A Primer on Instantons in QCD

Hilmar Forkel
Institut für Theoretische Physik, Universität Heidelberg,
Philosophenweg 19, D-69120 Heidelberg, Germany.

August 2000

Contents

Abstract

These are the (twice) extended notes of a set of lectures given at the “12th Workshop on Hadronic Interactions” at the IF/UERJ, Rio de Janeiro (31. 5. - 2. 6. 2000). The lectures aim at introducing essential concepts of instanton physics, with emphasis on the role of instantons in generating tunneling amplitudes, vacuum structure, and the induced quark interactions associated with the axial anomaly. A few examples for the impact of instantons on the physics of hadrons are also mentioned.

Chapter 1

Introduction

Yang-Mills instantons were discovered a quarter of a century ago [1]. They have furnished the first explicit (and still paradigmatic) example of a genuinely nonperturbative gauge field configuration and display a wealth of unprecedented geometrical, topological and quantum effects with fundamental impact on both ground state and spectrum of nonabelian gauge theories. In fact, one could argue that the discovery of instantons marked the beginning of a new era in field theory. In the words of Sidney Coleman [2]: "In the last two years there have been astonishing developments in quantum field theory. We have obtained control over problems previously believed to be of insuperable difficulty and we have obtained deep and surprising (at least to me) insights into the structure of ... quantum chromodynamics." Instantons have remained an active, fascinating and highly diverse research area ever since, both in mathematics and physics.

The following pages contain the notes of a series of elementary lectures on instantons in quantum mechanics and QCD which aim at introducing some of the underlying ideas to a non-expert audience.

1.1 Motivation and overview

QCD instantons are mediating intriguing quantum processes which shape the ground state of the strong interactions. These localized and inherently nonperturbative processes can be thought of as an ongoing "rearrangement" of the vacuum. In addition, there exist many other types of instantons in various areas of physics. This became clear shortly after their discovery when it was realized that instantons are associated with the semiclassical description of tunneling processes.

Since tunneling processes abound in quantum theory, so do instantons. We will take advantage of this situation by explaining the physical role of instantons in the simplest possible dynamical setting, namely in the one-dimensional quantum mechanics of tunneling through a potential barrier. Since the classical aspects of the problem are particularly simple and familiar here, we can focus almost exclusively on the crucial quantum features and their transparent treatment. Once the underlying physics is understood, this example can be more or less directly generalized to the tunneling processes in the QCD vacuum, where we can then concentrate on the additional complexities brought in by the field theoretical and topological aspects of classical Yang-Mills instantons. We will end our brief tour of instanton physics with a short section on the impact of instantons on hadron phenomenology.

As already mentioned, instantons occur in many guises and we can only give a small glimpse of this vast field here. (The SLAC high-energy preprint library lists far beyond 1000 preprints on instantons in the last decade.) Moreover, we will spend a large part of the available time on laying a solid conceptual foundation and will consequently have less time for applications. To do at least some justice to the diversity of the subject and to give an idea of its scope, let us mention a few actively studied topics not touched upon in these lectures: instantons have been found in many field theories, ranging from scalar field theory via supersymmetric Yang-Mills and gravity to string (or M-) theory. There are interesting relations and interactions between instantons and their topological cousins, the non-abelian monopoles and vortices. In several theories, probably including QCD, instantons are responsible for spontaneous chiral symmetry breaking. The role of instantons in deep inelastic scattering and other hard QCD processes has been examined, and also their impact on weak-interaction processes at RHIC, LHC and beyond. In inflationary cosmology and elsewhere relatives of instantons (sometimes called "bounces") describe the "decay of the false vacuum".

Moreover, there are fascinating mathematical developments in which instantons serve as tools. They have been instrumental, e.g., in obtaining profound results in geometry and topology, including Donaldson theory of four-manifolds (which led to the discovery of infinitely many new differentiable structures on R^4). Instantons play a particularly important role in supersymmetric gauge, string and brane theories. They saturate, for example, the nonperturbative sector of the low-energy effective (Seiberg-Witten) action of $N = 2$ supersymmetric Yang-Mills theory.

Theoretical approaches to QCD instantons include well-developed "instanton-liquid" vacuum models [3], a variety of hadron models at least partially based on instanton-induced interactions (see, e.g. [4]), a sum-rule approach based on a generalized operator product expansion (IOPE) [5], and an increasing amount of lattice simulations [6].

There is a vast literature on instantons. In preparing these lectures I have benefitted particularly from the articles and books by Coleman [2], Kleinert [7], Schulman [8], Sakita [9], Polyakov [10], Vainshtein et al [11], and Zinn-Justin [12]. More advanced and technical material on instantons is collected in [13]. A pedagogical introduction to instantons in supersymmetric field theories can be found in [14].

