one-instanton approximation with a cutoff of the integral over ρ at the upper limit at ρ_m gives for the same parameters a result proportional to

$$\frac{1}{\rho_m^4}d(\rho_m)\sim\frac{1}{\rho_m^4}\left[\frac{2\pi}{\alpha_s(\rho_m)}\right]^{2N}e^{-2\pi/\alpha_s(\rho_m)}, \tag{101}$$

where we have used the fact that the main contribution is made by the region of ρ near ρ_m. Comparing (101) and (99) for $\mu=1/\rho$, we see that the dependence on ρ_m does not agree with that required by renormalization invariance. The power of ρ_m is greater by the same $b/4=11N/12$ times, and the power of $\ln\rho_m$ also does not agree.

10. INSTANTONS IN THE QCD VACUUM

As we have already said, the main fluctuations in the QCD vacuum are those of large scales of the order of the confinement radius or, which is the same thing, the radius of hadrons. Unfortunately, we are not yet able to treat such fluctuations quantitatively.

The quasiclassical methods that have been developed apply to the study of nonperturbation-theoretical fluctuations of small scale, among which the instantons are dominant.

In this subsection we take into account the influence on the small-scale instantons of the fields due to the characteristic long-wavelength fluctuations in the vacuum.[18]

Since we distinguish fields of two types, namely, the fields of small-scale instantons and the fields of the characteristic vacuum fluctuations, it is convenient to introduce an effective Lagrangian. In it, as usual, the contribution of the rapidly varying fields is included in the coefficients of the various operators that act on the space of the slowly varying fields.

Thus, the effect of a distinguished instanton with scale ρ and center at x_0 reduces to the following correction to the effective Lagrangian of the long-wavelength fluctuations:

$$\Delta L(x_0)=\frac{d\rho}{\rho^5}\sum_n C_n(\rho)O_n(x_0),$$

where $C_n(\rho)$ are numerical coefficients and $O_n(x_0)$ are local operators constructed from the gluon fields (we consider pure gluodynamics; the changes introduced by fermions are discussed in the following section).

The probability of finding the instanton under consideration in the physical vacuum is given by averaging ΔL over this state. On the other hand, to find the coefficients C_n, it is necessary to consider the matrix elements of ΔL between perturbation-theory states (with different number of free gluons with momenta $q \ll 1/\rho$). These matrix elements can be calculated by quasiclassical methods.

Concretely, we consider the instanton contribution to the vacuum $\to n$ gluons transition and apply to it the reduction formula

$$\langle n \text{ gluons} \mid \Delta L \mid 0\rangle = \langle 0 \mid T \prod_{k=1}^{n} \int dx_k e^{iq_k x_k} \varepsilon^{a_k}_{\mu_k} q_k^2 A^{a_k}_{\mu_k}(x_k) \mid 0\rangle, \quad (102)$$

where q_k and $\varepsilon^{a_k}_{\mu_k}$ are the 4-momentum and the polarization of the kth gluon, and $A^a_\mu(x)$ is the operator of the gluon field. For $n=0$, i.e., for the vacuum–vacuum transition, the right-hand side of (102) was already calculated in Sec. 8 and is equal to $d\rho\rho^{-5}d(\rho)$; the left-hand side is obviously equal to the coefficient of the unit operator: $C_I d\rho/\rho^5$.

For $n \neq 0$, the prescription of the quasiclassical calculation of the expression (102) reduces to

a) the transition to the Euclidean space (see the equations of Sec. 5);

b) replacement of the Euclidean $\hat{A}^a_\mu(x)$ by the instanton field $\bar{A}^a_\mu(x-x_0)$ given by formula (65). The singular gauge is used because the reduction formula (102) is valid only for rapidly decreasing fields $A^a_\mu(x)$. For a nonsingular gauge, the reciprocal propagator q^2 is replaced by a more complicated expression;

c) multiplication by the $\langle 0 | 0_T\rangle_{\text{ins}}$ transition amplitude, which is equal to $d\rho\rho^{-5}d(\rho)$. Thus, for the matrix element (102) we obtain

$$\langle n \text{ gluons} \mid \Delta L(x) \mid 0\rangle = \frac{d\rho}{\rho^5} d(\rho) e^{-ix\Sigma q_k} \prod_{k=1}^{n} \left[\int dx_k e^{-iq_k x_k} (-q_k^2) \varepsilon^{a_k}_{\mu_k} \overline{A^{a_k}_{\mu_k}}(x_k) \right], \quad (103)$$

where all the quantities on the right-hand side are Euclidean.

