

## Chapter III

### ABC of Instantons

V. A. Novikov,<sup>a</sup> M. A. Shifman,<sup>b</sup> A. I. Vainshtein<sup>b</sup> and  
V. I. Zakharov<sup>c</sup>

<sup>a</sup>*Institute of Theoretical and Experimental Physics, Moscow 117259, Russia*

<sup>b</sup>*Theoretical Physics Institute, University of Minnesota, Minneapolis,  
MN 55455, USA*

<sup>c</sup>*Max-Planck Institute für Physik, 80805 München, Germany*

This lecture was first delivered at XVI Winter School of Physics of Leningrad Institute for Nuclear Physics in spring 1981. Revised April 1994. First published in Russian in *Usp. Phys. Nauk* **136** (1982) 553.

### Abstract

These lectures present an extended introduction to instantons in gauge theories. The lectures consist of several distinct parts. To reveal the physical meaning of instantons we consider in detail the simplest quantum-mechanical problem where they appear: tunneling in the double-well potential. This pedagogical example was suggested by Polyakov. Then we proceed to quantum chromodynamics (QCD). The discovery of instantons was instrumental in the understanding of the vacuum structure of QCD. The  $\theta$  vacuum is described from the quasiclassical perspective. The second part is devoted to the instanton formalism. We discuss various aspects of the instanton calculations: the solution *per se* in different gauges, the instanton measure in QCD and in the Higgs phase, the impact of external background fields. A related topic we dwell on is the sphaleron and its interpretation. Finally, the last part deals with the massless fermions in the instanton transitions. Their impact is drastic both at the conceptual and technical levels. We explain how the tunneling interpretation changes in the presence of the massless fermions. If the fermions are chiral rather than Dirac, under certain conditions the theory becomes ill-defined (Witten's global anomaly). Although these lectures are self-contained, they are best read in conjunction with Coleman's lecture *The Uses of Instantons* [S. Coleman, *Aspects of Symmetry* (Cambridge University Press, London, 1985), p. 265].

### Contents

Introduction	204
1. Quantum Mechanics, Imaginary Time, Path Integrals	206
2. Double-Well (Two-Hump) Potential. Tunneling	213
3. Determinant and Zero Modes	217
4. Instanton Gas	223
5. Tunneling in Quantum Chromodynamics	226
5.1. Nontrivial Topology in the Space of Fields in the Yang-Mills Theory	226
5.2. $\theta$ Vacuum	233
6. Euclidean Formulation of QCD	236

7. BPST Instantons. General Properties	238
7.1. Finiteness of the Action and the Topological Charge	238
7.2. The Distinguished Role of the Group SU(2)	241
7.3. Value of the Action for Instanton Solution	243
8. Explicit Form of the BPST Instanton	244
8.1. Solution with $Q = 1$	244
8.2. Singular Gauge. The 't Hooft <i>Ansatz</i>	246
8.3. Relations for the $\eta$ Symbols	247
8.4. Instanton in the $A_0 = 0$ Gauge	248
8.5. Instanton Collective Coordinates	249
8.6. Instantons in the Higgs Regime	253
8.7. The Height of the Barrier. Sphaleron	258
9. Calculation of the Pre-Exponential Factor for the BPST Instanton	262
9.1. Expansion Near a Saddle Point. Choice of the Gauge and Regularization	262
9.2. Zero Modes	263
9.3. Nonzero Modes. Effective Coupling Constant	264
9.4. Two-Loop Approximation	266
9.5. Density of Instantons (the Instanton Measure) in the Group SU( $N$ )	268
10. Instantons in the QCD Vacuum	270
10.1. Instantons in the Slowly Varying Background Fields	270
10.2. Instanton–Antiinstanton Interactions	274
11. Constrained Instantons and the Valley Method	278
12. Fermions in Instanton Field	282
12.1. Very Heavy Quarks	282
12.2. Light (Massless) Quarks	283
12.3. Tunneling Interpretation in the Presence of Massless Fermions. The Index Theorem	285
12.4. Instanton Density in the Theory with Light Quarks	288
12.5. Global Anomaly	292
13. Continuation in Euclidean Space in the Theories with Chiral Fermions and Supersymmetric Theories	292
References	296

## Introduction

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 octet of massless gluons. The idea of gauge invariance, however, is much older and derives from quantum electrodynamics, which was historically the first field-theoretical 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 in the mid-1970's 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. The latter, in turn, are due to 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 analytical “wave function” of the vacuum is a very difficult problem. Despite numerous attacks by theoreticians it still remains unsolved. Nevertheless, quite a lot is already known. The study of “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 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 [1]

$$\langle 0 | G_{\mu\nu}^a G_{\mu\nu}^a | 0 \rangle,$$

where  $G_{\mu\nu}^a$  is the gluon field strength tensor ( $a = 1, \dots, 8$  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 the “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 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 possible the reexamination of a number of problems.