The program of these lectures is as follows: first, we will discuss the semi-classical approximation (SCA) in quantum mechanics in the path integral representation. This will lead us to describe tunneling processes in imaginary time where we will encounter the simplest examples of instanton solutions. In the following we will turn to instantons in QCD, their topological properties, their role in generating the vacuum structure, and their impact on the physics of hadrons.

1.2 WKB reminder

As we have already mentioned, instantons mediate quantum-mechanical barrier penetration processes. The semiclassical approximation (SCA) is the method of choice for the treatment of such processes. It might therefore be useful to recapitulate some basic aspects of this technique, in the form established by Wenzel, Kramers and Brioullin (WKB).

The WKB method generates approximate solutions of the Schrödinger equation for wavefunctions with typical wavelength λ small in comparison to the spacial variations of the potential. This situation corresponds to the semiclassical limit $\hbar \to 0$ where

$$\lambda = \frac{2\pi\hbar}{p} \to 0. \tag{1.1}$$

Macroscopic structures, for example, behave normally according to classical mechanics because their wavefunctions vary extremely rapidly compared to the variations of the underlying potentials or any other length scale in the problem, even if of microscopic (e.g. atomic) origin. The semiclassical limit is therefore the analog of the geometrical (ray) limit of wave optics, where the scatterers are structureless at the scale of the typical wavelengths of light and where wave phenomena like diffraction disappear.

In order to set up the semiclassical approximation, the WKB method writes the wave function as

$$\psi(x) = e^{i\Phi(x)/\hbar}. \tag{1.2}$$

Since $\Phi \in C$ this ansatz is fully general. In scattering applications Φ is often called the "Eikonal". Inserting (1.2) into the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x) - E\right]\psi(x) = 0 \tag{1.3}$$

one obtains the WKB "master equation"

$$\frac{1}{2m}\Phi'^2(x) - \frac{i\hbar}{2m}\Phi''(x) = E - V(x). \tag{1.4}$$

This nonlinear differential equation can be solved iteratively. One expands

$$\Phi(x) = \Phi_0(x) + \hbar\Phi_1(x) + \hbar^2\Phi_2(x) + ..., \tag{1.5}$$

(the series is generally asymptotic) and equates terms of the same order of $\hbar$ on both sides. For each power of $\hbar$ one thereby obtains a corresponding WKB equation. The zeroth-order equation has the solution

$$\Phi_0(x) = \pm\int^x dx' p(x') \qquad \text{with} \qquad p(x) \equiv \sqrt{2m[E - V(x)]} \tag{1.6}$$

where $p(x)$ is the momentum of the particle in a constant potential $U_x = const$ whose value equals that of V at x. If $V(x)$ varies slowly compared to $\psi(x)$, ψ indeed experiences a locally constant U_x and the $O(\hbar^0)$ solution

$$\psi_0(x) = e^{i/\hbar\int^x dx'\sqrt{2m[E-V(x')]}} \tag{1.7}$$

becomes a useful approximation. By inserting it back into the Schrödinger equation (1.3),

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V(x) - E\right]\psi_0(x) = -\frac{i\hbar}{2}\frac{V'(x)\psi_0(x)}{\sqrt{2m[E-V(x)]}}, \tag{1.8}$$

we confirm that it is a solution up to the correction term on the right-hand side, which is indeed of $O(\hbar)$ and proportional to the variation V' of the potential.

The quality of the zeroth-order approximation can be gauged more systematically by checking whether the neglected term in (1.4) is small, i.e whether

$$\left|\frac{\frac{i\hbar}{2m}\Phi''}{\frac{1}{2m}\Phi'^2}\right| = \left|\hbar\frac{\Phi''}{\Phi'^2}\right| = \left|\frac{d}{dx}\frac{\hbar}{\Phi'}\right| \ll 1. \tag{1.9}$$

With $\lambda(x) \equiv 2\pi\hbar/p(x) = 2\pi\hbar/\Phi_0'$ and with $\Phi_0' \sim \Phi'$ (to leading order) this turns into

$$\frac{1}{2\pi}\left|\frac{d\lambda}{dx}\right| = \frac{1}{2\pi}\left|\frac{\frac{d\lambda}{dx}\cdot\lambda}{\lambda}\right| \equiv \frac{1}{2\pi}\left|\frac{\delta\lambda}{\lambda}\right| \ll 1 \tag{1.10}$$

where $\delta\lambda$ is the change of λ over the distance of a wavelength. From the above inequality we learn that the semiclassical expansion is applicable in spacial regions where the de-Broglie wavelength 1) is small compared to the typical variations of the potential (therefore highly excited states behave increasingly classical) and where it 2) changes little over distances of the order of the wavelength. (The latter condition alone is not sufficient since we have compared two terms in a differential equation, not in the solution itself.)