The Fourier transform of the instanton solution, which we want in the limit $q\rho \to 0$, is readily found:

$$\int dx e^{-iqx} (-q^2) A^a_\mu(x) = \frac{4\pi i}{g} \bar{\eta}_{a\mu\nu} q_\nu \rho^2. \quad (104)$$

After this, it is easy to recover the complete operator form of ΔL:

$$\Delta L(x) = \frac{d\rho}{\rho^5} d(\rho) \exp\left[-\frac{2\pi^2}{g} \rho^2 \bar{\eta}^M_{a\mu\nu} G^a_{\mu\nu}(x) \right],$$
$$\bar{\eta}^M_{a\mu\nu} = \begin{cases} \bar{\eta}_{amn}, & \mu=m, \ \nu=n; \quad m, n=1, 2, 3, \\ i\bar{\eta}_{a4n}, & \mu=0, \ \nu=n; \quad n=1\ 2, 3, \end{cases} \quad (105)$$

where $G^a_{\mu\nu}(x)$ is the operator of the large-scale gluon field. The factorials which occur in the expansion of the exponential cancel against the combinatorial coefficients when the matrix element (103) is taken.

The expression (105) for the interaction of an instanton with an external field was obtained for the first time by Callen, Dashen, and Gross[8] by a different and more complicated method. An important point is that we, in contrast to them, have not fixed the external $G^a_{\mu\nu}(x)$ "by hand" but have related it to the field of the large-scale fluctuations.

This is achieved by averaging the Lagrangian (105) over the physical vacuum. The term linear in $G^a_{\mu\nu}$ obviously vanishes as a result of such averaging, and the first nonvanishing correction to the effective density of the instantons is proportional to G^2:

$$\langle 0 \mid \Delta L \mid 0\rangle = \frac{d\rho}{\rho^5} d_{\text{eff}}(\rho) = \frac{d\rho}{\rho^5} d(\rho) \left[1 + \frac{\pi^3\rho^4}{(N^2-1)\alpha_s} \langle 0 \mid G^a_{\mu\nu} G^a_{\mu\nu} \mid 0\rangle + O(\rho^6) \right], \quad (106)$$

where in the averaging we have used the relation

$$\langle 0 \mid G^a_{\mu\nu} G^{a'}_{\mu'\nu'} \mid 0\rangle = \frac{\delta^{aa'}}{N^2-1} \cdot \frac{1}{12} (g_{\mu\mu'} g_{\nu\nu'} - g_{\mu\nu'} g_{\nu\mu'}) \langle 0 \mid G^b_{\alpha\beta} G^b_{\alpha\beta} \mid 0\rangle. \quad (107)$$

Note that the constant α_s and the operator $(G^a_{\mu\nu})^2$ which occur here are normalized at the point ρ. A quantity that does not depend on the renormalization point [to accuracy $\alpha_s(\rho)$] is the product $\alpha_s G^a_{\mu\nu} G^a_{\mu\nu}$ (see the previous section).

To obtain a quantitative estimate of the correction, it is necessary to know the mean square of the intensity of the gluon field in the physical vacuum. This was found in Refs. 19 by analyzing the influence of the vacuum fields on the charmonium states, and it was found to be

$$\langle 0 \mid \frac{\alpha_s}{\pi} G^a_{\mu\nu} G^a_{\mu\nu} \mid 0\rangle \approx 0.012 \text{ GeV}^4. \quad (107')$$

For the group SU(3), the relative correction to $d(\rho)$ can be written in the form

$$\frac{\pi^4\rho^4}{8\alpha_s^2(\rho)} \langle 0 \mid \frac{\alpha_s}{\pi} G^a_{\mu\nu} G^a_{\mu\nu} \mid 0\rangle. \quad (108)$$

It reaches unity at a value of ρ equal to

$$\rho_{\text{crit}} \approx \frac{1}{1.15 \text{ GeV}}, \quad (109)$$

if for α_s we take $\alpha_s(\rho) = 2\pi/9 \ln(1/\Lambda\rho)$ with $\Lambda = 100$ MeV. For $\rho = \rho_{\text{crit}}$, the interaction of the instanton with the vacuum fields of the other fluctuations becomes 100% important. This ρ_{crit} is very small compared with the characteristic hadron dimensions 1/(200–300) MeV. The word "very" can indeed be used if one bears in mind the fact that $d(\rho)$ is proportional to a high power of ρ; the $\rho \leq \rho_{\text{crit}}$ contribution to, say, the vacuum energy density is extremely small.

