Chapter III

ABC of Instantons

V. A. Novikov,[a] M. A. Shifman,[b] A. I. Vainshtein[b] and V. I. Zakharov[c]

[a] *Institute of Theoretical and Experimental Physics, Moscow 117259, Russia*
[b] *Theoretical Physics Institute, University of Minnesota, Minneapolis, MN 55455, USA*
[c] *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 θ 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. θ 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 *Ansatz* 246
   - 8.3. Relations for the η 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 quantitites 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]

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

where $G^a_{\mu\nu}$ is the gluon field strength tensor ($a = 1, \ldots, 8$ 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 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 θ 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 indispensible element of supersymmetric guage 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

$$\mathcal{L} = \frac{1}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 - V(x)\,. \tag{1}$$

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 [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

$$S = \int_{-t_0/2}^{t_0/2} \mathrm{d}t\, \mathcal{L}(x, \dot{x})\,. \tag{2}$$

Thus, the transition amplitude is

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

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

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

then, obviously,

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

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 (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 $t \to -i\tau$ 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

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

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

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

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

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

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:

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

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 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

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

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 expresses, in mathematical language, the fact that $X(\tau)$ realizes a minimum of the action:

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

where $V' = \mathrm{d}V/\mathrm{d}x$. The equation

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

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]

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:

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

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

[1] 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

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

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

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

(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

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

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 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 $m\omega^2x^2/2$),

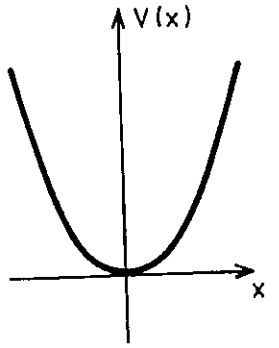


Fig. 1. The quantum-mechanical problem with the potential of the oscillator type.

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

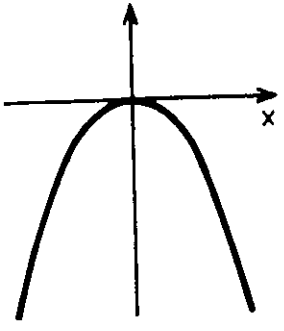


Fig. 2. The potential appearing in the same problem in the Euclidean time.

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

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

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

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

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:

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

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

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

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 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:

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

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

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

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.

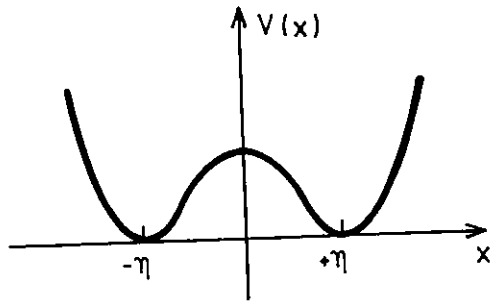


Fig. 3. The double-well potential.

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

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

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 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 $x \to -x$ 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 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.

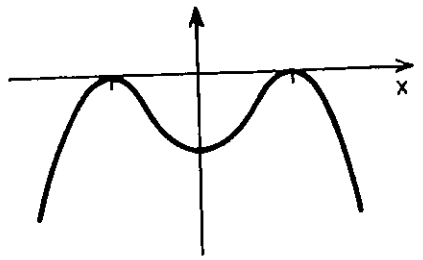


Fig. 4. The two-hump potential relevant to the motion in the Euclidean time.

We consider the calculation of the amplitudes

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

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 $\tau_0 \to \infty$, since such solutions are important in the quasiclassical approximation that is under discussion. Most paths correspond to trajectories on which $x \to \infty$ 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:

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

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

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

and readily integrated).

[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 $\pm\eta$.

Such a solution is called *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 antiinstantons, 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 $\frac{1}{2}\dot{x}^2 = V(x)$):

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

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.

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

(herself) all details of 't Hooft's work, which is central for the entire instanton problem.

The original formula (14) is conveniently rewritten as

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

We have multiplied and divided by a known number — the determinant for the harmonic oscillator see Eq. (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

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

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 ([10], pp. 73 and 80), 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

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

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.

[4]With our 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.

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 generic perturbation of $X(\tau)$ increases the action. Accordingly, ε_n is positive. However, we already know that there is one direction in the functional 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 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 mode has the form

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

or, which is the same as

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

The fact that the normalization factor reduces to $S_0^{-1/2}$ 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 mode (see Eqs. (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 $\mathrm{d}c_0$ is the same as the integration over $\mathrm{d}\tau_c$, 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 Eq. (8)). On the other hand, the change $\Delta x(\tau)$ under a shift $\Delta\tau_c$ of the center is

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

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 ensure that as c_0 varies from $-\infty$ to $+\infty$ the parameter τ_c changes in the same interval.) This is not yet the end of the story, 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 mode removed.

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

We now consider nonzero 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 need 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 interval $(0, \infty)$. 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

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

we have in the other asymptotic region the same exponential:

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

The second exponential, $e^{-ip\tau}$, which should in principle arise, is absent, and the entire dynamical effect is 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 [10] on p. 81). 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 $x(\pm\tau_0/2) = 0$. The equations for A and B,

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

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 as,

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

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

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

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

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

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:

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

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 antiinstantons, and this will be done in the following section.

It is appropriate here to make some comments. The factor in the first 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 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 mode. This circumstance is also of a general nature. In quantum chromodynamics too, each zero mode is associated with $\sqrt{S_0}$. Finally, the existence of the zero 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 is 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 results can be reconstructed almost at once, without calculations. We have given so much attention to 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 lowest 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 antiinstantons are important. If the distance between their centers is large, such a path is also a classical solution.

[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.

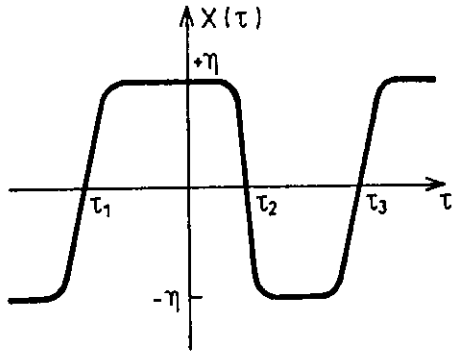


Fig. 5. The chain of n well-separated instantons (antiinstantons).

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

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

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:

$$\sqrt{\frac{\omega}{\pi}}e^{-\omega\tau_0/2} \to \sqrt{\frac{\omega}{\pi}}e^{-\omega\tau_0/2}\left(\sqrt{\frac{6}{\pi}}\sqrt{S_0}e^{-S_0}\right)^n \prod_1^n (\omega\,\mathrm{d}\tau_i)\,.$$

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

$$\sqrt{\frac{\omega}{\pi}}e^{-\omega\tau_0/2}d^n \int_{-\tau_0/2}^{\tau_0/2} \omega\mathrm{d}\tau_n \int_{-\tau_0/2}^{\tau_n} \omega\mathrm{d}\tau_{n-1}\cdots\int_{-\tau_0/2}^{\tau_2} \omega\mathrm{d}\tau_1$$

$$= \sqrt{\frac{\omega}{\pi}}e^{-\omega\tau_0/2}d^n\frac{(\omega\tau_0)^n}{n!}\,,$$

where d denotes the instanton density,

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

The amplitudes $\langle -n|e^{-H\tau_0}|\eta\rangle$ 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:

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

$$= \sqrt{\frac{\omega}{\pi}}e^{-\omega\tau_0/2}\sinh(\omega\tau_0 d)\,, \tag{36}$$

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

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 wells 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 antiinstantons 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. This has been exploited particularly by Callan, Dashen and Gross [11] in quantum chromodynamics. Unfortunately, in QCD we do not have free parameters like λ that can be kept small. Therefore, a dilute instanton gas is not suitable from the quantitative point of view in QCD, and the most we can extract from it are

heuristic indications. Further details regarding attempts to improve the instanton gas approximation (the so-called, instanton liquid) can be found in Ref. [4].

To conclude the section, we note that a somewhat more extensive exposition of the instanton approach to the double-well, potential problem is contained in Coleman's lecture [8]. The reader interested in special questions, for example, situations not covered by the gas approximation in quantum mechanics, must consult Ref. [12].

5. Tunneling in Quantum Chromodynamics

5.1. *Nontrivial Topology in the Space of Fields in the Yang-Mills Theories*

Now, when we are done with the toy model, we can pass to real QCD. The Lagrangian of the theory has the form

$$L = -\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} + \sum \bar{\psi}(iD_\mu\gamma^\mu - M)\psi$$

where the sum runs over all quark flavors, $G_{\mu\nu}$ is the gluon field strength tensor,

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu ,$$

g is the gauge coupling constant, f^{abc} stands for the structure constants of the gauge group. In QCD the gauge group is SU(3); the quarks are described by the Dirac fields ψ^i transforming according to the fundamental (triplet) representation of SU(3). The issue to be discussed in this section is independent of the particular choice of the gauge group and the presence (absence) of the quark fields. To make the picture as transparent as possible we will disregard, for the time being, the quarks, and consider the simplest non-Abelian group, SU(2). Of course, later on we will include quarks and pass from SU(2) to SU(3).

If tunneling is important for understanding the structure of QCD, the first question to be asked is from where to where does the system tunnel.

At first glance it is not obvious at all that the Lagrangian of gluodynamics has a discrete set of degenerate classical minima. In the double-well potential problem this was evident. The main distinction is due to the fact that in the field theory the number of degrees of freedom is infinite, while in the toy

example of the previous sections we dealt with only one degree of freedom. The space of fields in QCD is infinitely-dimensional. Most of these field-theoretical degrees of freedom are oscillator-like and, thus, "uninteresting." Our task is to single out such degree(s) of freedom which tunnel and are delocalized in the space of fields, much in the same way as the genuine ground state wave function in the double-well potential is not localized in any of the two minima, but is rather smeared along the x-axis.

Below, we will demonstrate that in QCD there exists one such direction in the infinitely-dimensional space of fields. If we forget for a while about all other degrees of freedom, and focus on this chosen degree of freedom, we will see that the corresponding dynamics, being somewhat more complicated than that of the double-well potential system, is similar in one aspect: it calls for the consideration of tunneling. The closest analogy one can keep in mind in this context is quantum mechanics of a particle living on a vertically oriented circle subjected to a constant gravitational force (Fig. 6). Classically the particle with the lowest possible energy (the ground state of the system) just stays at rest at the bottom of the circle. Quantum-mechanically, the zero-point oscillations come into play. Within the perturbative treatment, we deal exclusively with small oscillations near the equilibrium point at the bottom of the circle. For such small oscillations, the existence of the upper part of the circle plays no role. It could have been eliminated altogether with no impact on the zero-point oscillations.

From the courses of quantum mechanics it is known, however, that the genuine ground-state wave function is different. The particle oscillating near the

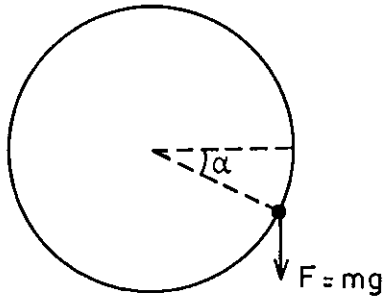


Fig. 6. Quantum mechanics of a particle (•) on a one-dimensional topologically nontrivial manifold, circle.

origin "feels" that it can wind around the circle to which it belongs, tunneling under the potential barrier it experiences at the top of the circle.

To single out the relevant degree of freedom in the infinitely-dimensional space of the gluon fields, it is necessary to proceed to the Hamiltonian formulation of the Yang-Mills theory which implies, of course, that the time component of the four-potential has to be gauged away, $A_0 = 0$. Then,

$$\mathcal{H} = \frac{1}{2}\int d^3x\{\mathbf{E}^a\mathbf{E}^a + \mathbf{B}^a\mathbf{B}^a\}$$

where $\mathcal{H}$ is the Hamiltonian and $E_i^a = \dot{A}_i^a$ are to be treated as canonical momenta.

Two subtle points are to be mentioned in connection with this Hamiltonian. First, the equation $\operatorname{div}\mathbf{E} = \rho$, inherent to the original Yang-Mills theory, does not stem from this Hamiltonian *per se*. This equation must be imposed as a constraint on the states from the Hilbert space by hand. Second, the gauge freedom is not fully eliminated. Gauge transformations which depend on $\mathbf{x}$ but not t are still allowed. This freedom is reflected in the fact that, instead of two transverse degrees of freedom $\mathbf{A}_\perp$, the Hamiltonian above has three (three components of $\mathbf{A}$). Imposing, say, the Coulomb gauge condition,

$$\partial_i A_i = 0$$

we could get rid of the "superfluous" degree of freedom, a procedure quite standard in perturbation theory (in the Coulomb gauge). Alas! If we want to keep and reveal the topologically nontrivial structure of the space of fields, the Coulomb gauge condition can *not* be imposed. We have to work, with certain care, with the "undergauged" Hamiltonian.

Quasiclassically the state of the system described by this Hamiltonian at any given moment of time is characterized by the field configuration $A_i^a(\mathbf{x})$. Since we are interested in the zero-energy states (classically, this is obviously the minimal energy) A_i^a must be pure gauge,

$$A_i^a(\mathbf{x})|_{\text{vac}} = \frac{i}{g}U(\mathbf{x})\partial_i U^{-1}(\mathbf{x})$$

where U is a matrix belonging to SU(2) and depending on the *spatial* components of the four-coordinate.

Moreover, we are interested only in those zero-energy states which might be connected with each other by tunneling transitions, i.e. the corresponding

classical action must be finite. The latter requirement results in the following boundary condition (see e.g. Ref. [11] for a detailed discussion)

$$U(|\mathbf{x}| \to \infty) = 1\,,$$

or any other constant matrix U_0 independent of the direction in the three-dimensional space along which $\mathbf{x}$ tends to infinity. This boundary condition compactifies our *three-dimensional* space which becomes topologically equivalent to three-dimensional sphere.

On the other hand, the group space of SU(2) is also a three-dimensional sphere. Indeed, any matrix belonging to SU(2) can be parametrized as

$$M = A + i\mathbf{B}\boldsymbol{\sigma}, \qquad M \in \mathrm{SU}(2)\,.$$

Here A and $\mathbf{B}$ are four real parameters; $\boldsymbol{\sigma}$ are the Pauli matrices. Both conditions, $M^+M = 1$ and $\det M = 1$, are met provided

$$A^2 + \mathbf{B}^2 = 1\,.$$

Since $U(\mathbf{x})$ is a matrix from SU(2), and the space of all coordinates $\mathbf{x}$ is topologically equivalent to a sphere (after compactification $U(|\mathbf{x}| \to \infty) = 1$), the function $U(\mathbf{x})$ realizes a mapping of the sphere in the coordinate space onto a sphere in the group space. Intuitively it is quite obvious that all continuous mappings $S_3 \to S_3$ are classified according to the number of coverings. This number will be the number of times we sweep the group sphere S_3 when the coordinate $\mathbf{x}$ sweeps the sphere in the coordinate space once. The number of coverings can be zero (topologically trivial mapping), one, two, and so on. The number of coverings can be negative too, since the mappings $S_3 \to S_3$ are orientable [13]. This fact is especially transparent for the mappings $S_1 \to S_1$, i.e. circle onto circle. If one circle is swept in the clockwise direction and the other one in the anticlockwise, we say that the number of coverings is -1.

In other words, the matrices $U(\mathbf{x})$ can be sorted out in distinct classes labeled by an integer number, $U_n(\mathbf{x})$, $n = 0, \pm1, \pm2, \ldots$, which is referred to as the *winding number*. All matrices belonging to a given class $U_n(\mathbf{x})$ are reducible to each other by a continuous $\mathbf{x}$-dependent gauge transformation. At the same time, no continuous gauge transformation can transform $U_n(\mathbf{x})$ into $U_{n'}(\mathbf{x})$ if $n \neq n'$. The unit matrix represents the class $U_0(\mathbf{x})$. For $n = 1$ one can take, for instance,[6]

[6] Let us note in passing that exactly the same topological classification is the basis of the theory of Skyrmions, see Ref. [14] for a review.

$$U_1(\mathbf{x}) = \exp\left[i\pi \frac{\mathbf{x}\sigma}{(\mathbf{x}^2+\rho^2)^{1/2}}\right],$$

where ρ is an arbitrary parameter. An example of the matrix from U_n is U_1^n.

Any field configuration $A_i^a(\mathbf{x})|_{\text{vac}} = (i/g)U_n(\mathbf{x})\partial_i U_n^{-1}(\mathbf{x})$, being pure gauge, corresponds to the lowest possible energy — the zero energy. As a matter of fact, the set of points $\{U_n\}$ in the space of fields, obviously consists simply of the gauge images of one and the same physical point (analogous to the bottom of the circle in Fig. 6). The fact that the matrices U_n from different classes are not continuously transformable to each other indicates the existence of a "hole" in the space of fields, with a noncontractible loop winding around this "hole".