The semiclassical treatment of *tunneling* processes, which will be a recurrent theme throughout these lectures, involves a characteristic additional step since tunneling occurs in potentials with classically forbidden regions, i.e. regions $x \in [x_l, x_u]$ where $E < V(x)$. The boundaries x_l, x_u are the classical turning points. Inside the classically forbidden regions, $p(x)$ becomes imaginary and the solution (1.7) decays exponentially as[1]

$$\psi_{0,tunnel}(x) = e^{-1/\hbar\int_{x_l}^{x_u} dx'\sqrt{2m[V(x')-E]}}. \tag{1.11}$$

A comparison of the two solutions (1.7) and (1.11) shows that, formally, the meaning of allowed and forbidden regions can be interchanged by the replacement

$$t \to -i\tau \qquad \Rightarrow \qquad E \to iE, \qquad V \to iV \tag{1.12}$$

which indeed transforms (1.7) into the tunneling amplitude (1.11). In other words, we can calculate tunneling amplitudes in SCA by the standard WKB methods, analytically continued to imaginary time. Below we will see that this works the same way in the path-integral approach.

[1] Since Gamov's classic treatment of nuclear α-decay as a tunneling process through the Coulomb barrier of the nucleus this amplitude is also known as the "Gamov factor".

Although the WKB procedure is simple and intuitive in principle, it becomes increasingly involved in more than one dimension (except for special cases [15]) and at higher orders. The $O(\hbar^n)$, $n > 0$ corrections are calculated by matching the solutions in the classically allowed and forbidden regions at the classical turning points. This leads to Bohr-Sommerfeld quantization conditions for $p(x)$ when integrated over a period of oscillation in the allowed regions. However, the generalization of this procedure to many-body problems and especially to field theory is often impractical [16]. Fortunately, there is an alternative approach to the SCA, based on the path integral, which can be directly applied to field theories. With our later application to QCD in mind, we will now consider this approach in more detail.

Chapter 2

Instantons in quantum mechanics

2.1 SCA via path integrals

One of the major advantages of the path integral representation of quantum mechanics is that it most transparently embodies the transition to classical mechanics, i.e. the semiclassical limit. In the following chapter we will discuss this limit and the associated SCA in the simplest dynamical setting in which instantons play a role, namely in one-dimensional potential problems with one degree of freedom (a spinless point particle, say).

2.1.1 The propagator at real times: path integral and SCA

The key object in quantum mechanics, which contains all the physical information about the system under consideration (spectrum, wavefunctions etc.), is the matrix element of the time evolution operator

$$\left\langle x_f \left| e^{-iHT/\hbar} \right| x_i \right\rangle , \tag{2.1}$$

i.e. the probability amplitude for the particle to propagate from

$$x_i \qquad \text{at} \qquad t = -\frac{T}{2} \qquad \text{to} \qquad x_f \qquad \text{at} \qquad t = \frac{T}{2}. \tag{2.2}$$

(In quantum field theory this matrix element, which is sometimes called the quantum-mechanical propagator, generalizes to the generating functional.) Its path integral representation is

$$\left\langle x_f \left| e^{-iHT/\hbar} \right| x_i \right\rangle = \mathcal{N} \int D[x] e^{i\frac{S[x]}{\hbar}} = \mathcal{N} \int D[x]_{\{x(-T/2)=x_i \,|\, x(T/2)=x_f\}} e^{\frac{i}{\hbar}\int_{-T/2}^{T/2} dt \mathcal{L}(x,\dot{x})} \tag{2.3}$$

where

$$S\left[x\right] = \int_{-T/2}^{T/2} dt \mathcal{L}\left(x, \dot{x}\right) = \int_{-T/2}^{T/2} dt \left\{ \frac{m}{2} \dot{x}^2 \left(t\right) - V\left[x\left(t\right)\right] \right\} \tag{2.4}$$

is the classical action along a given path. The integration over the paths can be defined, for example, by discretizing the time coordinate (Trotter formula) into intervals Δt so that $t_n = n\Delta t$ and

$$D[x] := \lim_{N\to\infty} \left(\frac{mN}{2\pi i \hbar t} \right)^{1/2} \prod_{n=1}^{N-1} \left(\frac{mN}{2\pi i \hbar t} \right)^{1/2} dx_n \tag{2.5}$$

where $x_n = x\left(t_n\right)$. An alternative representation of $D[x]$ in terms of a complete set of functions will be introduced below.