The smallness of ρ_{crit} given by the estimate (109) can also be seen in a different way by calculating the contribution of the instantons to the correlation function

$$i \int dx e^{iqx} \langle 0 \mid T A(x) B(0) \mid 0\rangle, \quad (110)$$

where A and B are certain local operators. At large Euclidean q, the instantons make contributions of two types to (110). First, there is the contribution of the fluctuations of a fixed (q-independent) scale to the coefficients of the regular expansion in powers of $1/q^2$. Second, there is the contribution from instantons with scales $\rho \sim \rho_{\text{eff}} = C/q$, which is proportional to $d(\rho_{\text{eff}})$, i.e., to a high (and not necessarily integral) power of $1/q^2$. The constant of proportionality C can be determined by the method of steepest descent and because of the high power of ρ in $d(\rho)$ is approximately equal to 5.

Thus, we can calculate the one-instanton contribution to (110) in terms of q using the ordinary expressions only when $q^2 > (5.5\ \mathrm{GeV})^2$. It is clear that such q^2 are considerably greater than the characteristic hadron masses.

We conclude this subsection by giving a formula that takes into account the higher powers of $G^a_{\mu\nu}$ in the effective instanton density. This formula is based on the hypothesis of dominance of the vacuum intermediate state, which makes it possible to reduce $\langle 0|(G^2)^n|0\rangle$ to $(\langle 0|G^2|0\rangle)^n$. This approximation is analogous to one used in many-body theory and for some 4-quark operators for which it can be verified has an accuracy of the order of a few percent.

The factorization leads to the relation

$$\langle 0|\left(\frac{2\pi^2}{g}\rho^2\bar{\eta}^M_{a\mu\nu}G^a_{\mu\nu}\right)^{2k}|0\rangle=(2k-1)!!\left[\frac{4\pi^4}{g^2}\rho^4\langle 0|(\bar{\eta}^M_{c\mu\nu}G^a_{\mu\nu})^2|0\rangle\right]^k,$$

by means of which we obtain for the effective instanton density the result

$$d_{\mathrm{eff}}(\rho)=d(\rho)\exp\left[\frac{\pi^4\rho^4}{(N^2-1)\alpha_s^2(\rho)}\langle 0|\frac{\alpha_s}{\pi}G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle\right], \qquad (111)$$

which can be represented as the replacement in the expression for $d(\rho)$ of $2\pi/\alpha_s(\rho)$ by

$$\frac{2\pi}{\alpha_s(\rho)}\to\frac{2\pi}{\alpha_s(\rho)}\left[1-\frac{\pi^3\rho^4}{2(N^2-1)\alpha_s^2(\rho)}\langle 0|\frac{\alpha_s}{\pi}(G^a_{\mu\nu})^2|0\rangle\right]. \qquad (112)$$

Using for $d(\rho)$ the expression (111), we can advance in ρ to $\rho > \rho_{\mathrm{crit}}$. However, when the interaction with the vacuum fields changes the classical action strongly, i.e., when (112) vanishes, the quasiclassical methods cannot be used. This limit under the same assumptions about α_s and $\langle 0|G^2|0\rangle$ is $\rho < 1/500$ MeV.

Despite the numerical uncertainties in the value of $\langle 0|G^2|0\rangle$ (which are of the order of a factor 2) and in α_s (the uncertainty in Λ is also of the order of a factor 2), it can be said that the vacuum fields deform the instantons at scales much smaller than the characteristic scales of the fluctuations that are dominant in the vacuum.

11. FERMIONS IN AN INSTANTON FIELD

In this section, we shall relatively briefly discuss how the instanton contribution to the vacuum–vacuum transition amplitude changes when fermions are included in the theory.