We are finally ready to identify the degree of freedom corresponding to the motion along this circle. Let us consider the vector K_μ,

$$K_\mu = 2\varepsilon_{\mu\nu\alpha\beta}\left(A_\nu^a\partial_\alpha A_\beta^a + \frac{g}{3}f^{abc}A_\nu^a A_\alpha^b A_\beta^c\right).$$

The vector K_μ is called the *Chern-Simons current*; it plays an important role in the instanton calculus. We will encounter it more than once in what follows. Now, define the charge $\mathcal{K}$ corresponding to the Chern-Simons current,

$$\mathcal{K} = \frac{g^2}{32\pi^2}\int K_0(x)d^3x\,.$$

It is not difficult to show that for any pure gauge field $A_i^a(\mathbf{x})$ the Chern-Simons charge $\mathcal{K}$ measures the winding number. For any field $A_i(\mathbf{x}) = (i/g)U_n(\mathbf{x})\partial_i U_n^{-1}(\mathbf{x})$ we have[7]

$$\mathcal{K} = n\,.$$

Summarizing, moving in the "direction of $\mathcal{K}$" in the space of fields we observe that this particular direction has the topology of circle. The points $\mathcal{K}$ and $\mathcal{K}+1$, and $\mathcal{K}-1,\ldots$ are physically one and the same point. The integer values of $\mathcal{K}$ correspond to the bottom of the circle in Fig. 6.

[7] *Exercise:* Prove the assertion formulated in the above paragraph. In case of difficulties consult Coleman's lecture [8].

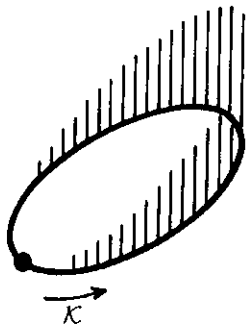


Fig. 7. Nontrivial topology in the space of gauge fields in the $\mathcal{K}$ direction. The length of the circle is 1. The vertical lines indicate the strength of a potential acting on the effective degree of freedom living on the circle.

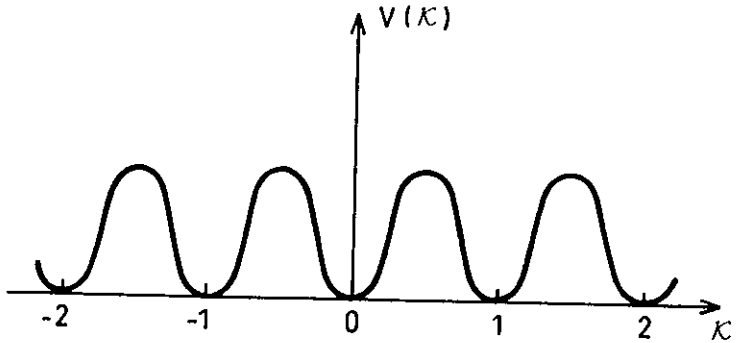


Fig. 8. If we unwind the circle of Fig. 7 onto a line we get a periodic potential.

It is convenient to visualize dynamics of the Yang-Mills system in the "direction of $\mathcal{K}$" as shown in Fig. 7. The vertical lines indicate the potential energy — the higher the line the larger the energy. It is well-known (see e.g. Ref. [15]) that the only consistent way of treating the quantum-mechanical systems living on a circle (angle-type degrees of freedom) is to cut the circle and map it many times onto a straight line. In other words, we pretend that the variable $\mathcal{K}$ lives on the line (Fig. 8). Any integer value of $\mathcal{K}$ in Fig. 8 corresponds to a pure gauge configuration with zero energy. On the other hand, if $\mathcal{K} \neq n$ the field strength tensor is nonvanishing and the energy of the field configuration is positive. Viewed as a function on the line, the potential energy $V(\mathcal{K})$ is, of

course, periodic — with the unit period. To take into account the fact that the original problem is formulated on the circle, we impose the (quasi)periodic Bloch boundary condition on the wave function Ψ,

$$\Psi(\mathcal{K}+1) = e^{i\theta}\Psi(\mathcal{K})\,.$$

The phase θ appearing in the Bloch quasiperiodic boundary condition is a hidden parameter of QCD, the *vacuum angle*. We will return to the issue of the vacuum angle later on. The classical minima of the potential of Fig. 8 can be called *pre-vacua*. The correct wave-function of the quantum-mechanical vacuum state of the Bloch form is built as a linear combination of these pre-vacua.

We would like to emphasize here a subtle point which in many presentations remains fogged. It might seem that the systems depicted in Figs. 7 and 8 (a particle on a circle and that in the periodic potential) are physically identical. This is not quite the case. In the periodic potentials, say in crystals, one can always introduce impurities that would slightly violate periodicity. For a system on a circle this cannot be done. The correct analog system for gluodynamics, where the gauge invariance is a sacred principle, is that of Fig. 7.

Assume that at $t = -\infty$ and at $t = +\infty$ our system is at the classical minimum (zero-energy state). Assume also that at $t = -\infty$ the winding number $\mathcal{K} = n$ while at $t = +\infty$ $\mathcal{K} = n \pm 1$. In Fig. 7 this means that our system tunnels from the point marked by the closed circle to the very same point under the hump of the potential energy.

Consider now a field configuration $A_\mu(t, \mathbf{x})$ continuously interpolating (with the minimal action) between these two states in the Euclidean time, the least-action tunneling trajectory. This is the BPST instanton.

The fact that the BPST instantons describe tunnelings in the space of fields, which possesses noncontractible paths, was realized by V. Gribov shortly after the discovery of the BPST instantons.[8] Almost simultaneously the tunneling picture was revealed in Refs. [16] and [17]. A very pedagogical and illustrative discussion of the tunneling interpretation given in the *Minkowski* space is presented in Ref. [18] which may be recommended to the reader just beginning to study instanton calculus.

[8]See e.g. a remark in Ref. [5] where Polyakov acknowledges Gribov's suggestion of the tunneling interpretation.

The analysis outlined above (the one based on the Hamiltonian formulation) is convenient for establishing the existence of a nontrivial topology, nonequivalent vacuum states and, hence, the existence of nontrivial interpolating field configurations corresponding to tunneling. As we already know, the minimal action is achieved for the solutions of the classical equations of motion (in Euclidean time) with the given boundary conditions. In practice, however, the Hamiltonian gauge $A_0 = 0$ is rarely used in constructing these solutions. This gauge is extremely inconvenient for this purpose.

Below we will describe a standard procedure based on a specific *ansatz* for $A_\mu(x)$ in which all four Lorentz components of A_μ are nonvanishing. This *ansatz* entangles the color and Lorentz indices; the field configurations emerging in this way are, following Polyakov, generically referred to as "hedgehogs".

Since the solutions we are going to deal with are those of the Euclidean equations of motion, the first question to be asked is what needs to be done with QCD in order to pass to the Euclidean time. One can choose two alternative routes. In pure Yang-Mills theory with no fermions, it is advantageous, from the very beginning, to formulate a Euclidean version of the theory, and work only with this version. With the instanton calculations we will be never required to return to the original Minkowski version of the theory. The Euclidean formulation can be also developed in the presence of fermions, provided all fermions in the theory are described by the Dirac fields, i.e. are nonchiral. We will follow this route almost up to the very end of the lecture. The approach does not work, however, for chiral fermions, and in many supersymmetric field theories. For such problems one must choose the second route, which will be discussed in Sec. 13.

5.2. *θ Vacuum*

The existence of a noncontractable loop in the space of fields A_μ leads to drastic consequences for the vacuum structure in non-Abelian gauge theories. Let us take a closer look at the potential of Fig. 8. The argument presented below is formulated in quasiclassical language. One should keep in mind, however, that the general conclusion is valid even though the quasiclassical approximation is inappropriate in quantum chromodynamics where the coupling constant becomes large at large distances.

The lowest-energy state of the system depicted in Fig. 8, classically, is in one of the minima of the potential. Quantum-mechanically the zero point

oscillations arise. The wave function[9] corresponding to oscillations near the nth zero-energy state, Ψ_n, is localized near the corresponding minimum. The genuine wave function is delocalized, however,

$$\Psi_\theta = \sum_{n=0,\pm1,\pm2,\ldots} e^{in\theta}\Psi_n\,,$$

where θ is a parameter,

$$0 \le \theta \le 2\pi\,,$$

analogous to the quasimomentum in the physics of crystals [15]. The nth term in the sum can be called a *pre-vacuum* while the total sum represents the θ vacuum of QCD. The vacuum angle θ is a global fundamental constant characterizing the boundary condition on the wave function. It does not make sense to say that in one part of the space θ takes some value, while in another part θ takes a different value, or that θ depends on time. The worlds with different values of θ have orthogonal wave functions; for any operator $\mathcal{O}$ from the Hilbert space of the physical states

$$\langle\Psi_\theta|\mathcal{O}|\Psi_{\theta'}\rangle = 0 \qquad \text{if } \theta \neq \theta'\,.$$

This property is referred to as the *superselection rule*. The energy of Ψ_θ can (and does) depend on θ. From the definition of the vacuum angle it is clear that the θ dependence of all physical quantities, including the vacuum energy, must be periodic, with the period 2π.

Since all states Ψ_n are degenerate in energy the question is often raised on why one should form a linear combination corresponding to the θ vacuum. Is it possible to take, say, Ψ_0 as the vacuum wave function?

This question can be answered at different levels. Purely theoretically, if we want to implement the full gauge invariance of the theory, including the invariance under "large" gauge transformations, we must pass from Ψ_n to Ψ_θ. At a more pragmatic level one can say that introduction of Ψ_θ is necessary to maintain the property of the cluster decomposition, which must take place in any sensible field theory. (We are reminded that the cluster decomposition means that the vacuum expectation value of the T product of any two operators, $\mathcal{O}_1(x_1)$ and $\mathcal{O}_2(x_2)$, at large separations $|x_1 - x_2| \to \infty$ must tend to $\langle\mathcal{O}_1\rangle\langle\mathcal{O}_2\rangle$.) If the vacuum wave function is chosen to be Ψ_n, this property would not be valid, as we will see below. Finally, by proceeding to Ψ_θ we ensure that the vacuum state is stable under small perturbations. This is not the case if the vacuum wave function is Ψ_n. A small mass term of the quark fields would then cause a drastic restructuring of the vacuum wave function.

[9] In application to QCD we should rather use the term wave functional; nevertheless, we will continue referring to the wave function.

Although the physical meaning of the parameter θ is absolutely transparent within the Hamiltonian formulation, when we speak practically of instantons in field theory we keep in mind the Lagrangian formulation based on the path-integral formalism. In the Lagrangian formalism the vacuum angle is introduced as the θ *term* in the Lagrangian,

$$L = -\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} + L_\theta\,, \qquad L_\theta = \theta\frac{g^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^a_{\mu\nu}\,,$$

where

$$\tilde{G}^a_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}G^a_{\alpha\beta}, \qquad \varepsilon^{0123} = 1\,.$$

Note that if $\theta \neq 0$ or π, the θ term violates P and T invariance.

Before the discovery of the instantons it was believed that QCD naturally conserves P and CP. Indeed, the only gauge invariant Lorentz scalar operator one could construct from the A_μ fields of dimension 4 violating P and T is $G\tilde{G}$. This operator is a full divergency, $G\tilde{G} = \partial_\mu K_\mu$ where K_μ is the Chern-Simons current. It was believed that full divergencies have no impact on the action.

In the instanton field, however, the integral of $G\tilde{G}$ does not vanish. The reasons for that will be explained below. What is important for us now is the fact that adding the θ term to the QCD Lagrangian we do break P and CP in the strong interactions if $\theta \neq 0$. Since it is known experimentally, that P and CP symmetries are conserved in the strong interactions, to a very high degree of accuracy, this means that in nature the vacuum angle is fine-tuned, and is very close to zero. (As a matter of fact, estimates show that $\theta \le 10^{-9}$ [19].) Thus, with the advent of instantons the naturalness of QCD is gone. Can this fine-tuning be naturally explained? There exist several suggestions as to how one could solve the problem of P and CP conservation in QCD in a natural way. One of the most popular is the axion conjecture [20]. This topic, however, definitely lies outside the scope of the present lecture. Interested readers are referred to Ref. [21] for a pedagogical review. We will simply assume that $\theta = 0$, although theoretically it could take any value from the interval $[0, 2\pi]$.

6. Euclidean Formulation of QCD

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. In Minkowski space one distinguishes between the contravariant and covariant vectors, v^μ and v_μ, respectively. ($\mu = 0, 1, 2, 3$.) The spatial vector $\mathbf{v}$ coincides with the components of the contravariant four-vector, $\mathbf{v} = \{v^1, v^2, v^3\}$. In Euclidean space the distinction between the lower and upper vectorial indices is immaterial, we consider just one vector $\hat{v}_\mu (\mu = 1, 2, 3, 4)$. (In this section the caret is used to denote all quantities on Euclidean space.) In transition to Euclidean the spatial coordinates are not changed, $\hat{x}_i = x^i$, $i = 1, 2, 3$. 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:

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

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 is introduced for convenience in the expression of the following formulas.[10]

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

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

For the field strength tensor $G_{\mu\nu}$ we obtain the formulas

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

[10] 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 $g \to -g$.

where the Euclidean field strength tensor $\hat{G}^a_{\mu\nu}$,

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

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, what remain are the formulas for the Fermi fields. We begin with the definition of four Hermitian γ matrices $\hat{\gamma}_\mu$:

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

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 $\hat{\bar{\psi}}$ as

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

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 varies as follows:

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

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

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

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:

$$iS = -\hat{S}\,,$$

$$S = \int \mathrm{d}^4x\left[-\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu} + \bar{\psi}(i\gamma_\mu D_\mu - M)\psi + \theta\frac{g^2}{32\pi^2}G^a_{\mu\nu}\tilde{G}^a_{\mu\nu}\right], \quad (45)$$

$$\hat{S} = \int \mathrm{d}^4\hat{x}\left[\frac{1}{4}\hat{G}^a_{\mu\nu}\hat{G}^a_{\mu\nu} + \bar{\hat{\psi}}(-i\hat{\gamma}_\mu\hat{D}_\mu - iM)\hat{\psi} + i\theta\frac{g^2}{32\pi^2}\hat{G}^a_{\mu\nu}\tilde{\hat{G}}^a_{\mu\nu}\right],$$

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. Note that in the Euclidean space the Levi-Civita tensor $\varepsilon_{\mu\nu\alpha\beta}$ is defined in such a way that $\varepsilon_{1234} = 1$.

Below, we shall use the Euclidean space and omit the caret. The formulas given above 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 values of the time-ordered products of currents for space-like 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.

7. BPST Instantons General Properties

7.1. *Finiteness of the Action and the Topological Charge*

We have learned that in quantitative description of tunneling 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 quantum-mechanical example we have analyzed in Sec. 2, 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 deformed 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 antiinstanton realize minima for $Q = \pm 1$.

We now turn to gluodynamics — the theory of non-Abelian vector fields — 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^a_μ as $x \to \infty$ if the action is to be finite? (We have in mind the Euclidean action $\hat{S}$; see Eq. (45).) It is clear that the field strength tensor $G^a_{\mu\nu}$ must decrease more rapidly than $1/x^2$. But this by no means implies that the fields A^a_μ must decrease faster than $1/x$. Indeed, suppose A^a_μ in the limit $x \to \infty$ has the form

$$A_\mu = g\frac{\tau^a}{2}A^a_\mu \xrightarrow[x\to\infty]{} iS\partial_\mu S^+ \quad (46)$$

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 field strength tensor $G^a_{\mu\nu}$ vanishes, since A^a_μ has a purely gauge form.

Thus, the behavior of A^a_μ at large x is determined by the matrix S, which depends on the angles. Under a gauge transformation of A_μ defined[11] 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 $1/x$

[11] Warning: Although one and the same letter U is used here and in Sec. 5, these are different gauge matrices which depend on different variables: on the spatial coordinate $\mathbf{x}$ in Sec. 5, and on the Euclidean four-coordinate x_μ here.

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^a_μ 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. [3], 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

$$S_n = (S_1)^n, \qquad n = 0, \pm 1, \pm 2, \ldots . \tag{48}$$

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 topologically trivial gauge transformation.

For the careful reader it should be clear already that there exist two related, but not identical topological arguments. The first argument, discussed in detail in Sec. 5, reveals the existence of the distinct topologically inequivalent zero-energy states. Outlined here is a four-dimensional topological aspect; it refers to the topology of the trajectories connecting (in the Euclidean space-time) the distinct zero-energy states discussed in Sec. 5. The field configurations $A_\mu(x_4, \mathbf{x})$ satisfying Eq. (46) with $S = S_1$ interpolate between the state with the winding number $\mathcal{K}$ and that with the winding number $\mathcal{K}+1$. (To see that this is indeed the case we must, of course, transform the instanton into the $A_0 = 0$ gauge,see Sec. 8.4.) For $S = S_2$ we deal with the trajectory $A_\mu(x_4, \mathbf{x})$ connecting $\mathcal{K}$ and $\mathcal{K}+2$, etc. The topological charge Q of any given field configuration $A_\mu(x_4, \mathbf{x})$ satisfying Eq. (46) is actually nothing but

$$Q = \mathcal{K}' - \mathcal{K},$$

where the prime marks the distant (Euclidean) future while the unprimed quantity refers to the distant (Euclidean) past. A gauge-invariant integral

representation exist for the topological charge Q (making unnecessary the transition to the $A_0 = 0$ gauge):

$$Q = \frac{g^2}{32\pi^2} \int \mathrm{d}^4 x G^a_{\mu\nu} \tilde{G}^a_{\mu\nu}, \tag{49}$$

where

$$\tilde{G}^a_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\gamma\delta} G^a_{\gamma\delta}, \qquad \varepsilon_{1234} = 1. \tag{50}$$

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

$$G_{\mu\nu}\tilde{G}_{\mu\nu} = \partial_\mu K_\mu,$$

$$K_\mu = 2\varepsilon_{\mu\nu\gamma\delta}\left(A^a_\nu \partial_\gamma A^a_\delta + \frac{1}{3} g\varepsilon^{abc} A^a_\nu A^b_\gamma A^c_\delta\right),$$

so that the volume integral (49) can be transformed into an integral over the surface of a large sphere embedded in four-dimensional space, where A^a_μ has the form (46).