In 1975 one of the most beautiful phenomena in quantum chromodynamics was discovered, instantons, classical solutions of the field equations with nontrivial topology [2]. 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 undoubted. Therefore, 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.

The original Belavin-Polyakov-Schwarz-Tyupkin solution [3] (BPST instanton) may or may not be the fluctuation which is dominant in the vacuum wave function. Although there are some numerical evidence in favor of the instanton dominance [4] the arguments are far from being conclusive. The instanton-based models of the QCD vacuum do exist, but the last word in this line of research is yet to be said. Therefore, we will not dwell on this issue. Instead, we will focus on those aspects of the instanton calculus which are completely settled and will stay with us forever.

We begin with a simple quantum-mechanical problem that illustrates the role of nonperturbative fluctuations. This example was analyzed in detail by Polyakov [5], who made a major contribution to the development of the entire subject.

A double-well potential will be considered, and the famous problem of the level splittings will be solved by exploiting an instanton approach, which is rather awkward in this particular problem, but has an important advantage over the standard WKB method: it can be directly extended to field theory, while the standard method cannot. All technical elements of the instanton calculus (the Euclidean time, classical solutions, zero modes and determinants) which we will encounter later in QCD are introduced in this setting. Having dealt with the toy model we proceed to QCD. General arguments are presented revealing a nontrivial topology in the space of the gauge fields. The existence of distinct classical minima of the “potential” is demonstrated. Classical trajectories interpolating between these distinct minima (“pre-vacua”) are BPST instantons. We discuss the explicit form of the instanton, and calculate,

in a pedagogical manner, the instanton density. The notion of the vacuum angle  $\theta$  is introduced.

We then briefly consider an applied aspect of the instanton calculus. Instantons submerged in background fields, produced by other fluctuations, deform. As a result of this deformation the instanton density changes. The change in the density caused by the gluon condensate is considered in some detail.

The second part of the lecture is devoted to the role of fermions. Massless fermions have a drastic impact both on interpretation of the instanton as a tunneling trajectory, and on all technical aspects of the instanton calculus. We first consider the Dirac fermions and explain how the instanton calculations must be modified. Then a more subtle problem of chiral fermions is addressed. Here we have to reanalyze anew the very foundations of the procedure, such as the Euclidean continuation. The chiral fermions are an indispensable element of supersymmetric gauge theories. A brief excursion in the topic of supersymmetric instantons concludes the lecture.

## 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 [1] are “inconvenient” for finding the eigenvalues of the 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

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

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 [6]. 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

$$e^{iS}.$$

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

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

Thus, the transition amplitude is

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

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 pass from states with a definite coordinate to states with a definite energy,

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

then, obviously,

$$\langle x_f | e^{-iHt_0} | x_i \rangle = \sum_n e^{-iE_n t_0} \langle x_f | n \rangle \langle n | x_i \rangle, \quad (4)$$

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  $t \rightarrow -i\tau$ . Then in the limit  $\tau_0 \rightarrow \infty$  only a single term survives in the sum (Eq. (4)), and this directly tells us what are the energy  $E_0$  and the wave function  $\psi_0(x)$  of the lowest level,  $e^{-E_0 \tau_0} \psi_0(x_f) \psi_0^*(x_i)$ .

In the literature, the transition to the imaginary time is frequently called the Wick rotation, and the corresponding version of the theory is referred to as the Euclidean version. Below, we shall see that the substitution  $t \rightarrow -i\tau$  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

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

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

We call

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

the Euclidean action. Since  $S_E \geq 0$ , we have acquired an exponentially decreasing weight on the right-hand side of Eq. (3). In the present lecture, we shall remain in the Euclidean space and shall not return to the Minkowski space (i.e. to real time) until Sec. 13; therefore, in what follows we shall omit the subscript  $E$ .