It is immediately clear that for a fluctuation with a given scale ρ the influence of "heavy" quarks with mass $m \gg \rho^{-1}$ is small; for in this case the quarks appear at times and distances $\sim 1/m \ll \rho$, at which perturbation theory can be used to calculate the quark loops of the form shown in Fig. 7. We give the first few terms of the effective Lagrangian that takes into account the fermion loops:

$$\Delta L_F=\frac{1}{2}\mathrm{Tr}\Big\{-\frac{1}{4}G^2_{\mu\nu}\times\frac{g^2}{24\pi^2}\ln\frac{M^2}{m^2}+\frac{1}{16\pi^2}\Big(\frac{ig^3}{180m^2}G_{\mu\nu}G_{\nu\gamma}G_{\gamma\mu}$$
$$+\frac{g^4}{288m^4}\Big[-(G_{\mu\nu}G_{\mu\nu})^2+\frac{7}{10}\{G_{\mu\gamma},G_{\gamma\nu}\}^2_+ +\frac{29}{70}[G_{\mu\gamma},G_{\gamma\nu}]^2_- -\frac{8}{35}[G_{\mu\nu},G_{\gamma\delta}]^2_-\Big]\Big)\Big\},$$
$$G_{\mu\nu}=G^a_{\mu\nu}t^a,\quad \mathrm{Tr}\,t^at^b=2\delta^{ab}. \qquad (113)$$

The first term in this expression contains the cutoff parameter M and, obviously, describes the contribution of the quark under consideration to the change in the charge g. Therefore, it is automatically taken into account when the result is expressed in terms of the charge at distances greater than $1/m$.

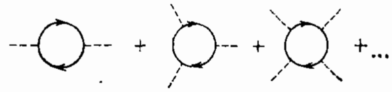


FIG. 7.

The following terms in (113) give a series in powers of $1/m^2\rho^2$ on the transition to the Euclidean space and substitution of the instanton field.

We now turn to the limiting case of "light" quarks, $m\rho \ll 1$. We note that for sufficiently small instantons all quarks are light. We calculate the integral over the Fermi fields in the path integral that determines the vacuum–vacuum transition: $\langle 0|0_T\rangle$. In the Euclidean action, a fermion with mass m adds a term of the form [see (45)]

$$S_F^{(E)}=\int d^4x\bar{\psi}(-i\gamma_\mu D_\mu-im)\mu,$$

and integration of this with respect to the anticommuting fields leads to

$$\mathrm{Det}\,(-i\gamma_\mu D_\mu-im).$$

The determinant can be understood as a product of the eigenvalues of the corresponding operator,

$$\mathrm{Det}\,(-i\gamma_\mu D_\mu-im)=\prod_n(\lambda_n-im),$$

where the real numbers λ_n are the eigenvalues of the Hermitian operator $-i\gamma_\mu D_\mu$:

$$-i\gamma_\mu D_\mu u_n(x)=\lambda_n u_n(x). \qquad (114)$$

Of fundamental importance in the study of the limit $m=0$ is the question of whether certain λ_n vanish, i.e., the question of zero-frequency modes of the fermion field. We shall show that the interaction with the instanton field leads to the appearance of one such mode u_0,

$$-i\gamma_\mu D_\mu u_0=0. \qquad (115)$$

We go over to two-component spinors $\chi_{L,R}$ (we use the standard representation for the γ matrices):

$$u_0=\begin{pmatrix}1\\-1\end{pmatrix}\chi_L+\begin{pmatrix}1\\1\end{pmatrix}\chi_R,\quad \sigma^+_\mu D_\mu\chi_L=0,\quad \sigma^-_\mu D_\mu\chi_R=0, \qquad (116)$$

where $\sigma^\pm_\mu=(\sigma,\mp i)$. To the equations for χ_L,χ_R we apply the operators $\sigma^-_\mu D_\mu, \sigma^+_\mu D_\mu$, respectively. Using the relations (57), the commutator $[D_\mu D_\nu]=-(ig/2)\tau^a G^a_{\mu\nu}$, and the explicit form of $G^a_{\mu\nu}$ [see (63)], we obtain

$$-D^2_\mu\chi_L=0,\qquad -D^2_\mu\chi_R=-4\sigma\tau\frac{\rho^2}{[(x-x_0)^2+\rho^2]^2}\chi_R.$$

The operator $-D^2_\mu$ is a sum of the squares of Hermitian operators: $-D^2=(-iD_\mu)^2$, i.e., it is positive definite. Therefore, it does not have vanishing eigenvalues (the boundary conditions are imposed at a large but finite distance R) and, therefore $\chi_L=0$.