7.2. *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. To fix the gauge 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

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

with matrix U_2 that does not depend on x.

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

$$S(x) \to U_1 S, \qquad \tilde{A}_\mu(x) \to \tilde{A}_\mu(x). \tag{53}$$

Thus, the color 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

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

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.) η and $\bar{\eta}$ are called the 't Hooft symbols. 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

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

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

$$e^{i\varphi_1^a I_1^a + i\varphi_2^a I_2^a} i\tau_\mu^+ x_\mu = e^{-i\varphi_1^a(\tau^a/2)}(i\tau_\mu^+ x_\mu)e^{i\varphi_2^a(\tau^a/2)},$$

where φ_1^a and φ_2^a are the parameters of rotations. In other words, a rotation of x_μ is equivalent to the multiplication of unitary unimodular matrices both 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 by 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 $I_1^a + T_2^a$, the angular momentum operators, the introduced object has spin zero (here $T_{1,2}^a$ are the operators of the infinitesimal transformations (52) and (53)).

Thus, we see that the group SU(2) is distinguished on account of the dimension of the coordinate space. Further clarifying remarks as to why the group SU(2) is singled out are presented in Sec. 8.5.

7.3. *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 Q, the Euclidean action can be rewritten in the form

$$\begin{aligned} S &= \int \mathrm{d}^4x \frac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a = \int \mathrm{d}^4x \left[\frac{1}{4}G_{\mu\nu}^a\tilde{G}_{\mu\nu}^a + \frac{1}{8}(G_{\mu\nu}^a - \tilde{G}_{\mu\nu}^a)^2\right] \\ &= Q\frac{8\pi^2}{g^2} + \frac{1}{8}\int \mathrm{d}^4x (G_{\mu\nu}^a - \tilde{G}_{\mu\nu}^a)^2. \end{aligned} \tag{58}$$

It is clear from this formula that in the class of functions with given positive Q the minimum of S is attained for $G_{\mu\nu}^a = \tilde{G}_{\mu\nu}^a$ and is equal to $(8\pi^2 Q/g^2)$. We recall the functions with different Q cannot be related by a continuous deformation if the action is to remain finite. Therefore, minimization of the action can be carried out separately in each class of functions with the given Q. The BPST instanton has $Q = 1$, and the action $S = 8\pi^2/g^2$.

The case of negative Q 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 $Q \to -Q$. Thus, the minimum of the action for negative Q is $(8\pi^2/g^2)|Q|$, and it is attained when $G^a_{\mu\nu} = -\tilde{G}^a_{\mu\nu}$.

As can be seen from this discussion, the fulfillment of the self-duality and antiself-duality conditions $G^a_{\mu\nu} = \pm\tilde{G}^a_{\mu\nu}$ 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

$$D_\mu G^a_{\mu\nu} = D_\mu \tilde{G}^a_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\gamma\delta} D_\mu G^a_{\gamma\delta}$$

$$= \frac{1}{6}\varepsilon_{\mu\nu\gamma\delta}(D_\mu G^a_{\gamma\delta} + D_\gamma G^a_{\delta\mu} + D_\delta G^a_{\mu\gamma}) = 0\,,$$

where we have used the Bianchi identity:

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

8. Explicit Form of the BPST Instanton

8.1. *Solution with $Q = 1$*

As discussed in the previous section, the asymptotic behavior of A^a_μ for this solution is

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

where the matrices $\tau^\pm_\mu$ 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 Sec. 8.3.

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

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

For an instanton with its 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

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

where

$$f(x^2) \xrightarrow[x^2\to\infty]{} 1\,, \qquad f(x^2) \xrightarrow[x^2\to 0]{} \text{const}\cdot x^2\,.$$

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 $G^a_{\mu\nu}$. From (60), we obtain for $G^a_{\mu\nu}$

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

Here the prime denotes differentiation with respect to x^2. 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 at the end in Sec. 8.3. Using the formula for $\varepsilon_{\mu\nu\gamma\delta}\eta_{a\delta\rho}$ from the same list, we obtain for $\tilde{G}^a_{\mu\nu}$ the expression

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

The condition of self-duality, $G^a_{\mu\nu} = \tilde{G}^a_{\mu\nu}$, requires the 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 instanton size or, the instanton radius. The translational invariance guarantees obtaining the solution with the center at an arbitrary point x_0, for which it is necessary to replace x by $x - x_0$. We will discuss ρ, x_0 and other collective coordinates in more detail later. Note that if $f - \frac{1}{2}$ is denoted as X and $x^2 = e^\tau$ the equation for f becomes identical to the first-order differential equation one obtains in the double-well potential problem, $\dot{X} = \frac{1}{4} - X^2$.

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

$$A^a_\mu = \frac{2}{g}\eta_{a\mu\nu}\frac{(x-x_0)_\nu}{(x-x_0)^2+\rho^2}\,, \qquad G^a_{\mu\nu} = -\frac{4}{g}\eta_{a\mu\nu}\frac{\rho^2}{[(x-x_0)^2+\rho^2]^2}\,. \tag{63}$$

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

8.2. *Singular Gauge. The 't Hooft Ansatz*

It is frequently convenient to use the expression for A^a_μ in the so-called singular gauge, when the "bad" behavior of A^a_μ 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 $x \to \infty$.[12] We write down the formulas of the gauge transformation,

$$\begin{aligned} g\frac{\tau^a}{2}\bar{A}^a_\mu &= U^+ g\frac{\tau^a}{2}A^a_\mu U + iU^+\partial_\mu U\,, \\ g\frac{\tau^a}{2}\bar{G}^a_{\mu\nu} &= U^+ g\frac{\tau^a}{2}G^a_{\mu\nu}U\,, \end{aligned} \tag{64}$$

and for an instanton with its 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 $\bar{A}^a_\mu$ and the field strength tensor $\bar{G}^a_{\mu\nu}$ in the singular gauge we obtain

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

It is obvious that the quantities $G^a_{\mu\nu}G^a_{\gamma\delta}$ are invariants of the gauge transformation (see, however, the last footnote). Note also the fact 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 saturated in the neighborhood of $x = x_0$ and not at infinity.

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

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

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

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

then for $G^a_{\mu\nu} - \tilde{G}^a_{\mu\nu}$ we obtain [see the properties of the η symbols in Sec. 8.3]

$$G^a_{\mu\nu} - \tilde{G}^a_{\mu\nu} = \frac{1}{g}\bar{\eta}_{a\mu\nu}\frac{\partial_\gamma\partial_\gamma W}{W}\,.$$

The self-duality of $G^a_{\mu\nu}$ requires fulfillment of the equation $\partial_\gamma\partial_\gamma W/W = 0$. The solution with topological charge Q 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

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

It should be noted that the choice of A^a_μ in the form (67) did not give the most general solution with charge Q, since all Q instantons have the same orientation in the color space (for the construction of the general solution, see Ref. [22]).

8.3. *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):

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

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

[12] 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, $G^a_{\mu\nu}G^a_{\mu\nu}$. To use such transformations, it is necessary to consider a space with punctured singular points. This we shall do, remembering that the physical quantitites are nonsingular at the singular points.

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

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

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

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

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

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

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

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

8.4. *Instanton in the $A_0 = 0$ Gauge*

In Sec. 7.1 it was mentioned that the relation between the instanton topological charge Q and the winding numbers of the zero-energy states in the distant past and distant future, between which it interpolates,

$$Q = \mathcal{K}' - \mathcal{K}\,,$$

is most transparently seen in the $A_0 = 0$ gauge. Now we can explicitly demonstrate that this relation does indeed take place.

Equations (59) and (62) imply that the instanton field

$$A_\mu = \frac{x^2}{x^2+\rho^2}\, iS_1\partial_\mu S_1^+$$

where $A_\mu = (g/2)\tau^a A_\mu^a$ and the matrix S_1 is defined in Eq. (59). Let us presume that the time component of the gauge-transformed field vanishes identically,

$$U^+A_\mu U + iU^+\partial_\mu U = 0 \text{ at } \mu = 0\,.$$

Substituting the expression for the instanton field we get the following equation for the gauge matrix U transforming the BPST instanton to the $A_0 = 0$ gauge,

$$\dot{U} + \frac{x^2}{x^2+\rho^2}(S_1\dot{S}_1^+)U = 0\,,$$

where

$$S_1\dot{S}_1^+ = i\,\frac{\mathbf{x}\boldsymbol{\sigma}}{x^2}\,.$$

The solution of this equation is obvious,

$$U(\tau,\mathbf{x}) = \left[\exp\int_{-\infty}^{\tau}\left(\frac{i\mathbf{x}\boldsymbol{\sigma}}{x^2+\rho^2}\right)d\tau\right]U(\tau=-\infty,\mathbf{x})\,.$$

The instanton field in the $A_0 = 0$ gauge takes the form

$$A_i(\tau,\mathbf{x}) = i\left\{\frac{x^2}{x^2+\rho^2}U^+(\tau,\mathbf{x})(S_1\partial_i S_1^+)U(\tau,\mathbf{x}) + U^+(\tau,\mathbf{x})\partial_i U(\tau,\mathbf{x})\right\}.$$

In the distant past and distant future

$$A_i \to i(U^+S_1)\partial_i(U^+S_1)^{-1}\,.$$

Moreover, $S_1 \to 1$ at $\tau \to = \pm\infty$. As for $U(\tau = +\infty)$, we have

$$\begin{aligned} U(\tau=+\infty,\mathbf{x}) &= \left[\exp\int_{-\infty}^{+\infty}\left(-\frac{i\mathbf{x}\boldsymbol{\sigma}}{x^2+\rho^2}\right)d\tau\right]U(\tau=-\infty,\mathbf{x}) \\ &= \left[\exp\left(-\frac{i\pi\mathbf{x}\boldsymbol{\sigma}}{\sqrt{\mathbf{x}^2+\rho^2}}\right)\right]U(\tau=-\infty,\mathbf{x})\,. \end{aligned}$$

The hedgehog matrix appears on the right-hand side. This obviously concludes the proof that the winding numbers of the filed configurations between which the instanton interpolates differ by unity.

8.5. *Instanton Collective Coordinates*

The instanton solution presented in Eq. (63) has the following collective coordinates: the instanton size ρ (associated with dilatations); and four parameters represented by the instanton center x_0 (associated with translations). The issue of the collective coordinates is important, since each of them gives rise to a zero mode. As we already know (Sec. 3), the latter play a special role in calculating the instanton determinant, and, eventually, the instanton measure. Thus, it is imperative to establish a complete set of the collective coordinates. In this section we will analyze the set of the collective coordinates for the SU(2) instanton.

The action of the pure Yang-Mills theory, Eq. (58), has no dimensional parameters, and is conformally invariant at the classical level. Since the instanton is the solution of the classical equations of motion (which are naturally conformally invariant too), the set of the collective coordinates appearing in the generic instanton solution is determined by the conformal group. Each given instanton solution breaks (spontaneously) some of the invariances. The conformal symmetry is restored only upon consideration of the family of the solutions, as a whole. Those symmetry transformations that act on the instanton solution nontrivially, generate another solution belonging to the same family, with "shifted" values of the collective coordinates. Thus, each symmetry transformation from the conformal group, which does not leave the instanton solution intact, requires a separate collective coordinate.

The conformal group in four dimensions includes 15 transformations (e.g. Ref. [23]): four translations, six Lorentz rotations (in the Euclidean space it is more appropriate to speak of six $O(4)$ rotations), four proper conformal transformations and one dilatation. Moreover, the Yang-Mills action is gauge invariant. We do not need to consider (small) gauge transformations of the instanton, since they produce just the same solution in a different gauge. Global rotations in the color space have to be considered, however. In the SU(2) theory there are three global rotations. Thus, *a priori* one could expect the generic instanton solution to depend on 18 collective coordinates. So far, we have only seen five. Where are the remaining collective coordinates?

The proper conformal transformations can be represented as a combination of translations and inversion. Under inversion

$$x_\mu \to x'_\mu = \frac{x_\mu}{x^2}\,, \qquad A_\mu(x) \to x'^2 A_\mu(x')\,.$$

Translations are already represented by the corresponding collective coordinate, x_0. Now, if we start from the original BPST instanton with the unit radius and make the inversion, we will obviously get an antiinstanton in the singular gauge,

$$\frac{2}{g}\eta_{a\mu\nu}\frac{x_\nu}{x^2+1} \xrightarrow{\text{inv.}} \frac{2}{g}\eta_{a\mu\nu}\frac{x_\nu}{x^2(x^2+1)}$$

(cf. Eqs. (63) and (65)). Thus, no new collective coordinates are associated with the proper conformal transformations.

What remains to be discussed? We must consider six rotations of the Euclidean space and three global color rotations. We will show that only three linear combinations of these nine generators act on the instanton solution nontrivially, resulting in three extra collective coordinates. As a matter of fact, the argument was already outlined in Sec. 7.2 where we explained why the gauge group SU(2) is singled out. Here we rephrase the argument [24] in a more explicit form by passing to a convenient spinorial formalism which we will need later on anyway. This formalism becomes practically indispensable in dealing with the chiral fermions.

For arbitrary vector V_μ from the Euclidean space-time introduce $V_{\alpha\dot\alpha}$,

$$V_{\alpha\dot\alpha} = (\sigma_\mu)_{\alpha\dot\alpha}V_\mu\,,\ (\mu = 1,\dots,4, \alpha,\dot\alpha = 1,2)\,,$$

where

$$\sigma_\mu \equiv \tau^-_\mu = \{\boldsymbol{\sigma}, i\}\,.$$

(The matrices $\tau^\pm_\mu$ were introduced in Eq. (57). Here we find it converient to pass to a more concise notation.) The undotted and dotted indices from the beginning of the Greek alphabet will denote the spinorial indices of the $\mathrm{SU}(2)_L$ and $\mathrm{SU}(2)_R$ subgroups of the Lorentz group in the Euclidean space ($O(4) = \mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$). A similar definition refers to the coordinate x_μ itself, $x_{\alpha\dot\alpha} = -(\sigma_\mu)_{\alpha\dot\alpha}x_\mu$. The lowering and raising of the spinorial indices are done by multiplying by antisymmetric tensor (the Levi-Civita tensor) from the left,

$$\chi^\alpha = \varepsilon^{\alpha\beta}\chi_\beta\,, \qquad \chi_\alpha = \varepsilon_{\alpha\beta}\chi^\beta\,,$$

and the same for the dotted indices, where

$$\varepsilon^{\alpha\beta} = -\varepsilon^{\beta\alpha}, \qquad \varepsilon^{12} = -\varepsilon_{12} = 1\,.$$

Then,

$$x^1_{\dot1} = -x_1 - ix_2,\ x^1_{\dot2} = -ix_4 + x_3,\ x^2_{\dot1} = ix_4 + x_3,\ x^2_{\dot2} = x_1 - ix_2\,.$$

In this notation the instanton field takes the form

$$A^{\eta\xi}_{\alpha\dot\beta} = -\frac{i}{g}\frac{1}{x^2+\rho^2}(x^\eta_{\dot\beta}\delta^\xi_\alpha + x^\xi_{\dot\beta}\delta^\eta_\alpha)$$

where the color index of A^a_μ is also converted into two spinorial indices according to the formula

$$A^{\eta\xi} = \frac{1}{2}(\tau^a)^\eta_\delta\varepsilon^{\xi\delta}\,,$$

and $\varepsilon^{\xi\delta}$ is the Levi-Civita tensor.

For completeness we give here the expression for the gluon field strength tensor,

$$G^{\gamma\delta}_{\alpha\beta} = \frac{4i}{g}\frac{\rho^2}{(x^2+\rho^2)^2}(\delta^\gamma_\alpha\delta^\delta_\beta + \delta^\gamma_\beta\delta^\delta_\alpha), \quad G^{\gamma\delta}_{\dot\alpha\dot\beta} = 0\,,$$

where the transition from the standard vectorial to spinorial notation is done in the following way

$$(\sigma_\mu)_{\alpha\dot\alpha}(\sigma_\nu)_{\beta\dot\beta}G_{\mu\nu} = \varepsilon_{\dot\alpha\dot\beta}G_{\alpha\beta} + \varepsilon_{\alpha\beta}G_{\dot\alpha\dot\beta}\,.$$

$G_{\alpha\beta}$ and $G_{\dot\alpha\dot\beta}$ are self-dual and anti-self-dual parts of the gluon field strength tensor,[13] so that the equation $G^{\gamma\delta}_{\dot\alpha\dot\beta} = 0$ is nothing but the condition of self-duality of the instanton solution.

Now we are ready to discuss what happens with the instanton when one does the Lorentz and/or color rotations. Note that if in the standard notation the global *color* rotation acts on the four-potential A as $A \to M^+AM$, then within our new convention

$$A^{\alpha\beta} \to (M^+)^\alpha_\gamma (M^+)^\beta_\delta A^{\gamma\delta}\,.$$

Here $M = \exp(i\omega^a\tau^a/2)$, and ω^a are three parameters corresponding to a global rotation in the SU(2) color group.