The Euclidean variant of Eq. (3) is

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

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

$$x(\tau) = X(\tau) + \sum_n c_n x_n(\tau), \quad (8)$$

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

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

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 in other words, the method of steepest descent (the latter, “mathematical” term may be more readily understood

by some 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

$$N \int [Dx] e^{-S} \sim e^{-S_0}. \quad (10)$$

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 expresses, 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

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

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*  $V(x)$ .<sup>1</sup>

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 “beam” of paths near the extremal path, i.e. by the paths with action that differs little from  $S_0$ . In other words, we take into account only the quadratic deviation:

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

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

<sup>1</sup>The minus sign is due to the fact that the Euclidean formulation is considered [see Ref. [6]].

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

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

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

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

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

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

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  $\varepsilon_n$  are positive. In most cases, this is so, but in the instanton example 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 yet to be fixed. We shall not 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  $m\omega^2 x^2/2$ ),

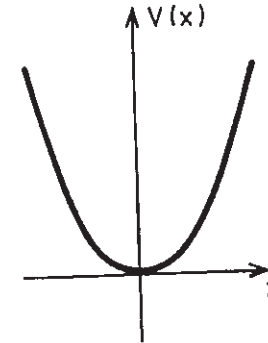


Fig. 1. The quantum-mechanical problem with the potential of the oscillator type.

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  $V(x)$ . 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  $\infty$ . It is also clear that the action on the path  $X(\tau) = 0$  vanishes.

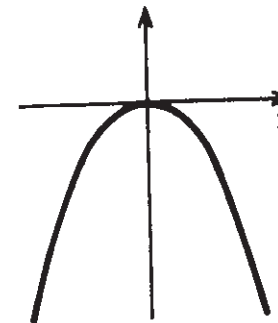


Fig. 2. The potential appearing in the same problem in the Euclidean time.

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

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

The eigenfunctions of the operator in the square brackets are the sine and cosine functions. For instance, the lowest eigenfunction is  $\cos(\pi\tau/\tau_0)$ , the next is  $\sin(2\pi\tau/\tau_0)$ , and so on. All the eigenvalues  $\varepsilon_n$  are immediately fixed by the boundary conditions  $x_n(\pm\tau_0/2) = 0$ :

$$\varepsilon_n = \frac{\pi^2 n^2}{\tau_0^2} + \omega^2, \quad n = 1, 2, \dots$$

We have now arrived at the point at which it is impossible to advance further without saying what is the value of  $N$ . To avoid the necessity of explicit determination of  $N$  we split the determinant into two factors:

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

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

$$\begin{aligned} N \left( \prod_{n=1}^{\infty} \frac{\pi^2 n^2}{\tau_0^2} \right)^{-1/2} &= \langle x_f = 0 | e^{-\hat{p}^2 \tau_0/2} | x_i = 0 \rangle = \sum_n |\langle p_n | x = 0 \rangle|^2 e^{-p_n^2 \tau_0/2} \\ &= \int_{-\infty}^{+\infty} \frac{dp}{2\pi} e^{-p^2 \tau_0/2} = \frac{1}{\sqrt{2\pi\tau_0}}. \end{aligned} \quad (17)$$

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 factor in Eq. (16). For the infinite product which occurs in it we have the well-known formula [see e.g. formula (1.431.2) in Ref. [7]]

$$\pi y \prod_{n=1}^{\infty} \left( 1 + \frac{y^2}{n^2} \right) = \sinh \pi y,$$

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

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

$$\begin{aligned} \langle x_f = 0 | e^{-H\tau_0} | x_i = 0 \rangle &= N \left[ \det \left( -\frac{d^2}{d\tau^2} + \omega^2 \right) \right]^{-1/2} \\ &= \frac{1}{\sqrt{2\pi\tau_0}} \left( \frac{\sinh \omega\tau_0}{\omega\tau_0} \right)^{-1/2} \\ &= \left( \frac{\omega}{\pi} \right)^{1/2} (2 \sinh \omega\tau_0)^{-1/2}. \end{aligned} \quad (18)$$

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

$$\langle x_f = 0 | e^{-H\tau_0} | x_i = 0 \rangle \xrightarrow{\tau_0 \rightarrow \infty} \left( \frac{\omega}{\pi} \right)^{1/2} e^{-\omega\tau_0/2} \left( 1 + \frac{1}{2} e^{-2\omega\tau_0} + \dots \right),$$

from which it follows that for the lowest state  $E_0 = \omega/2$  and  $[\psi_0(0)]^2 = (\omega/\pi)^{1/2}$ . The next term in the expansion corresponds to the level of the harmonic 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. Double-Well (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 oscillations — 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 oscillations; 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 QCD 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.