In the equation for χ_R, we use a basis in the space of spinor and color indices that diagonalizes the matrix $\sigma\tau$. We recall that σ acts on the spinor indices, and τ on the color indices. This basis corresponds to addition of the ordinary spin and the color spin to a total angular momentum equal to zero (when $\sigma\tau=-3$) or unity ($\sigma\tau=+1$). It again follows from the positive definiteness of $-D_\mu^2$ that the only suitable case for us is when the total spin is equal to zero, which completely determines the dependence of χ_R on the indices:

$$(\sigma+\tau)\chi_R=0, \quad \chi_R^{\alpha m}\sim\varepsilon^{\alpha m}, \tag{117}$$

where $\alpha=1,2$ and $m=1,2$ are the spin and color indices, respectively.

The dependence on the coordinates can be readily found from the explicit form of D_μ^2, and the final result for the zero-frequency mode $u_0(x-x_0)$ (normalized by the condition $\int u^+u\,dx=1$) has the form

$$u_0(x)=\frac{1}{\pi}\frac{\rho}{(x^2+\rho^2)^{3/2}}\begin{pmatrix}1\\1\end{pmatrix}\varphi, \quad \varphi^{\alpha m}=\frac{1}{\sqrt{2}}\varepsilon^{\alpha m}. \tag{118}$$

We also write down the expression for the zero-frequency mode in the singular gauge, $u_0^{\text{sing}}(x-x_0)$ (which we shall require),

$$u_0^{\text{sing}}(x)=\frac{1}{\pi}\frac{\rho}{(x^2+\rho^2)^{3/2}}\frac{x_\mu\gamma_\mu}{\sqrt{x^2}}\begin{pmatrix}1\\-1\end{pmatrix}\varphi, \tag{119}$$

it being obtained by multiplication of (118) by the gauge transformation matrix (64a).

We now turn to the instanton part of the vacuum–vacuum transition amplitude. In it, we have the factor

$$F=\frac{m}{M}\frac{\mathrm{Det}'(-i\gamma_\mu \hat{D}_\mu)}{\mathrm{Det}'(-i\gamma_\mu D_\mu-iM)}\frac{\mathrm{Det}(-i\gamma_\mu\partial_\mu-iM)}{\mathrm{Det}(-i\gamma_\mu\partial_\mu)},$$

where Det′ denotes the determinant without the zero-frequency mode and we have taken into account the regularization and also the normalization by perturbation theory. In all the positive-frequency modes, m is taken equal to zero, so that after the separation in F of the factor m/M the remaining part depends only on the dimensionless parameter $M\rho$. As in pure gluodynamics (see Sec. 8), this dependence must be such that the cutoff parameter M is removed by a renormalization of the coupling constant, i.e., the dependence of F on $M\rho$ must give the renormalization of the coupling constant due to the fermions in the factor $e^{-8\pi^2/g_0^2}$,

$$\Delta_F\frac{8\pi^2}{g^2}=-\ln\frac{F}{m\rho\cdot\mathrm{const}}=\ln M\rho-\frac{1}{3}\ln M\rho. \tag{120}$$

The first logarithm derives from the zero-frequency mode, the second from the positive-frequency modes. Comparing the result with formula (81) for gluons, we see that the situation has been changed because of the anticommutativity: The zero-frequency modes of the light quarks lead to screening of the charge, and the positive-frequency modes to antiscreening.

In ordinary perturbation theory, the splitting in (120) can be associated with the spin-dependent part of the interaction (the first logarithm) and the "charge" part, which is not associated with the spin (the second logarithm). Indeed, the imaginary part of the gluon polarization operator, which derives from the intermediate $q\bar{q}$ state, can be represented in the form

$$\mathrm{Im}\,\Pi_{\mu\nu}^{Fab}=\delta^{ab}\frac{g^2}{2}\int\frac{do}{32\pi^2}[q_\mu q_\nu-g_{\mu\nu}q^2-(p_1-p_2)_\mu(p_1-p_2)_\nu]$$
$$=\delta^{ab}\frac{g^2}{16\pi}(q_\mu q_\nu-q_{\mu\nu}q^2)\left(1-\frac{1}{3}\right). \tag{121}$$

In this formula, p_1 and p_2 are the particle and antiparticle momenta, $q=p_1+p_2$, and the integration is over the directions of $\mathbf{p}_1=-\mathbf{p}_2$ in the center-of-mass system. The second term in (121) differs by only the factor -2 from the contribution of a spinless color doublet. The factor 2 corresponds to the two polarization states, and the minus to the anticommutativity.