Let the left-handed rotation act on the vector with the upper undotted index as

$$x^\eta_{\dot\beta} \to L^\eta_\alpha x^\alpha_{\dot\beta}\,,$$

where L is a matrix from $\mathrm{SU}(2)_L$. Then for the vectors with the lower undotted index

$$x_{\alpha\dot\beta} \to (L^+)^\gamma_\alpha x_{\gamma\dot\beta}\,.$$

Transformations from $\mathrm{SU}(2)_R$ (they act on the dotted indices) rotate x and A in the same way. In other words, the form of the instanton solution does not change at all. No collective coordinates corresponding to the $\mathrm{SU}(2)_R$ rotations emerge.

[13] From this expression for $G^{\gamma\delta}_{\alpha\beta}$ it is clear that the field configuration under discussion is a hedgehog in the $\mathrm{SU}(2)_L$ Lorentz subgroup. According to the nomenclature suggested by 't Hooft, it should be actually called *anti*instanton. Indeed, passing to the standard vectorial notation we get $\bar\eta_{a\mu\nu}$ rather than $\eta_{a\mu\nu}$.

Transformations from $\mathrm{SU}(2)_L$ do change the form of the instanton solution, and so do the global color rotations. Under the combined action of these two transformations the instanton solution becomes

$$A^{\eta\xi}_{\alpha\dot\beta} = \frac{i}{x^2+\rho^2}(M^+)^\eta_{\eta'}(M^+)^\xi_{\xi'}(L^{\eta'}_\gamma L^{\xi'}_\alpha x^\gamma_{\dot\beta} + L^{\xi'}_\gamma L^{\eta'}_\alpha x^\gamma_{\dot\beta})\,.$$

We see that if $L = M$ the solution remains intact. This means that out of six transformations (three global color rotations and three $\mathrm{SU}(2)_L$ rotations) only three are independent, giving rise to three collective coordinates. We can choose them to be associated either with the global color rotations (as is usually assumed) or with the rotations from $\mathrm{SU}(2)_L$. In the conventional formalism these three collective coordinates are introduced as

$$\eta_{a\mu\nu} \to O_{ab}\eta_{b\mu\nu}\,, \qquad \bar\eta_{a\mu\nu} \to O_{ab}\bar\eta_{b\mu\nu}$$

where the three-by-three matrix O_{ab} is defined as

$$O_{ab} = \frac{1}{2}\,\mathrm{Tr}(M\tau^a M^+\tau^b)\,.$$

The advantage of the spinorial formalism is obvious — there is no need to introduce the 't Hooft symbols, and the hedgehog nature of instanton is most transparent.

Summarizing, there are eight collective coordinates characterizing the SU(2) instanton. Correspondingly, we will observe eight zero modes. For higher gauge groups the number of the collective coordinates corresponding to global color rotations increases. Altogether, in the $SU(N)$ group the BPST instanton has $4N$ collective coordinates. The counting was first carried out in Ref. [25]. We will return to the discussion of the $SU(N)$ instanton in Sec. 9.5.

8.6. *Instantons in the Higgs Regime*

Quantum chromodynamics is not the only gauge theory of practical importance. The Standard model of the electroweak interactions is a gauge theory too. A drastic distinction in their dynamical behavior is due to the fact that the non-Abelian gauge group is spontaneously broken in the standard model due to the Higgs mechanism, the coupling constant is frozen at the values of momenta of order of the W boson mass. It never becomes strong. Correspondingly, the color confinement and other peculiar phenomena of QCD do not take place. Since the nontrivial topology in the space of the gauge fields is not affected by

the Higgs phenomenon, instantons (as the tunneling trajectories) exist in the standard model too, leading to certain nonperturbative effects. The one which was under the most intense scrutiny is the baryon number violations at high energies. We will not dwell on this, applied, aspect of the instanton calculus, referring the reader to Ref. [26] for a detailed review. Instead, we will focus on theoretical issues. As a matter of fact, as we will see below, consideration of instantons in the Higgs regime, even has certain advantages over the QCD instantons. Since the coupling constant never becomes large the quasiclassical approximation in the description of the tunneling phenomena, based on instantons, is always justified, in clear distinction with QCD, where the instanton contribution is dominated by large-size instantons, which are obviously outside the scope of the applicability of quasiclassical methods.

Since we are not concerned with applications, we will limit ourselves to a truncated standard model: SU(2) gauge group, with the minimal Higgs sector — one complex Higgs doublet χ^i, $i = 1, 2$. The $U(1)$ subgroup, as well as fermions, present in the standard model, are discarded.

If the Higgs field is in the fundamental representation of the color group, there is no clear-cut distinction between the confinement phase and the Higgs phase. As the vacuum expectation value (VEV) of the Higgs field χ continuously changes from large values to smaller ones, we continuously flow from the weak coupling regime to the strong coupling one. The spectrum of all physical states, and all other measurable quantities, change smoothly [27].

One can argue that the case is such in many different ways. Perhaps, the most straightforward line of reasoning is as follows. Using the Higgs field in the fundamental representation one can build gauge invariant interpolating operators for *all* possible physical states. The Källén-Lehmann spectral function corresponding to these operators, which carries complete information on the spectrum, depends smoothly on $\langle\chi^*\chi\rangle$. When the latter parameter is large the Higgs description is more convenient, when it is small it is more convenient to think in terms of the bound states. There is no sharp boundary. We deal with a single Higgs/confining phase [27].

It is convenient for our purposes to write the Lagrangian in the Higgs sector in a slightly nonconventional form.

The model at hand has a *global* SU(2) symmetry, associated with the possibility of rotating the doublet χ^i into the conjugated doublet $\varepsilon^{ij}\chi_i^\dagger$. (This global symmetry is responsible for the fact that all three W bosons are degenerate if the $U(1)$ interaction is switched off in the standard model.) The

SU(2) symmetry of the χ sector becomes explicit if we introduce a matrix field

$$X = \begin{pmatrix} \chi^1 & -\chi^{2\dagger} \\ \chi^2 & \chi^{1\dagger} \end{pmatrix},$$

and identically rewrite the standard Higgs Lagrangian in terms of this matrix,

$$\Delta\mathcal{L}_\chi = \frac{1}{2}\,\mathrm{Tr}\, D_\mu X^\dagger D_\mu X - \frac{\lambda}{4}\left(\frac{1}{2}\,\mathrm{Tr}\, X^\dagger X - \eta^2\right)^2$$

where $D_\mu X = (\partial_\mu - iA_\mu)X$. The complex doublet field χ^i develops a vacuum expectation value η which can be varied continuously.

All physical states form representations of the global SU(2). Consider, for instance, the SU(2) triplets produced from the vacuum by the operators

$$W_\mu^a = -\frac{i}{2}\mathrm{Tr}(X^\dagger \overleftrightarrow{D}_\mu X\tau^a), \qquad a = 1, 2, 3\,.$$

The lowest-lying states produced by these operators in the weak coupling regime (i.e. when $\langle\chi^\dagger\chi\rangle \gg \Lambda^2$) coincide with the conventional W bosons of the Higgs picture, up to a normalization constant. The mass of the W bosons is $\sim g\eta$. On the other hand, if $\langle\chi^\dagger\chi\rangle \lesssim \Lambda^2$ it is more appropriate to think of the bound states of the χ "quarks" forming vector mesons, triplet with respect to the global SU(2) ("ρ mesons"). Their mass is $\sim \Lambda$. The continuous evolution of η results in the continuous evolution of the mass of the corresponding states. It is easy to check that the complete set of the gauge invariant operators that one can build in this model spans the whole Hilbert space of the physical states.

By the same token, one can treat, in a gauge-invariant manner, elements of the instanton calculus. We will not pursue this line of reasoning in detail, referring the reader to the original publications (e.g. Ref. [24]). Instead, we will dwell on two problems: calculation of the instanton action in the Higgs regime, and the height of the barrier in Fig. 8.

Instanton action

Strictly speaking, if the scalar field develops a vacuum expectation value, the only exact solution of the classical equation of motion is the zero-size instanton. For each given value of ρ one can make the action of the tunneling trajectory smaller by choosing a smaller value of ρ, so that $8\pi^2/g^2$ is achieved asymptotically (see below). Since the nontrivial topology remains intact (one direction

in the space of fields forms a circle), for proper understanding of the tunneling phenomena one cannot disregard the trajectories connecting the zero-energy gauge copies (pre-vacua) in the Euclidean time, even though they are not exact solutions any more. Following 't Hooft [9], we will consider *constrained* instantons — trajectories that minimize the action under the condition that the size ρ is fixed. Our analysis will be somewhat heuristic. More rigorously the construction is described, for example, in Ref. [28].

Technically the procedure can be summarized as follows. We first find the solutions of the classical (Euclidean) equations of motion for the gauge field ignoring the scalar field altogether. The solution is of course the familiar instanton. Then we look for the solution of the equations of motion for the χ field in the given instanton background. This solution minimizes the Higgs part of the action. A nonvanishing scalar field, in turn, induces a source term in the equation for the gauge field, which is neglected. This source term will push the instanton towards smaller sizes, in particular, by cutting off the tails of the A_μ field at large distances (where they should become exponentially small). The distance where this occurs is of order $1/(g\eta)$. If we are interested in distances of order $1/\eta$ — and the instanton contributions are indeed saturated at such distances — then we can neglect this effect and continue to disregard the back reaction of the scalar field in considering instantons whose sizes are fixed by hand.

To keep our analysis as simple as possible we will further assume that the scalar self-coupling $\lambda \to 0$. The only role of the scalar self-interaction then is to provide the boundary condition at large distances,

$$\frac{1}{2}\,\mathrm{Tr}(X^+X) \to \eta^2\,.$$

The equation of motion of the scalar field is completely determined by the kinetic term in the Lagrangian,

$$D^2X = 0\,.$$

If the instanton field is written as

$$A_\mu = \frac{ix^2}{x^2+\rho^2} S_1 \partial_\mu S_1^+$$

(for the antiinstanton $S_1 \leftrightarrow S_1^+$; the matrix S_1 is defined in Eq. (59)), it is not difficult to check that the solution of the equation $D^2X = 0$ takes the form[14]

$$X = \eta\left(\frac{x^2}{x^2+\rho^2}\right)^{1/2} S_1\,.$$

This expression is properly normalized. Asymptotically the modulus of the Higgs field approaches its value in the "empty" vacuum, while the "phase" part of the scalar field has a hedgehog winding. At small x the VEV is suppressed.

Using the fact that the Higgs field satisfies the equation of motion we readily rewrite the contribution of the Higgs kinetic term in the action as

$$\int d^4x \partial_\mu \frac{1}{2}\,\mathrm{Tr}(X^+D_\mu X)\,.$$

Moreover,

$$X^+D_\mu X = \eta^2\rho^2 \frac{x_\mu}{(x^2+\rho^2)^2} + \eta^2\rho^2 \frac{x^2}{(x^2+\rho^2)^2}(S_1^+\partial_\mu S_1)\,.$$

The last bracket, being an element of the algebra, is proportional to σ^a and, hence, vanishes upon taking the color trace. Therefore, the trace is determined entirely by the first term. Exploiting now the Gauss theorem and rewriting the volume integral as that over the surface of the large sphere dS_μ we arrive at

$$\int d^4x \partial_\mu \frac{1}{2}\,\mathrm{Tr}(X^+D_\mu X) = \int dS_\mu \eta^2\rho^2 \frac{x_\mu}{(x^2+\rho^2)^2} = 2\pi^2\eta^2\rho^2\,.$$

Summarizing, the extra term in the action induced by a nonvanishing vacuum the expectation value of the Higgs field has the form

$$\Delta S = 2\pi^2\eta^2\rho^2\,.$$

This term is called the 't Hooft interaction, since 't Hooft was the first to calculate it [9]. The 't Hooft interaction explicitly exhibits the feature we anticipated earlier — the smaller the instanton size ρ the smaller is the instanton action. It is clear that the instanton contribution to physical quantities is determined by an integral over ρ (in Sec. 9 we will calculate the instanton density

[14] Exercise: Verify that the expression presented here is indeed a solution.

which will give us the measure of integration). The exponential suppression of the instanton density at large ρ due to the 't Hooft term, $\exp(-2\pi^2\eta^2\rho^2)$, guarantees that $\rho \sim \eta^{-1}$. This, in turn, justifies the approximations made: the back reaction of the scalar field on the gauge field becomes important at much larger distances, $x \sim (g\eta)^{-1}$.

In the SU(2) theory the 't Hooft interaction depends only on one collective coordinate, ρ. In more complicated examples it may acquire dependences on other collective coordinates. For instance, if we consider an SU(3) model with one Higgs triplet (breaking SU(3) down to SU(2)) then the 't Hooft interaction will depend, roughly speaking, on the orientation of the instanton in the color space relatively to the direction of the Higgs VEV. The 't Hooft term becomes $2\pi^2\eta^2\rho^2\cos^2(\alpha/2)$ where α is a certain angle. If the instanton under consideration resides in the corner of SU(3) corresponding to the unbroken SU(2), then $\alpha = \pi$ and the 't Hooft term vanishes. Further details can be found in Ref. [24].

8.7. *The Height of the Barrier. Sphaleron*

Let us return to the tunneling interpretation of instantons discussed in Sec. 5 and ask the question on what is the height of the barrier in Figs. 7 or 8, under which the tunnelings described by the instanton trajectory take place. This issue is not so simple as it might seem at first sight.

Indeed, the QCD Lagrangian at the classical level contains no dimensional parameters. Since the instantons are solutions of the classical equations of motion (in the Euclidean space), they do not carry dimensional constants other than the instanton size which is a variable parameter. The height of the barrier must have a dimension of mass. Therefore, the smaller the size ρ, the higher the barrier the instanton with the given ρ "sees", so that the classical action stays constant, $8\pi^2/g^2$. This is possible, of course, due to the fact that the space of fields is actually infinitely-dimensional. The one-dimensional plot depicted in Fig. 8 is symbolic. Since the infrared limit of QCD is not tractable quasiclassically it is impossible to determine the lowest possible height of the barrier under which the system tunnels. All we can say is that it is of order $\Lambda_{\rm QCD}$.

The situation drastically changes in the Higgs regime considered in the previous section. The vacuum expectation value of the Higgs field provides masses to all gauge bosons. If the vacuum expectation value is much larger than $\Lambda_{\rm QCD}$, the coupling constant always stays small, and the quasiclassical

picture is fully applicable. Under the circumstances the question on what is the minimal height of the barrier in Fig. 8 becomes amenable to quantitative analysis. From this figure it is clear that when the system sits right on top of the barrier, this is a solution of the *static* equations of motion, since the position on top of the barrier is an equilibrium. It is also clear that the equilibrium is unstable since it corresponds to a maximum of energy rather than a minimum.

Thus, we will look for the solution of the static equations of motion of the Yang-Mills Lagrangian in the $A_0 = 0$ gauge. By inspecting the structure of these equations it is easy to guess an *Ansatz* which untangles the color and Lorentz indices,

$$A_i^a = \frac{1}{g}\varepsilon_{iak}\frac{x^k}{r}f(r)\,, \qquad X = \frac{\boldsymbol{\sigma}\mathbf{x}}{r}h(r)\,,$$

where $r = \sqrt{\mathbf{x}^2}$ and f, h are profile functions to be determined from the equations. The boundary conditions are quite obvious: at $r \to 0$ both functions f and h must tend to zero to avoid singularities; at $r \to \infty$ the function h tends to η while $f(r) \to -2/r$. The latter condition is necessary to ensure that $A_i(r)$ becomes pure gauge at infinity. Then the energy density of the gauge field vanishes at large r. Simultaneously, the energy density of the scalar field also vanishes in spite of the winding of the field X. The overall energy of the field configuration under consideration can be expected to be finite if both conditions are met.

Technically, instead of solving the equations of motion it is more convenient to write out the energy functional and minimize it with respect to f and h under the given boundary conditions. Substituting our *Ansatz* in the Lagrangian presented in the previous section we readily obtain

$$\mathcal{H} = 4\pi\int_0^\infty r^2 dr\left\{\frac{1}{g^2}\left[f'^2 + \frac{2}{r^2}f^2 + \frac{2}{r}f^3 + \frac{1}{2}f^4\right] + h'^2 + 2h^2\left(\frac{1}{r} + \frac{f}{2}\right)^2\right\}.$$

The terms in the square brackets are from the gauge part (integration by parts is carried out in one of the terms). The second term represents the Higgs part. Since all terms in $\mathcal{H}$ are positive-definite it is clear that a minimum of this functional exists. It can be found numerically; the corresponding profile functions are depicted in Fig. 9.

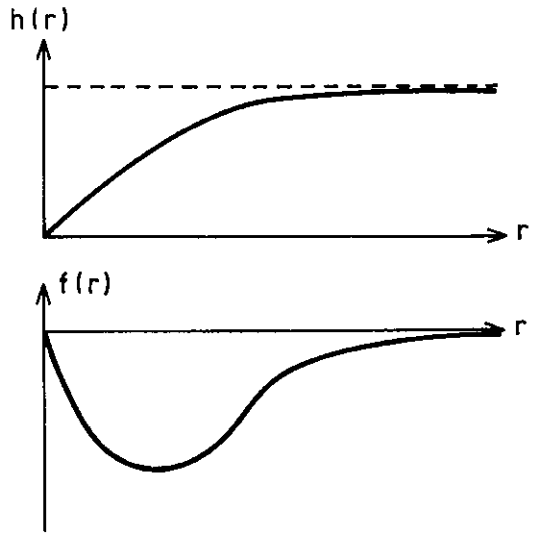


Fig. 9. The solutions for the sphaleron profile functions.