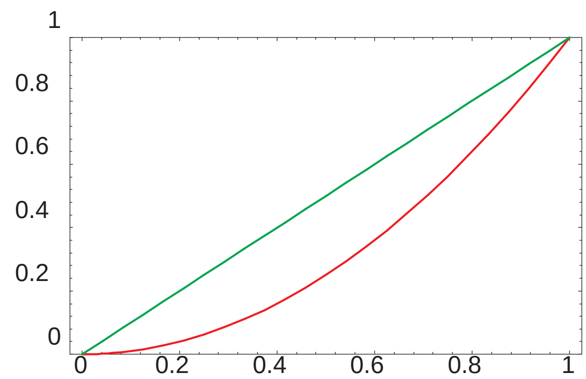

Figure 2.1: The two paths from $(x, t) = (0, 0)$ to $(1, 1)$ whose action is compared.

In the semiclassical limit, i.e. for $\hbar \to 0$, the classical action can become much larger than $\hbar$. As a consequence, the path integral is dominated by the paths in the vicinity of the stationary point(s) $x_{cl}(t)$ of the action (if they exist), which satisfy

$$\frac{-\delta}{\delta x(t)} S[x] = m\ddot{x} + \frac{\partial V}{\partial x} = 0 \tag{2.6}$$

with the boundary conditions

$$x_{cl}\left(-\frac{T}{2}\right) = x_i, \qquad x_{cl}\left(\frac{T}{2}\right) = x_f. \tag{2.7}$$

These classical paths are important not because they themselves give dominant contributions to the integral (in fact, their contribution vanishes since the set of classical paths is of measure zero) but rather because the action of the neighboring paths varies the least around them. Thus, infinitely many neighboring paths, which lie in the so-called "coherence region" with similar phase factors $\exp(iS/\hbar)$, add coherently. For the paths outside of the coherence region the phases vary so rapidly that contributions from neighboring paths interfere destructively and become irrelevant to the path integral. This exhibits in a transparent and intuitive way the relevance of the classical paths in quantum mechanics.

To get a more quantitative idea of the coherence region, let us approximately define it as the set of paths whose phases differ by less than π from the phase of the classical path (the "stationary phase"), which implies

$$\delta S[x] = S[x] - S[x_{cl}] \leq \pi\hbar. \tag{2.8}$$

Thus, for a macroscopic particle with

$$S \simeq 1 \text{ erg sec } = 1 \frac{\text{g cm}^2}{\text{sec}^2} \text{sec} \simeq 10^{27}\hbar \tag{2.9}$$

only an extremely small neighborhood of the classical path contributes, since $\delta\phi = \delta S/\hbar$ gets very sensitive to variations of the path. A numerical example (borrowed from Shankar [17]) makes this more explicit: compare two alternative paths from $(x, t) = (0, 0)$ to $(x, t) = (1\,\text{cm}, 1\,\text{s})$ (see Fig. 2.1) for a noninteracting particle: the classical path $x_{cl}(t) = t$ with $S[x_{cl}] = \int_0^T dt\,(m/2)\,v[x_{cl}]^2 = m\,\text{cm}^2/\left(2\,\text{sec}^2\right)$ and the alternative path $x_{alt}(t) = t^2$ with $S[x_{alt}] = 2m\,\text{cm}^2/\left(3\,\text{sec}^2\right)$. For a classical particle with

$$m = 1g \qquad \Rightarrow \delta S = S[x_{alt}] - S[x_{cl}] = \frac{m\ \text{cm}^2}{6\,\text{sec}^2} \simeq 1.6 \times 10^{26}\hbar \qquad \Rightarrow \delta\phi \simeq 1.6 \times 10^{26}\,\text{rad} \ggg \pi \tag{2.10}$$

the alternative path is extremely incoherent and therefore irrelevant, while for an electron with

$$m = 10^{-27}g \qquad \Rightarrow \delta S \simeq \frac{1}{6}\hbar \qquad \Rightarrow \delta\phi \simeq \frac{1}{6}\,\text{rad} \ll \pi \tag{2.11}$$

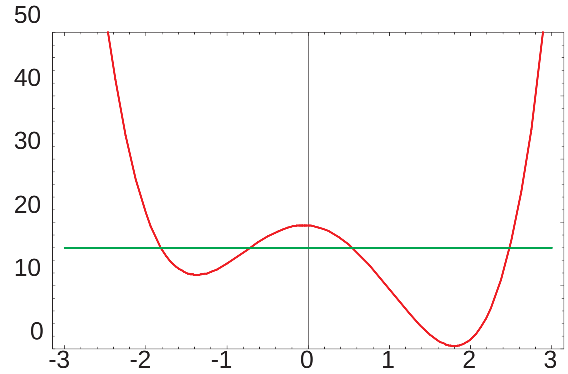

Figure 2.2: A typical tunneling potential with nondegenerate minima. The total energy (horizontal line) is smaller than the hump so that a classically forbidden region exists.

it is well within the coherence region and makes an important contribution to the path integral!