We note that for the vacuum polarization there is also an analogous relation between the spin part of the polarization and the zero-frequency modes. This is readily seen in the "background" gauge obtained by adding the term (71) to the action. In perturbation theory, one can take as the "external" field, for example, a potential that has only a third color component, and in the loop only "charged" components will propagate. The three-gluon vertex in this gauge has the form of a sum of a color part and a magnetic part, which do not interfere in the polarization operator. The spin part gives the "antiscreening" logarithm, and the charge part (together with the Higgs particles) the "screening" part.

What has been obtained from the inclusion in the theory of a light quark? In the limit $m\to 0$, the vacuum–vacuum transition amplitude tends to zero. Does this mean that for $m=0$ there are no tunnel transitions? By no means. The point is that now the instanton fluctuation couples the vacuum to states of a quark–antiquark pair.

To see this, we consider the crossing process—the transition from a single-quark state to a single-quark state; we shall assume that the quark momenta p and p' are small compared with $1/\rho$. Proceeding as in Sec. 10, we use the reduction formula

$$\langle p'|p_T\rangle=-\int dx\,dx'\,e^{ip'x'-ipx}\bar{v}_\alpha^m(\hat{p}')_{\alpha\gamma}\langle 0|\mathrm{T}\{q_\gamma^m(x')\bar{q}_\beta^k(x)\}|0\rangle_{\text{ins}}(\hat{p})_{\beta\delta}v_\delta^k, \tag{122}$$

where $\bar{v}_\alpha^m$ and v_δ^k are the spinors that describe the final and the initial quark (the superscript is the color index, the subscript the spinor index).

We find the instanton contribution to the fermion Green's function by using the relation

$$\langle 0|\mathrm{T}\{q_\gamma^m(x')\,q_\beta^k(x)\}|0\rangle_{\text{ins}}\xrightarrow[x_0\to -ix_4]{}\sum_n\frac{u_{(n)\gamma}^m(x')\,u_{(n)\beta}^{+k}(x)}{m+i\lambda_n}\langle 0|0_T\rangle_{\text{ins}}. \tag{123}$$

In the limit $m\to 0$, the zero-frequency mode makes the main contribution, and (123) is finite at $m=0$.

Using the explicit form (119) of the zero-frequency mode in the singular gauge, we can now readily obtain the result. We formulate it in the form of the expression for the effective Lagrangian that describes all transitions which arise from an instanton fluctuation with scale ρ:

$$\Delta L(x)=\prod_q\left[m_q\rho-2\pi^2\rho^3\bar{q}_R\left(1+\frac{i}{4}\tau^a\bar{\eta}_{a\mu\nu}\sigma_{\mu\nu}\right)q_L\right]$$
$$\times\exp\left(-\frac{2\pi^2}{g}\rho^2\bar{\eta}_{b\gamma\delta}G_{\gamma\delta}^b\right)d_0(\rho)\frac{d\rho}{\rho^5}d\hat{o}. \tag{124}$$

This contains a product over all species of light ($m_q\rho\ll 1$) quarks, and $\sigma_{\mu\nu}=(\gamma_\mu\gamma_\nu-\gamma_\nu\gamma_\mu)/2$. In Minkowski space, the symbols $\eta_{a\mu\nu}$ differ from the Euclidean symbols only when μ or $\nu=0$, and then by a factor i. By $d\hat{o}$ we denote the differential corresponding to the color orientation of the instanton, and it is normalized to unity, $\int d\hat{O}=1$. A dependence on the orientation enters through the substitution $\bar{\eta}_{\alpha\mu\nu}\to h_{aa'}\bar{\eta}_{a'\mu\nu}$ (h is the matrix of rotations in the color space), which must be made in (124). The quantity $d_0(\rho)$ differs from $d(\rho)$ (85) in pure gluodynamics by multiplication by the factor

$$\exp F\left[-\frac{1}{3}\ln 2-\frac{17}{36}+\frac{1}{3}(\ln 2\pi+\gamma)+\frac{2}{\pi^2}\sum_{s=1}^{\infty}\frac{\ln s}{s^2}\right]=e^{0.292F},$$

where F is the number of light fermions. This is for Pauli–Villars regularization; for the MS scheme, the 0.292 is replaced by −0.495 and by 0.153 for the $\overline{\rm MS}$ scheme. In addition, in the expression (82) for $8\pi^2/g^2(\rho)$ it is necessary to include the fermion contribution.