Before minimization it is convenient to rescale the fields and the variable r to make them dimensionless,

$$f = g\eta F, \qquad h = \eta H, \qquad r = R(g\eta)^{-1}.$$

In terms of the rescaled fields the energy functional takes the form

$$\mathcal{H} = 4\pi\frac{\eta}{g}\int_0^\infty R^2 dR\Big\{\Big[F'^2 + \frac{2}{R^2}F^2 + \frac{2}{R}F^3 + \frac{1}{2}F^4\Big] + H'^2 + 2H^2\Big(\frac{1}{R} + \frac{F}{2}\Big)^2\Big\},$$

where the prime here denotes differentiation over R. The expression in the braces contains no parameters, neither does the boundary conditions for the dimensionless fields F, H (at $R \to \infty$ the function H approaches unity and the function F tends to $-2/R$). The only parameter of the problem, η/g, is an overall factor. This obviously means that the energy of the solution obtained

by minimizing the energy functional $\mathcal{H}$ is

$$E \equiv \mathcal{H}_{\min} = \text{constant of order unity} \times \frac{\eta}{g}.$$

The numerical value of the constant is not very important for our illustrative purposes. It can be found in the original papers, see e.g. Ref. [29].

The static solution outlined above, corresponding to the top of the barrier, is called *sphaleron*, from the Greek adjective *sphaleros* meaning unstable, ready to fall. It was found in the SU(2) theory in Ref. [30] and rediscovered later in the context of the standard model by Klinkhamer and Manton ([29]; see also Ref. [31]), who were the first to interpret the sphaleron energy as the height of the barrier separating distinct pre-vacua of the Yang-Mills theory in the Higgs regime. It is instructive to examine the position of the sphaleron on the plot of Fig. 8 directly, by calculating the winding number of the corresponding gauge field. Note that at large distances

$$(A_i)_{\text{sph}} \to iU\partial_i U^+, \qquad U = \frac{\sigma \mathbf{x}}{r}.$$

The matrix U takes different values as we approach infinity from different directions. Thus, the condition of compactification which we impose on the vacuum gauge field, does not hold for the sphaleron. Correspondingly, the winding number $\mathcal{K}[(A_i)_{\text{sph}}]$ need not be an integer. A direct calculation (left as an exercise for the reader) readily yields

$$\mathcal{K}[(A_i)_{\text{sph}}] = \frac{1}{2},$$

demonstrating that the sphaleron sits right in the middle between two classical minima with $\mathcal{K} = 0$ and $\mathcal{K} = 1$.

To give a well-defined quantitative meaning to the height of the barrier in the absence of the Higgs field we must regularize the Yang-Mills theory in the infrared domain. One of the possible regularizations was suggested in Ref. [32] where the Yang-Mills fields were put on a three-dimensional sphere of a finite radius, instead of the flat space of conventional QCD. The radius of the sphere plays essentially the same role as $(g\eta)^{-1}$ in the Higgs picture. If this radius is small, the quasiclassical consideration becomes closed, and one naturally discovers analogs of the sphaleron solution. The advantage of this regularization over the Higgs field regularization is the existence of analytic expressions. Both, the sphaleron field configuration and its energy can be found analytically [32]. In particular, the sphaleron energy turns out to be $3\pi^2/g^2$ times the inverse radius of the sphere.

9. Calculation of the Pre-Exponential Factor for the BPST Instanton

9.1. *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 $A_\mu^{a(\mathrm{ins})}$:

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

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 the 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

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

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

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

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

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

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 appearing 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 cut-off 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

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

or, more precisely, the ratio of $\langle 0|0_T\rangle_{\mathrm{ins}}^{\mathrm{Reg}}$ to the corresponding perturbative 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 bare coupling constant normalized at the cut-off 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. The study of zero modes plays the main part in obtaining the result.

9.2. *Zero Modes*

As shown in the one-dimensional example, each zero mode 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)?

The issue has been discussed in Sec. 8.5. 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 color space. The spatial rotations need not be counted, since they are equivalent to isorotations (see Sec. 8.5).

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 mode gives rise to a factor M. Thus, from all (since we have listed *all* collective coordinates) zero modes, there arises in $\langle 0|0_T\rangle_{\text{ins}}^{\text{Reg}}$ the factor

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

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.

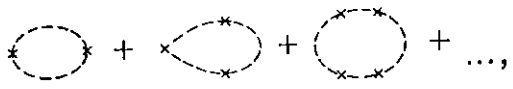
Using (75) we rewrite $\langle 0|0_T\rangle_{\text{ins}}^{\text{Reg}}/\langle 0|0_T\rangle_{\text{p.th}}$ in the form

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

where $\exp\Phi_1$ denotes the contribution of the nonzero modes.

9.3. *Nonzero Modes. Effective Coupling Constant*

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 factor by retaining the terms quadratic in the deviation from the external field corresponds to the calculation of the one-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 $A_\mu^{a(\text{ins})}$.

It is clear that complete calculation of the contribution of the zero modes requires summation of a complete chain of diagrams — the zero 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 $\ln\langle 0|0_T\rangle_{\text{ins}}$. It is also clear that there is no nonanalyticity of this kind for the nonzero 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 cut-off 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 nonzero 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 the spatial transverse polarization states; the unphysical polarizations and ghosts are not necessary. Omitting the details of this simple calculation, we give the result for $\operatorname{Im}\Pi_{\mu\nu}$:

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

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

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

where we cut off the integration over s at M^2, since the regularization involves a subtraction of an analogous contribution with the Pauli-Villars regulators of mass M.

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

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

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). Passing 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 and nonzero modes has the consequence that $8\pi^2/g_0^2$ in the argument of the exponential (76) is replaced by the effective coupling constant $8\pi^2/g^2(\rho)$:

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

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 the non-Abelian theory which arises when the zero modes are considered. In the instanton calculation the antiscreening is entirely due to zero modes.

Indeed, both the sign and the magnitude of the coefficient of the "antiscreening" logarithm in Eq. (76), 8 $\ln M\rho$, are obvious consequences of the consideration above — the coefficient is simply the number of the zero modes.

In the framework of the perturbative calculations, the "antiscreening" result can be most clearly explained in the framework of the ghost-free Coulomb gauge, which was used in calculations by Khriplovich [33] as early as 1969 (see also Ref. [34]). 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 imaginary part and arises when one of the virtual quanta has the spatial 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 the Coulomb quantum and the transverse quantum (electric forces repel charges of the same sign, while magnetic forces attract currents of the same type). A more detailed pedagogical discussion of the issue can be found in Ref. [35].

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 modes. With this we conclude our panegyric to the zero modes.

9.4. *Two-Loop Approximation*

The above calculations led to the replacement of the bare coupling constant g_0 in the classical action by the effective constant $g(\rho)$. However, the bare 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 one-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.

From the two-loop calculation we actually only need the expression for the effective coupling constant; such an expression as known from the perturbation theory [36],

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

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

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

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 one-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 the determination of their contribution to the coefficient of the unit operator. The calculation of the coefficients of other operators will be considered in Sec. 10. A specific feature of the unit operator is the fact that its matrix elements are independent of the normalization point; i.e. it has the vanishing anomalous dimension. Therefore, the coefficient of it, expressed in terms of $g(\rho)$, cannot contain ρ_c (for operators with a nonvanishing anomalous dimension the factor $[g^2(\rho_c)/g^2(\rho)]^\delta$ arises).

Now, it only remains to express $g^2(\rho)$ in terms of $g^2(\rho_0)$ by means of the renormalization-group equations; the argument above proves that the retention of the two-loop correction in (82) is fully legitimate.

9.5. *Density of Instantons (The Instanton Measure) in the Group SU(N)*

How does the number of zero 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 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. 10). Those of the remaining generators that occur in the $(N-2) \times (N-2)$ matrix at the bottom right obviously do not rotate the instanton field. Thus, to the three SU(2) rotations a further $4(N-2)$ unitary rotations are added. The total number of zero 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 instanton density $d(\rho)$, which is defined as follows:

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

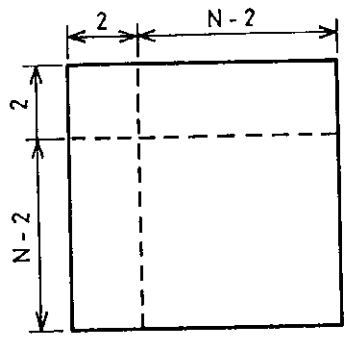


Fig. 10. Counting generators of group rotations in SU(N).

Since the SU(N) group space is finite it is assumed that the integration over the collective coordinates associated with the instanton orientation in the SU(N) group is carried out. 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 [37] of the 't Hooft's calculations [9]. Concretely, it is necessary to take into account additional $4(N-2)$ vector fields with the above quantum numbers in the contribution of both the zero and nonzero modes. In addition, we need 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. [37]). 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 bare constant. Equation (86) refers to the Pauli-Villars regularization. Instead of the Pauli-Villars regularization (PV scheme), the so-called dimensional regularization is frequently used. Instead of logarithms of the cut-off parameter, poles with respect to the dimension of space arise in this method, $\ln M \to 1/(4-D)$. Use of the minimal subtraction scheme [38] (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 the perturbation theory the so-called modified minimal subtraction ($\overline{\text{MS}}$) scheme has proved helpful, since in it, too large coefficients of the expansion in g^2 do not arise [39]. The difference between the $\overline{\text{MS}}$ scheme and the MS scheme reduces to

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

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

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

10. Instantons in the QCD Vacuum

10.1. *Instantons in the Slowly Varying Background Fields*

The quasiclassical methods that have been developed apply to the study of nonperturbative fluctuations of a 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 QCD vacuum [40].

Since we distinguish fields of two types, namely, the fields of small-scale instantons and the fields of the large-sized 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 in the space of the slowly varying fields.

Thus, the effect of a selected instanton with scale ρ and center at x_0 reduces to the following extra term in 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 convenient to consider the matrix elements of ΔL between the 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

$$\langle n \text{ gluons } |\Delta L|0\rangle = \langle 0|\,T\prod_{k=1}^{n}\int \mathrm{d}x_k e^{iq_kx_k}\varepsilon^{a_k}_{\mu_k}q_k^2A^{a_k}_{\mu_k}(x_k)|0\rangle\,, \tag{90}$$

where q_k and $\varepsilon^{a_k}_{\mu_k}$ are the four-momentum and the polarization of the kth gluon, and $A^a_\mu(k)$ is the operator of the gluon field. For $n = 0$, i.e. for the vacuum–vacuum transition, the right-hand side of (90) was already calculated in Sec. 9 and is equal to $\mathrm{d}\rho\rho^{-5}d(\rho)$; the left-hand side is obviously equal to the coefficient of the unit operator: $C_I\mathrm{d}\rho/\rho^5$.

For $n \neq 0$, the prescription of the quasiclassical calculation of the expression (90) reduces to

(a) the transition to the Euclidean space (see the equations in Sec. 6);

(b) replacement of the Euclidean $\hat{A}^a_\mu(x)$ by the instanton field $\bar{A}^a_\mu(x - x_0)$ given by the formula (65). The singular gauge is used because the reduction formula (90) is valid only for rapidly decreasing fields $A^a_\mu(x)$. For the nonsingular gauge, the inverse 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 $\mathrm{d}\rho\rho^{-5}d(\rho)$. Thus, for the matrix element (90) we obtain

$$\langle n \text{ gluons } |\Delta L(x)|0\rangle = \frac{d\rho}{\rho^5}\mathrm{d}(\rho)e^{-ix\Sigma q_k}\prod_{k=1}^{n} \times\left[\int \mathrm{d}x_k e^{-iq_kx_k}(-q_k^2)\varepsilon^{a_k}_{\mu_k}\overline{A^{a_k}_{\mu_k}}(x_k)\right], \tag{91}$$

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:

$$\int \mathrm{d}x e^{-iqx}(-q^2)\bar{A}^a_\mu(x) = \frac{4\pi i}{g}\bar{\eta}_{a\mu\nu}q_\nu\rho^2 . \tag{92}$$

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

$$\Delta L(x) = \frac{d\rho}{\rho^5}d(\rho)\exp\left[-\frac{2\pi^2}{g}\rho^2\bar{\eta}^M_{a\mu\nu}G^a_{\mu\nu}(x)\right],$$
$$\bar{\eta}^M_{a\mu\nu} = \begin{cases} \bar{\eta}_{amn}, & \mu = m,\ \nu = n;\ m,n = 1,2,3, \\ i\bar{\eta}_{a4n}, & \mu = 0,\ \nu = n;\ n = 1,2,3, \end{cases} \tag{93}$$

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 (91) is taken.

The expression (93) for the interaction of an instanton with an external field was obtained for the first time by Callan, Dashen and Gross [11] 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 (93) 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:

$$\langle 0|\Delta L|0\rangle = \frac{\mathrm{d}\rho}{\rho^5}d_{\mathrm{eff}}(\rho)$$
$$= \frac{\mathrm{d}\rho}{\rho^5}d(\rho)\left[1 + \frac{\pi^3\rho^4}{(N^2-1)\alpha_s}\langle 0|G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle + O(\rho^6)\right], \tag{94}$$

where in the averaging we used the relation

$$\langle 0|G^a_{\mu\nu}G^{a'}_{\mu'\nu'}|0\rangle = \frac{\delta^{aa'}}{N^2-1}\cdot\frac{1}{12}(g_{\mu\mu'}g_{\nu\nu'} - g_{\mu\nu'}g_{\nu\mu'})\langle 0|G^b_{\alpha\beta}G^b_{\alpha\beta}|0\rangle . \tag{95}$$

Note that the coupling 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 $\alpha_s G^a_{\mu\nu}G^a_{\mu\nu}$.

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. If we accept a "canonic" [1] value of the gluon condensate,

$$\langle 0|\frac{\alpha_s}{\pi}G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle \approx 0.012\ \mathrm{GeV}^4 , \tag{96}$$

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

$$\frac{\pi^4\rho^4}{8\alpha_s^2(\rho)}\langle 0|\frac{\alpha_s}{\pi}G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle . \tag{97}$$

It reaches unity at a value of ρ equal to

$$\rho_{\mathrm{crit}} \approx \frac{1}{1.15\ \mathrm{GeV}} , \tag{98}$$

if for α_s we take $\alpha_s(\rho) = 2\pi/9\ln(1/\Lambda\rho)$ with $\Lambda = 100$ MeV. For $\rho = \rho_{\mathrm{crit}}$, the interaction of the instanton with the vacuum fields of the other fluctuations becomes 100% important. This ρ_{crit} is rather small.

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 the one used in the many-body theory and for some four-quark operators for which it can be verified to have an accuracy of the order of a few percent.

The factorization leads to the relation

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

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

$$d_{\mathrm{eff}}(\rho) = d(\rho)\exp\left[\frac{\pi^4\rho^4}{(N^2-1)\alpha_s^2(\rho)}\langle 0|\frac{\alpha_s}{\pi}G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle\right], \tag{99}$$

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

$$\frac{2\pi}{\alpha_s(\rho)} \to \frac{2\pi}{\alpha_s(\rho)}\left[1 - \frac{\pi^3\rho^4}{2(N^2-1)\alpha_s^2(\rho)}\langle 0|\frac{\alpha_s}{\pi}(G^a_{\mu\nu})^2|0\rangle\right]. \tag{100}$$

Using for $d(\rho)$ the expression (99), we can advance in ρ to $\rho > \rho_{\rm crit}$. However, when the interaction with the vacuum fields changes the classical action strongly, i.e. when (100) 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.

10.2. *Instanton–Antiinstanton Interactions*

In the first part of the lecture devoted to the quantum-mechanical problem of the double well potential it was explained that the instanton contribution to the vacuum energy is determined by a chain of instantons. The chain was treated as an ensemble of well-separated instantons unaffected by the presence of others. Similar approximation in QCD is called the dilute instanton gas [11]. The instanton gas is not a good approximation in QCD, even at the qualitative level. In the absence of the instanton interactions, there is no way one can make the instantons gas self-consistent. Thus, the situation differs drastically from that we encountered in the quantum-mechanical example discussed in the beginning of this lecture. The basic difference is due to the fact that an intrinsic mass scale controlling the gas "diluteness" appears already at the classical level in the quantum-mechanical problem, while QCD at the classical level has no dimensional parameters. The only mass scale, $\Lambda_{\rm QCD}$, emerges as a result of the dimensional transmutation, at the quantum level. It is not surprising then, that the gas model, with no intrinsic mass scale, turns out to be inadequate.

One could try to amend the instanton gas model, by including interactions of instantons between each other. If a classical (or semiclassical) interaction of instantons becomes important in QCD at such values of ρ where $8\pi^2/g^2(\rho)$ is still a large parameter, an instanton-based picture of the QCD vacuum could survive. A crude picture can then be formulated as follows. Instantons and antiinstantons of a relatively small size form an interacting liquid. The "atoms" of this liquid are instantons and antiinstantons. The "atoms" act as potential wells for the light quark propagating in this "medium" [41] (Fig. 11).

This heuristic picture serves as a basis for various versions of the instanton liquid model [41]. We will not go into details of this model for two reasons: (i) it is too closely related to the applications of the instanton calculus while in this lecture we are mostly preoccupied with the basics of the formalism *per se*; (ii) exhaustive reviews exist in the literature (e.g. Ref. [4]). Instead, we will dwell on a general element from which each model of this type begins. If

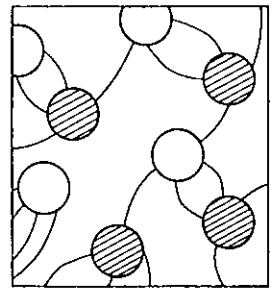


Fig. 11. A schematic picture of the instanton liquid model of the QCD vacuum. The shaded objects are antiinstantons. The lines connecting instantons (antiinstantons) denote interactions and quark exchanges.

one deals with more than one instanton the first issue to be addressed is the question of the instanton–antiinstanton (I–A) interactions at large separations. We speak about the instanton–antiinstanton interaction since two instantons, or two antiinstantons, being the exact solution of the duality equations, do not interact classically — the action of two pseudoparticles is twice the action of one.