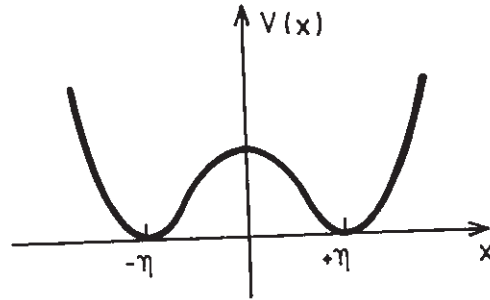


Fig. 3. The double-well potential.

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

$$V(x) = \lambda(x^2 - \eta^2)^2 \quad (19)$$

is shown in Fig. 3. We fix the parameters  $\lambda$  and  $\eta$  in such a way that

$$8\lambda\eta^2 = \omega^2,$$

where  $\omega$  is the frequency introduced in the previous section. Then near each 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 lowest 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 oscillations 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  $x \rightarrow -x$  is broken,  $E_0 = (\omega/2)(1 + \text{corrections})$  in both cases, and at small  $\lambda$  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:

$$E_0 = \frac{\omega}{2} - \sqrt{\frac{2\omega^3}{\pi\lambda}} e^{-\omega^3/12\lambda} \frac{\omega}{2}, \quad (20)$$

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

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  $\lambda$ . [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 equations of motion. Quantum-mechanical tunneling transfers the wave function from one well to the other and, in Polyakov’s terminology, “smears” the ground states. The correct wave function of the ground state is an even superposition of the wave functions in each well.

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 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 topologically nontrivial 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 one minimum to the other occurs below the barrier and is classically forbidden.

The solutions arise only after the Euclidean rotation. The double-well potential becomes a two-humped potential of Fig. 4.



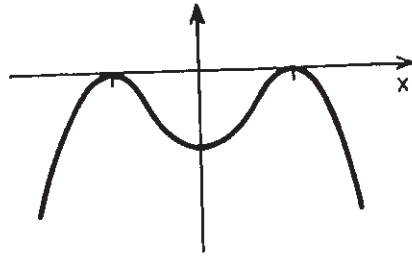


Fig. 4. The two-hump potential relevant to the motion in the Euclidean time.

We consider the calculation of the amplitudes

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

The first step 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  $\tau_0 \rightarrow \infty$ , since such solutions are important in the quasiclassical approximation that is under discussion. Most paths correspond to trajectories on which  $x \rightarrow \infty$  as  $\tau \rightarrow \infty$ , and they have infinite action.

A finite action in the limit  $\tau_0 \rightarrow \infty$  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  $\tau_0 \rightarrow \infty$  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  $\eta$  at the time  $\tau_0/2$ .<sup>3</sup> Physical intuition suggests that such trajectories exist, though their explicit form for finite  $\tau_0$  is complicated. We are always interested in only the lowest state, and therefore we can directly assume that  $\tau_0 \rightarrow \infty$ . In this limit, the solution is very simple:

$$X(\tau) = \eta \tanh \frac{\omega(\tau - \tau_c)}{2} \quad (21)$$

(it corresponds to mechanical motion with zero energy,  $E = (1/2)\dot{x}^2 - V(x) = 0$ , so that the equation can be rewritten as a first-order equation

$$\dot{X} = -\sqrt{2\lambda}(X^2 - \eta^2)$$

and readily integrated).

<sup>3</sup>Here we have allowed a slight inaccuracy. If  $\tau_0$  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  $\tau_0 \rightarrow \infty$  that the end points coincide with  $\pm\eta$ .

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

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  $\frac{1}{2}\dot{x}^2 = V(x)$ ):

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

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.

We draw attention to an additional property of the 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 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 been found already, 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 [8]. However, we shall proceed in a “brute force” manner, which is the closest approximation to the method used by 't Hooft [9] to calculate the instanton determinant in QCD. We hope that this will subsequently enable the reader to reproduce for himself