For the anti-instanton, ΔL is obtained from (124) by the substitution $\bar{\eta}_{a\mu\nu}\to\eta_{a\mu\nu}$, $q_{L,R}\to q_{R,L}$. Note also that all the operators, the constant g, and the masses m_q in ΔL are normalized at the point ρ, so that besides the dependence given explicitly there is a logarithmic dependence on ρ, which is determined by the anomalous dimension of the operator term in ΔL under consideration.

Of particular interest are the instanton-generated fermion vertices; this interaction is frequently called the 't Hooft determinant interaction. The point is that it explicitly demonstrates the breaking of the $U(1)$ symmetry associated with transformations of the form $q'=e^{i\alpha\gamma_5}q$. Naively, such a symmetry holds in a theory with massless quarks. The nontrivial nature of the breaking of this symmetry can be seen from the fact that, for example, in a theory with one quark ΔL describes the transition of a "left-handed" quark into a "right-handed" one, which is impossible in any finite order of perturbation theory for $m=0$.

[1]A. A. Belavin, A. M. Polyakov, A. S. Schwartz (Shvarts), and Yu. S. Tyupkin, Phys. Lett. **B59**, 85 (1975).
[2]A. M. Polyakov, Nucl. Phys. **B120**, 429 (1977).
[3]R. P. Feynman and A. R. Hibbs, Quantum Mechanics and Path Integrals, New York (1965) [Russian translation published by Mir, Moscow (1968)].
[4]I. Gradshtein and I. A. Ryzhik, Tablitsy integralov, summ, ryadov i proizvedenii, Fizmatgiz, Moscow (1962); English translation: Tables of Integrals, Series, and Products, Academic Press, New York and London (1965).
[5]S. Coleman, "The uses of instantons," Preprint HUTP-78/004 (1977).
[6]G. 't Hooft, Phys. Rev. D **14**, 3432 (1976).
[7]L. D. Landau and E. M. Lifshitz, Kvantovaya mechanika, Fizmatgiz, Moscow (1963); English translation: Quantum Mechanics, 2nd ed., Pergamon Press, Oxford (1965).
[8]C. G. Callan, R. Dashen, and D. J. Gross, Phys. Rev. D **17**, 2717 (1978); **19**, 1826 (1979).
[9]J. F. Willemsen, Phys. Rev. D **20**, 3292 (1979).
[10]M. F. Atiah, N. J. Hitchin, V. G. Drinfeld, and Yu. I. Manin, Phys. Lett. **A65**, 185 (1978); V. G. Drinfel'd and Yu. I. Manin, Yad. Fiz. **29**, 1646 (1979) [Sov. J. Nucl. Phys. **29**, 845 (1979)].
[11]I. B. Khriplovich, Yad. Fiz. **10**, 409 (1969) [Sov. J. Nucl. Phys. **10**, 235 (1970)].
[12]D. R. T. Jones, Nucl. Phys. **B75**, 531 (1974).
[13]C. Bernard, Phys. Rev. D **19**, 3013 (1979).
[14]G. 't Hooft and M. Veltman, Nucl. Phys. **B144**, 189 (1972); G. 't Hooft, Nucl. Phys. **B162**, 444 (1973).
[15]W. A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta, Phys. Rev. D **18**, 3998 (1978).
[16]V. A. Fateev, I. V. Frolov, and A. S. Schwartz (Shvarts), Nucl. Phys. **B154**, 121 (1979); Yad. Fiz. **30**, 1134 (1979) [Sov. J. Nucl. Phys. **30**, 590 (1979)].
[17]R. Crewther, Phys. Rev. Lett. **28**, 1421 (1972); M. Chanovitz and J. Ellis, Phys. Lett. **B40**, 397 (1972); J. Collins, L. Dunkan, and S. Joglekar, Phys. Rev. D **16**, 438 (1977).
[18]M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. **B163**, 46 (1980); **165**, 45 (1980).
[19]M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. **B147**, 385 448 (1979).

Translated by Julian B. Barbour