The master formula for the instanton of size ρ centered at x "living" in an external background field $G^a_{\mu\nu}(x)$, we have obtained above, will allow us to easily find the leading term in the I–A interaction at large distances, the so-called dipole–dipole interaction.

No assumption is made in Eq. (93) as to the nature of the background field. In particular, this field can be caused by a distant antiinstanton of a larger size. If we substitute in Eq. (93) the value of the gluon field strength tensor induced by the antiinstanton centered at y we will get a formula describing the instanton–antiinstanton (I–A) interaction at a large separation. Since the instantons belong to the Euclidean space, strictly speaking we need a Euclidean analog of Eq. (93). It is quite obvious that the corresponding expression takes the form

$$\mathcal{L}_\rho(x_0) \sim \exp\left(-\frac{2\pi^2\rho^2}{g} O^{ab}\bar{\eta}_{b\mu\nu} G^a_{\mu\nu}(x_0)\right) + \text{h.c.}\,, \qquad (101)$$

where O^{ab} is a global color rotation matrix reflecting the relative orientations of the I–A pair in the color space, (it was not explicitly written in Eq. (93) since it was irrelevant for Sec. 10.1), x_0 is the coordinate of the instanton center, the subscript ρ reminds us of its radius, and $\bar{\eta}_{b\mu\nu}$ is the 't Hooft symbol. $G^a_{\mu\nu}(x_0)$ is the operator of the gluon field strength tensor. In principle, beyond the

leading approximation, the exponent in Eq. (101) will contain other operators, say with derivatives $D_\alpha G_{\mu\nu}$ or with two or more G's, along with a series in g.

In this way one can determine the I–A interaction as a systematic double expansion, in the ratio ρ/R and in the coupling constant, where R is the distance between instanton and antiinstanton. Thus, the leading dipole–dipole term is obtained from Eq. (101) by substituting the operator $G^a_{\mu\nu}(x_0)$ by the antiinstanton field at x_0. The antiinstanton centered at y_0 should be taken in the singular gauge, see Eq. (65) (where $\bar\eta$ must be substituted by η). It is assumed that $R = |x_0 - y_0| \gg \rho$, of course.

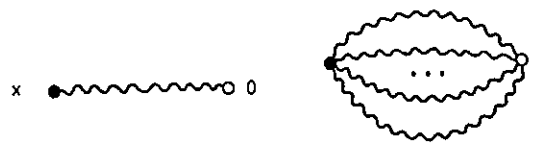
This exercise is pretty straightforward. For pedagogical reasons we find it more instructive to present a somewhat different (and less known) derivation of the I–A interaction that does not use the language of the classical fields at all. It operates with amplitudes of particle emission and allows us to connect the classical problem of the I–A interaction energy with the quantum problem of the instanton-induced cross-sections.

To illustrate how it works we will do the calculation of the I–A interaction in the leading (dipole) approximation. Relevant graphs are depicted in Fig. 12. The instanton with size ρ_1 is placed at x and the antiinstanton with size ρ_2 at the origin; $|x| \gg \rho_{1,2}$ so that the whole approach makes sense.

Figure 12(a) is the basic element of the calculation. We expand the exponent in Eq. (101) and a similar one for the antiinstanton, keep the linear in $G^a_{\mu\nu}$ terms and contract $G(x)$ and $G(0)$ to get

$$\left(\frac{4\pi^2}{g^2}\rho_1^2\rho_2^2\right) O_I^{ab}\bar\eta_{b\mu\nu}O_A^{cd}\eta_{d\alpha\beta}\langle G^a_{\mu\nu}(x)G^c_{\alpha\beta}(0)\rangle\,,$$

where $\langle G^a_{\mu\nu}(x)G^c_{\alpha\beta}(0)\rangle$ is the free Green function of the gauge field. Moreover, in the Green function $\langle A_\mu(x)A_\nu(0)\rangle$ we can only retain the $g_{\mu\nu}$ part, since



(a) One-gluon exchange (b) Multigluon exchange

Fig. 12. The I–A interaction from the instanton-induced effective Lagrangian (101). The instanton (•) is at the point x while the antiinstanton (o) is at the origin. The vertices in diagrams (a) and (b) are generated by expanding the exponent in Eq. (101) and keeping only the linear part in each of the $G_{\mu\nu}$ operators appearing in the expansion.

the part $x_\mu x_\nu$ drops out,

$$\langle G^a_{\mu\nu}(x)G^c_{\alpha\beta}(0)\rangle = \frac{2\delta^{ab}}{\pi^2}(g_{\nu\alpha}x_\mu x_\beta + g_{\mu\beta}x_\nu x_\alpha - g_{\nu\beta}x_\mu x_\alpha - g_{\mu\alpha}x_\nu x_\beta)\frac{1}{x^6} + \cdots\,, \tag{102}$$

where the dots denote the terms which are not contributed due to the fact $\eta_{a\mu\nu}\bar\eta_{b\mu\nu} = 0$.

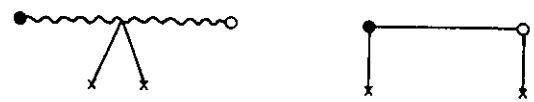
Now, it is not difficult to see that Fig. 12(b) just exponentiate the result $[1/(n!)^2$ from the expansion of the effective Lagrangians is supplemented by $n!$ coming from combinatorics] and we finally get for the I–A amplitude

$$\exp\left[-\frac{8\pi^2}{g^2} - \frac{32\pi^2}{g^2}\rho_1^2\rho_2^2\eta_{a\lambda\mu}\bar\eta_{b\lambda\nu}\Omega^{ab}\frac{x_\mu x_\nu}{x^6}\right], \tag{103}$$

where Ω^{ab} is the matrix of the relative orientation of the pseudoparticles, $\Omega^{ab} = O_I^{cb}O_A^{ca}$.

The x^{-4} term above is the dipole–dipole interaction sought for [42]. The procedure can be continued further [43]. The very same term in Eq. (101) generates, through Fig. 13, a part $\propto g^{-2}R^{-6}$ (the graph 13(a) is due to the quadratic term in $G^a_{\mu\nu}$ in one of the vertices). If the Higgs field is included, it generates the terms of order v^2R^{-2}, v^2R^{-4}, etc. through the graphs depicted in Fig. 14. The last term which has been explicitly computed is of order $g^{-2}R^{-8}\ln R$ corresponding to Fig. 15.

Fig. 13. Higher order terms in ρ/R in the I–A interaction due to the quantum corrections in the gluon exchanges. The R^{-6} term is generated by the same effective Lagrangian (101). The vertices attached to the instanton (•) in diagrams (a) and (b) and to the antiinstanton (o) on diagram (b) are generated by expanding the exponent in Eq. (101) and keeping the quadratic in A_μ part in $G_{\mu\nu}$.



(a) Mass term of the gauge boson (b) Higgs exchange

Fig. 14. Higher order terms in the I–A interaction due to the Higgs field. The crosses denote the vacuum expectation value of the Higgs field.

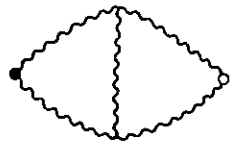


Fig. 15. An example of the two-loop diagram contributing to the I–A interaction at the level $g^{-2}R^{-8}\ln R$.

11. Constrained Instantons and the Valley Method

In the several problems discussed above we dealt with the field configurations that are *not* the exact solution of the classical equations of motion. The instanton in the theories with the spontaneously broken gauge symmetry (the Higgs regime) is an important example. Another example is the ensemble of the I–A "atoms": each given pseudoparticle in the ensemble is affected by the presence of others, and it is quite obvious that the topologically neutral gas cannot, strictly speaking, correspond to a minimum of the action and, hence, is not the exact solution. In other words the action of n pseudoparticles in the I–A ensemble $S_n \neq (8\pi^2 n/g^2)$; instead

$$S_n = \frac{8\pi^2}{g^2} n - S_{\rm int} \tag{104}$$

where generically the interaction "energy" $S_{\rm int} > 0$. (From now on we will use the terminology of a static problem in four-dimensional space; the action will be reinterpreted as the energy, and the word "pseudoparticle" becomes appropriate.)

The small-size instanton $\rho \ll m_W^{-1}$, is a legitimate contribution in the theory with the spontaneously broken gauge symmetry. Likewise, in pure gluodynamics well separated I–A pairs should be included in the partition function. It is clear that if the field configuration under consideration is very close to the exact solution (stationary point in the functional integral) it must be taken into account.

Thus, the question arises as to whether one can make quantitative the notion of proximity of the given field configuration to the exact solution and — if yes — what small parameter measures this proximity. A related crucial question is as follows: "If the selection criteria are relaxed and the functional integral is not represented exclusively by the stationary points of the classical action how far can one distance oneself from the exact solutions?"

It is very hard to answer these questions in a fully quantitative manner. The understanding existing in the current literature is heuristic, at best, and since we agreed to avoid controversial issues in this lecture, we will limit ourselves to a discussion of the approximate solutions when they are arbitrarily close to the exact ones; the question of their fate with worsening of the proximity parameter is left aside.

The approach allowing one to deal with the approximate solutions is usually referred to in the literature as *constrained instantons* [28]. Assume we have a field configuration $\phi(\beta)$ continuously depending on some parameter(s) β; ϕ is a generic notation for the set of all relevant fields. Assume that at $\beta \to \beta_0$ the field ϕ tends to become an exact solution. Let us call $\beta = \beta_0$ the limiting point. In the vicinity of the limiting point $\phi(\beta)$ is almost an exact solution. The fact that it is still not the exact solution means that there exist such deformations of ϕ that decrease the action. Typically such "decreasing" deformations are inherent to one — at most, several — directions in the functional space, call them destabilizing directions. Deformations along all other directions increase the action ("energy"). The basic idea of the constrained instantons is as follows. One introduces a constraint in the functional integration measure in such a way as to lock up all destabilizing directions. Then one minimizes the action subject to this constraint. Only those variations of ϕ are allowed which go in the directions perpendicular to the destabilizing ones. In this way one arrives at the constrained instanton. Dynamics in the destabilizing directions is studied separately.

To make a simple physical picture lying behind the program graphic, let us turn to the example of the instanton in the Higgs phase. At $\rho \to 0$ the BPST instanton becomes the exact solution. This is the limiting point which one can approach arbitrarily closely. If $\rho \neq 0$ there exists one direction in the functional space along which the action ("energy") slowly decreases. This direction corresponds to rescaling the instanton solution as a whole to smaller radii. By imposing a constraint we forbid the movement in the functional space along this direction. In the orthogonal subspace any deformation of the field configuration only increases the action, so it is possible to find one which minimizes the action, the constrained instanton. We then calculate the contribution of the constrained instanton in physically observable quantities. At the last stage we eliminate the constraint by integrating the result over all possible values of ρ.

It is worth noting that there is no unambiguous prescription as to how to choose the constraint. Usually, in each particular problem the most convenient and adequate choice is pretty clear from the physical context. For instance in the example above the appropriate constraint must fix the size of the instanton. The classical equations of motion are changed, of course, once the constraint is introduced. New equations do have a solution which at small x behaves like that of Belavin *et al.* while at large distances it decays exponentially, $\mathcal{O}(\exp(-m_W x))$ for the gauge field and $\mathcal{O}(\exp(-m_\chi x))$ for the Higgs field. The corresponding action becomes now ρ dependent and can be readily calculated in the form of expansion in $\rho^2 v^2$, see Sec. 8.6. It is important that the 't Hooft term is $\mathcal{O}(g^2)$ compared to the BPST term, so that the ρ dependence is indeed weak provided that $\rho \ll m_W^{-1}$.

(As a matter of fact, in Sec. 8.6 we managed to obtain the 't Hooft term without explicitly invoking the constrained instanton technique, a fact explainable by specific features of the instanton expression for the scalar field. The heavy artillery of the constrained instantons become relevant at the next-to-leading order.)

A version of the constrained instanton technique most often exploited in practice in connection with the approximate solutions, is the valley or streamline method [44] (for reviews referring to the Yang-Mills theory see e.g. Ref. [45]). The valley method is a variant of the constrained instanton approach, with a specific prescription as to how one should choose the constraints. It is most suitable to the problem of I–A pairs. Let us sketch a physical picture lying behind the valley method in the simplest example, one I–A pair.

We start from a pair of pseudoparticles at very large separations, large compared to their sizes which for simplicity are assumed to coincide for both pseudoparticles. This configuration — our boundary condition — is not an exact solution but is arbitrarily close to that. The field equations will experience a force tending to change the field in the direction of lowering of the total energy of two pseudoparticles. Let us do a *gedankenexperiment* — introduce an auxiliary fifth coordinate, "fifth time", and trace the evolution of the original configuration in the fifth time. In other words, our I and A atoms are set free to move as they wish at $t_5 = 0$.

To visualize the picture further it is instructive to discretize the fifth time. Then at step zero we have $\phi_0 = \phi_I^{R/2} + \phi_A^{-R/2}$ where the superscripts indicate the position of the I (A) centers.

At step one the field is deformed. The variation of ϕ is obtained in the following way. Let us try to change ϕ along different directions in the functional space. Almost every such attempt will lead to a higher energy, with an exception of, say, one direction[15] where the energy of the field configuration $\phi_1 = \phi_0 + \delta\phi_1$ is smaller than that of ϕ_0. At step two we take ϕ_1 as an input and repeat the procedure. In this way we get a chain of field configurations ϕ_i which becomes, in the limit of the continuous fifth time, a one-parametric family $\phi(\beta)$, the bottom of the valley.

In the quantitative terms the bottom of the valley is defined by the following requirement: as we move along the bottom $\phi(\beta)$ at each point the variation of the coordinate of the bottom must be proportional to the force at the given point,

$$\frac{\partial\phi(\beta)}{\partial\beta} \propto \frac{\delta S}{\delta\phi}\Big|_{\phi=\phi(\beta)} . \tag{105}$$

The proportionality coefficient is, generally speaking, a function of β sensitive to particular parametrizations of the bottom of the valley. (The coordinates along the bottom can be introduced in different ways.)

To get a clearer picture of the valley method it may be instructive to compare the infinitely-dimensional functional space to an analog mechanical motion of a stream (with a large friction) on two-dimensional surface with a trough (Fig. 16) in the gravitational field. The bottom of the valley in this case is a one-dimensional curve in the three-dimensional space, $\mathbf{x}(\beta)$. The force at

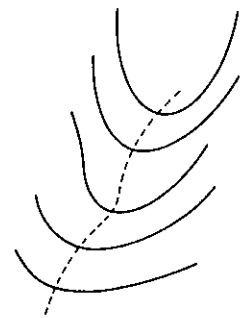


Fig. 16. Mechanical motion of a stream (---) on the bottom of a two-dimensional surface, a trough. The surface is steep in the direction perpendicular to the stream trajectory.

[15] It may well happen that the number of such directions is more than one; for clarity we confine ourselves to the simplest case.

the bottom is the gradient of the potential energy V, $F_i = -\partial_i V(\mathbf{x})|_{\mathbf{x}=\mathbf{x}(\beta)}$. The velocity of the stream at the given point at the bottom $\dot{\mathbf{x}}$ is proportional to the force,

$$\dot{\mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \beta}\dot{\beta} \propto -\partial_i V\,. \tag{106}$$

This analogy explains the origin of the name "streamline method".

If the beginning of the valley is a well-defined construction — and one can hardly doubt that — say, a well-separated I–A pair contributes to the observable effects, the question of where and how the valley ends up is rather obscure. Indeed, when the I and A "atoms" approach each other so that the defect of the action becomes comparable to $8\pi^2/g^2$ they tend to annihilate each other. Continuing the journey along the bottom of the valley we smoothly interpolate to a point where the field is weak, the action is $\mathcal{O}(1)$. This point, of course, belongs to ordinary perturbation theory. To avoid double-counting this part of the valley (a flat part) should be definitely excluded from consideration based on instantons. The quasiclassical approximation certainly fails here. So preferably we must stop earlier. Where exactly? Nobody knows. The collapse of the instanton-based approximations might manifest itself as an occurring phase transition when we approach a critical point from the "other" side of the valley (the one corresponding to large separations).

12. Fermions in Instanton Field

12.1. *Very Heavy Quarks*

In this section, we shall 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. 17. The first few terms of the effective Lagrangian that takes into account the fermion

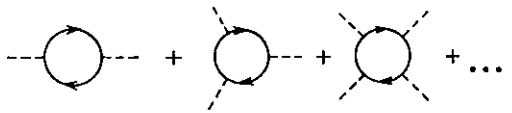


Fig. 17. A set of diagrams that one has to calculate to obtain the effective Lagrangian (107).

loops [46] are given below:

$$\Delta L_F = 2\,\mathrm{Tr}\left\{ -\frac{1}{4}G_{\mu\nu}^2 \times \frac{g^2}{24\pi^2}\ln\frac{M^2}{m^2} + \frac{1}{16\pi^2}\left(\frac{ig^3}{90m^2}G_{\mu\nu}G_{\nu\gamma}G_{\gamma\mu}\right.\right.$$
$$+ \frac{g^4}{72m^4}\left[-(G_{\mu\nu}G_{\mu\nu})^2 + \frac{7}{10}\{G_{\mu\gamma}, G_{\gamma\nu}\}_+^2 + \frac{29}{70}[G_{\mu\gamma}, G_{\gamma\nu}]_-^2\right.$$
$$\left.\left.\left. - \frac{8}{35}[G_{\mu\nu}, G_{\gamma\delta}]_-^2\right]\right)\right\}, \tag{107}$$

$$G_{\mu\nu} = G_{\mu\nu}^a T^a, \qquad \mathrm{Tr}\,T^aT^b = \frac{1}{2}\delta^{ab}\,.$$

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

The second and higher terms in (107) give a series in powers of $1/m^2\rho^2$ on the transition to the Euclidean space and the substitution of the instanton field for $G_{\mu\nu}$. Explicit expressions for all gluon operators in Eq. (107) in terms of the instanton size ρ can be found in Ref. [46].

12.2. *Light (Massless) Quarks*

We now turn to the limiting case of "light" quarks, $m\rho \ll 1$, whose impact on instantons is more radical. 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)]

$$S_F^{(E)} = \int d^4x\bar{\psi}(-i\gamma_\mu D_\mu - im)\psi\,, \tag{108}$$

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

$$\mathrm{Det}(-i\gamma_\mu D_\mu - im)\,. \tag{109}$$

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

$$\mathrm{Det}(-i\gamma_\mu D_\mu - im) = \prod_n(\lambda_n - im)\,, \tag{110}$$

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

$$-i\gamma_\mu D_\mu u_n(x) = \lambda_n u_n(x)\,. \tag{111}$$

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 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,

$$-i\gamma_\mu D_\mu u_0 = 0\,. \tag{112}$$

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

$$u_0 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}\chi_{\mathrm{L}} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}\chi_{\mathrm{R}}\,, \qquad \sigma_\mu^+ D_\mu \chi_{\mathrm{L}} = 0, \quad \sigma_\mu^- D_\mu \chi_{\mathrm{R}} = 0\,, \tag{113}$$

where $\sigma_\mu^\pm = (\boldsymbol{\sigma}, \mp i)$. To the equations for $\chi_{\mathrm{L}}, \chi_{\mathrm{R}}$ 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

$$-D_\mu^2 \chi_{\mathrm{L}} = 0, \qquad -D_\mu^2 \chi_{\mathrm{R}} = -4\boldsymbol{\sigma}\boldsymbol{\tau}\frac{\rho^2}{[(x-x_0)^2+\rho^2]^2}\chi_{\mathrm{R}}\,. \tag{114}$$

The operator $-D_\mu^2$ 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, $\chi_{\mathrm{L}} = 0$.

In the equation for χ_{R}, we use a basis in the space of spinor and color indices that diagonalizes the matrix $\boldsymbol{\sigma}\boldsymbol{\tau}$. We recall that $\boldsymbol{\sigma}$ acts on the spinor indices, and $\boldsymbol{\tau}$ on the color indices. This basis corresponds to the addition of the ordinary spin and the color spin to a total angular momentum equal to zero (when $\boldsymbol{\sigma}\boldsymbol{\tau} = -3$) or unity ($\boldsymbol{\sigma}\boldsymbol{\tau} = +1$). 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:

$$(\boldsymbol{\sigma}+\boldsymbol{\tau})\chi_{\mathrm{R}} = 0, \qquad \chi_{\mathrm{R}}^{\alpha m} \sim \varepsilon^{\alpha m}\,, \tag{115}$$

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 mode $u_0(x - x_0)$ (normalized by the condition $\int u^+ u dx = 1$) has the form

$$u_0(x) = \frac{1}{\pi}\frac{\rho}{(x^2+\rho^2)^{3/2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\varphi, \qquad \varphi^{\alpha m} = \frac{1}{\sqrt{2}}\varepsilon^{\alpha m}\,. \tag{116}$$

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

$$u_0^{\mathrm{sing}}(x) = \frac{1}{\pi}\frac{\rho}{(x^2+\rho^2)^{3/2}}\frac{x_\mu\gamma_\mu}{\sqrt{x^2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}\varphi\,, \tag{117}$$

which is obtained by the multiplication of (116) by the gauge transformation matrix (64′).

12.3. *Tunneling Interpretation in the Presence of Massless Fermions. The Index Theorem*

Since the instanton contribution is proportional to $\det(-i\gamma_\mu D_\mu - im)$, and the operator $i\gamma_\mu D_\mu$ has a zero mode in the instanton field, it is tempting to conclude that in the massless limit the instanton contribution vanishes. How can one reconcile this result with the tunneling interpretation?

Introduction of fermions certainly does not affect the nontrivial topology in the space of the gauge fields. The existence of a noncontractable loop remains intact, and with this loop comes the necessity of considering the wave function of the Bloch type (Sec. 5). The instanton trajectory connects Ψ_n and Ψ_{n+1} under the barrier, and is related to the probability of tunneling. If this probability were to vanish, what could have gone wrong with this picture?

To answer this question we must expand the picture of "tunneling in the $\mathcal{K}$ direction" by coupling the variable $\mathcal{K}$ to (an infinite number of) the fermion degrees of freedom. In order to make the situation more transparent we will slightly distort some details. We will assume that the motion of the system in the $\mathcal{K}$ direction is slow, while the fermion degrees of freedom are fast, so that the approximation of the Born-Oppenheimer type is applicable. In this approximation the motion in the $\mathcal{K}$ direction is treated adiabatically. We first freeze $\mathcal{K}$, then consider the dynamics of the fermion degrees of freedom, integrate them out, and at the last stage return to the evolution of the variable

$\mathcal{K}$. Certainly, in QCD all degrees of freedom are equally fast and no Born-Oppenheimer approximation can be developed. The general feature of the underlying dynamics that we are interested in does not depend, however, on this approximation.

Thus for each given value of $\mathcal{K}$ we must determine the fermion component of the wave function. This is done by building the Dirac sea in the fermion sector, with $\mathcal{K}$ frozen. The structure of the Dirac sea depends on the value of $\mathcal{K}$.

When $\mathcal{K}$ varies adiabatically, the energy of the fermion levels continuously evolves. The points $\mathcal{K} = n$ and $\mathcal{K} = n + 1$, being the gauge copies of each other, are physically identical. This means that the set of the energy levels of the Dirac sea at $\mathcal{K} = n$ is identical to the set at $\mathcal{K} = n + 1$.

This does not mean, however, that the individual levels do not move. When $\mathcal{K} =$ changes by one unit, some fermion levels with positive energy can dive into the negative-energy sea, while those from the sea, with the negative energies, can appear at levels with positive energies. As a whole the set will be intact, but, some levels interchange their positions (Fig. 18).

For each value of $\mathcal{K}$ we build the Dirac sea by filling in all negative-energy states. Let us say at $\mathcal{K} = n$ we built it properly. If in the process of motion

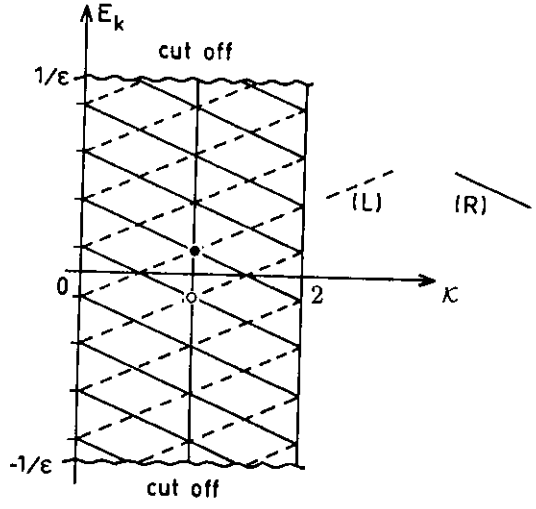


Fig. 18. The fermion energy levels *versus* $\mathcal{K}$.

in the $\mathcal{K}$ direction, at $\mathcal{K} = n + 1/2$, say, one level dives into the sea and one jumps out, this must be interpreted as the fermion production, since the state we end up with at $\mathcal{K} = n + 1$ is an excited state with respect to the proper Dirac sea at $\mathcal{K} = n + 1$. Thus, the tunneling trajectory connects the states $\Psi_n\Phi_n$(ferm) with $\Psi_{n+1}\Phi_{n+1}$(ferm) where the fermion components Φ_n(ferm) and Φ_{n+1}(ferm) differ by the quantum numbers of the fermion sector. We tried to calculate the probability of the tunneling transition with no change of the fermion state, and got zero. We now understand that we should not be discouraged. This zero could have been expected. The tunnelings occur in such a way that the fermion quantum numbers are forced to change in the process of the tunneling.

The consideration presented above is exact in the two-dimensional Schwinger model (spinor electrodynamics), see the review paper [47] for a pedagogical discussion. As was noted, in QCD a qualitatively similar picture is believed to take place. The argument demonstrating the validity of this picture in QCD is based on the so-called triangle anomaly. Assume for simplicity that we have one massless quark, q. At the classical level both the vector and axial currents

$$V_\mu = \bar{q}\gamma_\mu q, \qquad A_\mu = \bar{q}\gamma_\mu\gamma_5 q$$

are conserved,

$$\partial_\mu V_\mu = 0, \qquad \partial_\mu A_\mu = 0\,.$$

At the quantum level the axial current is anomalous,

$$\partial_\mu A_\mu = \frac{g^2}{16\pi^2} G^a_{\mu\nu}\tilde{G}^a_{\mu\nu}\,.$$

Let us now integrate over x and evaluate both sides of this equation in the instanton field. On the left-hand side we first integrate over the spatial variables $\mathbf{x}$. Then, the left-hand side reduces to

$$\int_{-\infty}^{\infty} dt\partial_0 \int A_0 d^3x = Q_5(t=\infty) - Q_5(t=-\infty)\,. \tag{118}$$

The right-hand side is

$$\frac{g^2}{16\pi^2}\int d^4x (G^a_{\mu\nu}\tilde{G}^a_{\mu\nu})_{\text{inst.}} = 2Q = 2(\mathcal{K}(t=\infty) - \mathcal{K}(t=-\infty))\,.$$

We see that in theory with one massless quark in the instanton transition the chiral charge is forced to change by two units — say, a left-handed quark is converted into a right-handed quark with the unit probability. If we want to get a nonvanishing tunneling probability we have to incorporate this feature. This will be done in the next section.

The variation of the chiral charge, $\Delta Q_5 \neq 0$, in the tunneling transition is in one-to-one correspondence with the occurrence of the zero modes in the Dirac equation in the self-dual fields. The number of the fermion zero modes is related to the topological charge of the gauge field by the so-called Atiyah-Singer (or the index) theorem [48], which has been derived in the instanton context in Ref. [49] (see also Refs. [8, 50, 51]). Specifically, if the number of the normalizable zero modes of positive (negative) chirality is n_+ (n_-) then

$$n_+ - n_- = Q \tag{119}$$

for each Dirac fermion field Ψ in the fundamental representation (since the operator $i\not{D}$ is Hermitian, the equation $\not{D}\Psi = 0$ implies that the equation on $\bar{\Psi}$ has a zero-eigenvalue solution as well). A brief but illuminating discussion of the derivation of Eq. (119) can be found in the review paper of Coleman [8]. As a matter of fact, this theorem is equivalent to the consideration of the triangle anomaly in the axial-vector current presented above.

Let us note in passing that the presence of massless fermions, combined with the triangle anomaly in $\partial_\mu A_\mu$, results in another drastic consequence: the θ term becomes unobservable even if $\theta \neq 0$. Indeed, one can rewrite L_θ as

$$L_\theta = \frac{\theta}{2}\partial_\mu A_\mu\,,$$

i.e. a full derivative of the gauge invariant quantity. Such full derivatives drop out from the action. This is in sharp distinction with the full derivative of the Chern-Simons current, which, as we know, gives a nonvanishing contribution in the action once we switch on the instanton field. The Chern-Simons current is not gauge-invariant.

This argument implies that in the theory with light quarks all θ dependent effects must be proportional to the quark mass.

12.4. *Instanton Density in the Theory with Light Quarks*

To calculate the instanton density we must consider the vacuum-to-vacuum transition amplitude in the theory with light quarks. In addition to the factors

we have obtained in pure gluodynamics this amplitude acquires the following factor due to the fermion field (we introduce one such field for the time being)

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

where Det′ denotes the determinant without the zero mode and we take into account the regularization and also the normalization by perturbation theory. In all the nonzero modes, m is taken as 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. 9), this dependence must be such that the cut-off 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 mode, the second from the nonzero modes. Comparing the result with formula (81) for gluons, we see that the situation has been changed because of the anticommutativity of the fermion fields: the zero modes of the light quarks lead to the screening of the charge, and the nonzero modes to antiscreening.

In the ordinary perturbation theory, the decomposition 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

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

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 gluon vacuum polarization there is also an analogous relation between the spin part of the polarization and the zero 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 (magnetic) part gives the "antiscreening" logarithm, and the charge (color) part (together with the Higgs particles) the "screening" part.

As we already know, in the limit $m \to 0$ the vacuum-to-vacuum transition amplitude tends to zero, and the instanton fluctuation couples the "vacuum" with the state with the one left-handed quark and the one right-handed antiquark. Alternatively, one can say that the instanton fluctuation couples the state with the left-handed quark with that with the right-handed quark. (Note that if $m = 0$ the left-handed quark does not become the right-handed quark in any order of the perturbation theory.) Let us calculate this coupling. To this end we consider 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

$$\langle p'|p_T\rangle = -\int dx dx' e^{ip'x'-ipx}\bar{v}^m_\alpha(\not{p}')_{\alpha\gamma}\langle 0|T\{q^m_\gamma(x')\bar{q}^k_\beta(x)\}|0\rangle_{\rm ins}(\not{p})_{\beta\delta}v^k_\delta\,, \tag{122}$$

where $\bar{v}^m_\alpha$ and v^k_δ 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

$$\langle 0|T\{q^m_\gamma(x')q^k_\beta(x)\}|0\rangle_{\rm ins} \underset{x_0\to -ix_4}{\longrightarrow} \sum_n \frac{u^m_{(n)\gamma}(x')u^{+k}_{(n)\beta}(x)}{m+i\lambda_n}\langle 0|0_T\rangle_{\rm ins}\,. \tag{123}$$

In the limit $m \to 0$, the zero mode makes the main contribution, and (123) is finite at $m = 0$ (since $\langle 0|0_T\rangle_{\rm ins}$ contains $F \propto m$).

Using the explicit form (117) of the zero 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 the instanton fluctuation with scale ρ:

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

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 $\bar{\eta}_{a\mu\nu} \to O_{aa'}\bar{\eta}_{a'\mu\nu}$ (O is the matrix of rotations in the color space), which must be made in (124). Further details and explicit color factors for the gauge group SU(3) can be found in Ref. [52]. The quantity $d_0(\rho)$ differs from $d(\rho)$ in Eq. (85) in pure gluodynamics by the multiplication by the factor

$$\exp F\left[-\frac{1}{3}\ln 2 - \frac{17}{36} + \frac{1}{3}(\ln 2\pi + \gamma) + \frac{2}{\pi^2}\sum_{s=1}^{\infty}\frac{\ln s}{s^2}\right] = e^{0.292F}\,, \tag{125}$$

where F is the number of light fermions. This is for the 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 antiinstanton, ΔL is obtained from (124) by the substitution $\bar{\eta}_{a\mu\nu} \to \eta_{a\mu\nu}$, $q_{\rm L,R} \to q_{\rm R,L}$. 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 $m = 0$.

12.5. *Global Anomaly*

The effective Lagrangian in Eq. (124) presents a concise formula summarizing the effects due to light quarks in QCD. A thorough inspection of this Lagrangian leads one to a perplexing question. Indeed, let us assume that, instead of QCD, we consider an SU(2) theory with one massless left-handed Weyl fermion transforming as a doublet with respect to SU(2). Usually we deal with the Dirac fermions; one Dirac fermion is equivalent to two Weyl fermions. Now we want to consider a chiral theory. Before the advent of the instantons this theory was believed to be perfectly well-defined. It has no internal anomalies; moreover, in perturbation theory, order by order, one encounters no reasons to make the theory sick. And yet, this theory is pathological. The 't Hooft interaction helps us reveal the pathology.

Indeed, if we start building an effective Lagrangian analogous to Eq. (124) in the SU(2) theory with one massless left-handed Weyl fermion we will immediately discover that this Lagrangian must be *linear* in the fermion field. In the instanton transition for one Dirac fermion $\Delta Q_5 = 2$, but the Weyl fermion $= 1/2$ of the Dirac fermion, and hence $\Delta Q_5 = 1$!

It was obvious to many that something was unusual in this theory. The intuitive feeling of pathology was formalized by Witten who showed [53] that this theory is ill-defined because of the *global anomaly*. Such theory is mathematically inconsistent. It simply does not exist.

One of the possible proofs of the global anomaly is based on the fermion level restructuring in the instanton transition. The key elements are the following: (i) the vacuum-to-vacuum amplitude in the theory with one Weyl fermion is proportional to $\sqrt{\det(i\not{D})}$; (ii) Only one pair of the fermion levels exchange their positions when $\mathcal{K} = n$ goes in $\mathcal{K} = n+1$. For further details see Ref. [53].

13. Continuation in Euclidean Space in the Theories with Chiral Fermions and Supersymmetric Theories

Usually, the first step of every instanton practitioner is to rewrite the theory under consideration as a Euclidean theory. We have done this too in Sec. 6.

This is possible and convenient if the theory one deals with contains only Dirac fermions. For chiral fermions this is not possible. The reason lies in the fact that one cannot define for the chiral Fermi fields the operation of involution (complex conjugation) in the Euclidean space. For this reason one cannot formulate, in particular (unextended) supersymmetric theories (which always involve chiral fermions) in the Euclidean space. In slightly different terms, one cannot introduce the notion of the Majorana spinor in the Euclidean space. Therefore, the problem of the continuation is common to both chiral and supersymmetric theories.

If we want to consider the instanton effects in these theories we have to indicate how the Euclidean continuation is to be done. Here we would like to show that Euclidean continuation is not at all equivalent to constructing the Euclidean version of the theory. One can keep the theory in the Minkowski space-time, and instead of continuing the theory, analytically continue just the solutions of the equations of motion. The fields, the Lagrangian and the action remain Minkowskean. What is really needed for the instanton calculations is the saddle point method of computation of path integrals based on analytic continuation in t for the tunneling processes.

Our method of Euclidean continuation of functional integrals follows the works of Berezin [54]. Let us recall the basic procedure for constructing the functional integrals. If we have a *quantum* Hamiltonian given as a function of the coordinate operator $\hat{q}_i$ and momentum operator $\hat{p}_i$, $\hat{H}(\hat{q}_i, \hat{p}_i)$, then we can unambiguously build a representation for the evolution operator $\exp(-i\hat{H}t)$ in terms of the functional integral

$$\langle q_i^{(+)}|e^{-i\hat{H}T}|q_i^{(-)}\rangle = \int \prod_i Dp_i Dq_i \exp\left\{ i \int_{-T/2}^{T/2} dt \times \left[\sum p_i \dot{q}_i - H(p_i(t), q_i(t)) \right] \right\}, \qquad (126)$$

where the integration runs over all trajectories satisfying the following conditions

$$q_i\left(t = \pm\frac{1}{2}T\right) = q_i^{(\pm)}, \qquad p_i\left(t = \frac{1}{2}T\right) = p_i\left(t = -\frac{1}{2}T\right).$$

Quite an analogous representation is valid for Fermi systems with the only difference that the integration runs over anticommuting Grassmann parameters.

Now, continuation to the Euclidean space reduces to a single step, transition to an imaginary value of the parameter T, $T = -i\tau$. In other words, instead of $\exp(-iHT)$ we consider $\exp(-H\tau)$. The functional integral representing this operator is built in the same way as above,

$$\langle q_i^{(+)}|e^{-H\tau}|q_i^{(-)}\rangle = \int \prod_i Dp_i Dq_i \exp\left\{ - \int_{-\tau/2}^{+\tau/2} dt \right.$$
$$\left. \times \left[- i \sum_i p_i \dot{q}_i + H(p_i(t), q_i(t)) \right] \right\}. \qquad (127)$$

Performing integration over $Dp_i(t)$ we arrive at the ordinary expression for the Euclidean action. This is for the Bose fields. For the Fermi fields one should calculate the integral over $D\bar{\psi}$, the step usually avoided.

On the other hand, arranging the expression (127) in the Euclidean form is not necessary at all. Indeed, quantization of the theory in the Minkowski space determines the operator $\hat{H}(\hat{p}_i, \hat{q}_i)$ as well as all the conserved charges (commuting operators). The action of these charges on $\hat{p}_i$ and $\hat{q}_i$ fixes their transformation law, which, in turn, generates the corresponding symmetry transformations for integration variables in the expressions like (127), in particular, SUSY transformations (for more details see Chapter VI).

Algebraically, all expressions for the transformations differ from the original ones (defined in the Minkowski space) only by the substitution $x_0 = -ix_4$.

To elucidate our procedure [55] and notations let us consider the vector potential for the instanton solution (more exactly, antiinstanton)

$$A^{\alpha\gamma}_{\beta\dot{\beta}} = -\frac{i}{g}\frac{1}{x^2+\rho^2}(\delta^\alpha_\beta x^\gamma_{\dot{\beta}} + \delta^\gamma_\beta x^\alpha_{\dot{\beta}}). \qquad (128)$$

Here spinor indices are introduced running over 1, 2 both in the color and coordinate spaces. The connection with the ordinary color triplet index is given by the following relation

$$A^{\alpha\gamma} = A^a \left(\frac{\tau^a}{2}\right)^\alpha_\rho \varepsilon^{\gamma\rho}, \qquad \begin{pmatrix} \alpha,\gamma,\rho = 1,2 \\ a = 1,2,3 \end{pmatrix}, \qquad (129)$$

where $(\tau^a)^\alpha_\rho$ stand for the Pauli matrices, $\varepsilon^{\gamma\rho}$ is the antisymmetric Levi-Civita symbol, which plays the role of metric:

$$F^\alpha = \varepsilon^{\alpha\beta}F_\beta, \qquad F_\alpha = \varepsilon_{\alpha\beta}F^\beta, \qquad \varepsilon_{\alpha\beta} = -\varepsilon^{\alpha\beta},$$
$$\varepsilon^{12} = -\varepsilon^{21} = 1. \qquad (130)$$

The relation between the vector and spinor indices in the coordinate space is as follows

$$A_{\beta\dot{\beta}} = (\sigma_\mu)_{\beta\dot{\beta}} A_\nu g^{\mu\nu}, \qquad g_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1),$$
$$(\sigma^\mu)_{\beta\dot{\beta}} = (\delta_{\beta\dot{\beta}}, \boldsymbol{\sigma}_{\beta\dot{\beta}}). \qquad (131)$$

(It is instructive to compare the expression for the σ^μ matrices in Eq. (131) with their Euclidean analogs which are defined after Eq. (113).) Furthermore, in particular, this relation refers also to x_μ; explicitly

$$x_{1\dot{1}} = x^{2\dot{2}} = x_0 - x_3,$$
$$x_{1\dot{2}} = -x^{2\dot{1}} = -x_1 + ix_2,$$
$$x_{2\dot{2}} = x^{1\dot{1}} = x_0 + x_3, \qquad (132)$$
$$x_{2\dot{1}} = -x^{1\dot{2}} = -x_1 - ix_2.$$

All the above notations are obviously taken from the Minkowski space. Their Euclidean nature is revealed only in the fact that x_0 is imaginary, $x_0 = -ix_4$. Another compromise to the Euclidean notation, which is rather unnecessary though, is in the definition of x^2,

$$x^2 = -x_\mu x_\nu g^{\mu\nu} = -\frac{1}{2}x^{\alpha\dot{\alpha}}x_{\alpha\dot{\alpha}} = -x_0^2 + x^2 = x_4^2 + x^2. \qquad (133)$$

One can readily convince oneself that the expression (128) coincides with the standard form for the BPST instanton. To this end one should take into account, apart from the expressions given above, the relation between the Minkowskian and Euclidean fields, A_μ and $\hat{A}_\mu$

$$A_0 = i\hat{A}_4, \qquad A^m = -\hat{A}_m \qquad (m = 1,2,3), \qquad (134)$$

(the latter are used in the standard approach, see Sec. 6).

The solution of the Dirac equation $\not{D}\Psi = 0$ (the 't Hooft zero mode) has the following form:

$$\Psi^\alpha_\gamma \propto \delta^\alpha_\gamma \frac{\rho}{(x^2+\rho^2)^{3/2}} \qquad (135)$$

where the lower (undotted) subscript refers to the left-handed SU(2) subgroup of the Lorentz $O(4)$ while the upper one is the spinorial index of the color SU(2).

References

[1] M. Shifman, A. Vainshtein and V. Zakharov, *Nucl. Phys.* **B147** (1979) 385; 448 [Reprinted in *Vacuum Structure and QCD Sum Rules*, Ed. M. Shifman (North-Holland, 1992) p. 24].

[2] A. Polyakov, *Phys. Lett.* **59B** (1975) 82 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 19].

[3] A. Belavin, A. Polyakov, A. Schwarz and Yu. Tyupkin, *Phys. Lett.* **59B** (1975) 85 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 22].

[4] T. Schäfer and E. V. Shuryak, *Instantons in QCD*, *Rev. Mod. Phys.* **70** (1998) 323.

[5] A. Polyakov, *Nucl. Phys.* **B120** (1977) 429.

[6] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw Hill, New York, 1965).

[7] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, New York, 1980).

[8] S. Coleman, in *Aspects of Symmetry* (Cambridge University Press, 1985), p. 265.

[9] G. 't Hooft, *Phys. Rev.* **D14** (1976) 3432; (E) **D18** (1978) 2199.

[10] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Third Edition (Pergamon Press, New York, 1989).

[11] C. Callan, R. Dashen and D. Gross, *Phys. Rev.* **D17** (1978) 2717; *Phys. Rev.* **D19** (1979) 1826.

[12] J. F. Willemsen, *Phys. Rev.* **D20** (1979) 3292.

[13] A. S. Schwarz, *Topology for Physicists* (Springer, Berlin, 1994).

[14] A. P. Balachandran, *Skyrmions*, in *High Energy Physics 1985*, Proc. Yale Theoretical Advanced Study Institute on High Energy Physics, New Haven, Conn., June-July, 1985, Eds. M. J. Bowick and F. Gursey (World Scientific, Singapore, 1985), Vol. 1, p. 1.

[15] S. Flüge, *Practical Quantum Mechanics* (Springer, Berlin, 1971), Vol. 1, Problem 28; C. Kittel, *Quantum Theory of Solids* (Wiley & Sons, New York, 1963), Chap. 9; N. Ashcroft and N. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976), Chap. 8.

[16] R. Jackiw and C. Rebbi, *Phys. Rev. Lett.* **37** (1976) 172 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 25].

[17] C. Callan, R. Dashen and D. Gross, *Phys. Lett.* **B63** (1976) 334 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 29].

[18] K. M. Bitar and S.-J. Chang, *Phys. Rev.* **D17** (1978) 486.

[19] R. J. Crewther, P. Di Vecchia, G. Veneziano and E. Witten, *Phys. Lett.* **B88** (1979) 123; (E)**B91** (1980) 487.

[20] S. Weinberg, *Phys. Rev. Lett.* **40** (1978) 223; F. Wilczek, *Phys. Rev. Lett.* **40** (1978) 279.

[21] R. D. Peccei, *QCD, Strong CP and Axions*, hep-ph/9606475; M. Davier, *Axions: a Review*, in *Hadrons, Quarks and Gluons*, Proc. XXII Rencontre de Moriond, Ed. J. Tran Thanh Van (Editions Frontieres, 1987).

[22] M. Atiyah, N. Hitchin, V. Drinfeld and Yu. Manin, *Phys. Lett.* **65A** (1978) 185 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 133; V. G. Drinfeld and Yu. I. Manin, *Yad. Fiz.* **29** (1979) 1646 [*Sov. J. Nucl. Phys.* **29** (1979) 845].

[23] R. Jackiw, *Field Theoretic Investigations in Current Algebra*, Sec. 7, in *Current Algebra and Anomalies*, Eds. S. B. Treiman, R. Jackiw, B. Zumino and E. Witten (Princeton University Press, 1985), p. 81; P. Fayet and S. Ferrara, *Phys. Rep.* **32** (1977) 249.

[24] M. Shifman and A. Vainshtein, *Nucl. Phys.* **B362** (1991) 21 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 97].

[25] R. Jackiw and C. Rebbi, *Phys. Lett.* **B67** (1977) 189; C. Bernard, N. Christ, A. Guth and E. Weinberg, *Phys. Rev.* **D16** (1977) 2967 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 149; 153].

[26] M. P. Mattis, *Phys. Rep.* **214** (1992) 159; V. A. Rubakov and M. E. Shaposhnikov, *Usp. Fiz. Nauk* **166** (1996) 493 [*Phys. Usp.* **39** (1996) 461]; hep-ph/9603208.

[27] T. Banks and E. Rabinovici, *Nucl. Phys.* **B160** (1979) 349; E. Fradkin and S. Shenker, *Phys. Rev.* **D19** (1979) 3682.

[28] The idea of the constrained instanton was first put forward in Y. Frishman and S. Yankielowicz, *Phys. Rev.* **D19** (1979) 540. The method was later developed by I. Affleck, *Nucl. Phys.* **B191** (1981) 429 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 247].

[29] F. R. Klinkhamer and N. S. Manton, *Phys. Rev.* **D30** (1984) 2212.

[30] R. Dashen, B. Hasslacher and A. Neveu, *Phys. Rev.* **D10** (1974) 4138.

[31] L. G. Yaffe, *Phys. Rev.* **D40** (1989) 3463; F. Klinkhamer, *Sphalerons and Energy Barriers in the Weinberg–Salam Model*, in Proc. XXV Int. Conf. on High Energy Physics, Singapore, 1990, Ed. K. K. Phua and Y. Yamaguchi (World Scientific, Singapore, 1991), p. 913.

[32] A. Smilga, *Nucl. Phys.* **B459** (1996) 263.

[33] I. Khriplovich, *Yad. Fiz.* **10** (1969) 409 [*Sov. J. Nucl. Phys.* **10** (1970) 235].

[34] T. Appelquist, M. Dine and I. Muzinich, *Phys. Lett.* **B69** (1977) 231.

[35] V. Novikov *et al.*, *Phys. Rep.* **41** (1978) 1, Sec. 1.3.

[36] D. R. T. Jones, *Nucl. Phys.* **B75** (1974) 531; for a mini-review see I. Hinchliffe, *Phys. Rev.* **D54** (1996) 77.

[37] C. Bernard, *Phys. Rev.* **D19** (1979) 3013 [Reprinted in *Instantons in Gauge Theories*, Ed. M. Shifman (World Scientific, Singapore, 1994), p. 109].

[38] G. 't Hooft and M. Veltman, *Nucl. Phys.* **B44** (1972) 189; G. 't Hooft, *Nucl. Phys.* **B62** (1973) 444.

[39] W. A. Bardeen, A. J. Buras, D. W. Duke and T. Muta, *Phys. Rev.* **D18** (1978) 3998.

[40] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Nucl. Phys.* **B165** (1980) 45.

[41] The instanton liquid model gradually developed in the 1980s. The key works are E. V. Shuryak, *Nucl. Phys.* **B198** (1982) 83; D. Diakonov and V. Petrov, *Nucl. Phys.* **B245** (1984) 259; **B272** (1986) 457; E. V. Shuryak, *Nucl. Phys.* **B302** (1988) 559; 574; 599. For references to further works and an exhaustive review see Ref. [4].

[42] D. Förster, *Phys. Lett.* **B66** (1977) 279; C. Callan, R. Dashen and D. Gross, *Phys. Rev.* **D17** (1978) 2717.

[43] A. Yung, *Instanton-Induced Effective Lagrangian in the Gauge-Higgs Theory*, Preprint SISSA 181/90/EP, 1990, unpublished; D. Diakonov and M. Polyakov, *Nucl. Phys.* **B389** (1993) 109; I. Balitsky and A. Schäfer, *Nucl. Phys.* **B404** (1993) 639.

[44] I. Balitsky and A. Yung, *Phys. Lett.* **B168** (1986) 113; I. Balitsky and A. Yung, *Nucl. Phys.* **B274** (1986) 475.

[45] A. Yung, *Valley Method for Instanton Induced Effects in Quantum Field Theory*, in *1991 Summer School in High Energy Physics and Cosmology*, Proc. ICTP 1991 Summer School on High Energy Physics and Cosmology, Ed. E. Gava, K. Narain, S. Randjbar-Daemi, E. Sezgin and Q. Shafi (World Scientific, Singapore, 1992), p. 580; I. Balitsky and A. Schäfer, *Nucl. Phys.* **B404** (1993) 639.

[46] V. A. Novikov, M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Fortsch. Phys.* **32** (1985) 585.

[47] M. Shifman, *Phys. Rep.* **209** (1991) 341.

[48] M. Atiyah and I. Singer, *Ann. Math.* **87** (1968) 484; **93** (1971) 119.

[49] A. Schwarz, *Phys. Lett.* **B67** (1977) 172.

[50] L. Brown, R. Carlitz and C. Lee, *Phys. Rev.* **D16** (1977) 417.

[51] D. Friedan and P. Windey, *Nucl. Phys.* **B235** (1984) 395 [Reprinted in Reprint Volume "Supersymmetry", Ed. S. Ferrara (North-Holland/World Scientific, 1987), p. 572.

[52] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Nucl. Phys.* **B163** (1980) 46.

[53] E. Witten, *Phys. Lett.* **B117** (1982) 324 [Reprinted in *Current Algebra and Anomalies*, Ed. S. Treiman *et al.* (Princeton Univ. Press, 1985), p. 429].

[54] F. A. Berezin, *The Method of Second Quantization* (Academic Press, New York, 1966). V. A. Novikov, M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, *Nucl. Phys.* **B260** (1985) 157.

Recommended Literature

Reviews

S. Coleman, *Aspects of Symmetry* (Cambridge University Press, 1985), p. 265.

T. Schäfer and E. V. Shuryak, *Instantons in QCD*, *Rev. Mod. Phys.* **70** (1998) 323.

Books

Instantons in Gauge Theories, Ed. M. Shifman (World Scientific, Singapore, 1994). I bring my apology to E. F. Corrigan, D. B. Fairlie, S. Templeton and P. Goddard whose work (*Nucl. Phys.* **B140** (1987) 31) was inadvertently omitted from Sec. IV of the book. M.S.

R. Rajaraman, *Solitons and Instantons* (North-Holland, Amsterdam, 1982).