Obviously, then, the movement of a free electron (even a very fast one) cannot be described classically. One has to resort to quantum mechanics where the electron's path is much more uncertain and, in fact, not an observable. There also exist microscopic situations, on the other hand, where the quantum fluctuations do not totally wash out the classical results and which can therefore be treated semiclassically. Typical examples are the scattering off a slowly varying potential (Eikonal approximation) or the highly excited electronic orbits in atoms (Rydberg atoms).

However - and this is a crucial point for our subsequent discussion - the stationary-phase approximation fails to describe tunneling processes! The reason is that such processes are characterized by potentials with a classically forbidden region (or barrier) which cannot be transgressed by classical particles. In other words, there exist no classical solutions of (2.6) with boundary conditions corresponding to barrier penetration in a potential of the type shown in Fig. 2.2. As a consequence, the action $S[x]$ has no extrema with tunneling boundary conditions, and therefore the path integral (2.3) has no stationary points.

Does this mean that the SCA becomes unfeasible for tunneling processes? Fortunately not, of course, as the standard WKB approximation shows. The problem is particular to the stationary-phase approximation in the path integral framework, which indeed ceases to exist. The most direct way to overcome this problem would be to generalize the familiar saddle-point approximation for ordinary integrals. This method works by deforming the integration path into the complex plane in such a way that it passes through the saddle points of the exponent in the integrand. And since there indeed exist complex solutions of the combined equations (2.6), (2.7) in tunneling potentials (in other words, the complex action still obeys the Hamilton-Jacobi equation), a generalization of the complex saddle-point method to integrals over complex paths would seem natural. But unfortunately integration over complex paths has not been sufficiently developed (although a combined approach to both diffraction and ray optics in this spirit was initiated by Keller [18], and McLaughlin [19] has performed a complex saddle point evaluation of the Fourier-transformed path integral. See also the recent work of Weingarten [20] on complex path integrals.).

However, in many instances - including our tunneling problem - an analytical continuation of the path integral to imaginary time, as already encountered in the WKB framework in (1.12), serves the same purposes. This is the way in which we will perform the SCA to tunneling problems in the following sections.

2.1.2 Propagator in imaginary time

As just mentioned, the paths which form the basis of the SCA to tunneling processes in the path-integral framework can be identified by analytical continuation to imaginary time. This approach was apparently first used by Langer [22] in his study of bubble formation processes at first-order phase transitions[1]. In the context of the SCA, tunneling problems in field theory, and instantons it was introduced by Polyakov et al. [2, 24], building on a suggestions by Gribov.

In order to prepare for the use of this approach, let us first see how the ground-state energy and wave-function in tunneling situations can be obtained from the matrix element

$$Z(x_f, x_i) = \left\langle x_f \left| e^{-HT_E/\hbar} \right| x_i \right\rangle \tag{2.12}$$

which is obtained from (2.1) by analytically continuing

$$t \to -i\tau \qquad (\Rightarrow T \to -iT_E). \tag{2.13}$$

This procedure (together with its counterpart $p_0 \to ip_4$ in momentum space) is often called a "Wick rotation". (Note, incidentally, that imaginary-"time" evolution as generated by $e^{-HT_E/\hbar}$ is not unitary and therefore does not conserve probability. It might also be worthwhile to recall that the matrix element (2.12) plays a fundamental role in statistical mechanics (its trace over state space, corresponding to the sum over all periodic paths, is the partition function) although this is not the angle from which we will look at it in the following.)

The matrix element $Z(x_f, x_i)$ has the energy representation (in the following we drop the subscript E of T_E)

$$Z(x_f, x_i) = \sum_n e^{-E_n T/\hbar} \langle x_f | n \rangle \langle n | x_i \rangle \tag{2.14}$$

in terms of the spectrum

$$H|n\rangle = E_n |n\rangle, \qquad 1 = \sum_n |n\rangle\langle n| \tag{2.15}$$

of the (static) Hamiltonian H. The energies E_n are the usual real eigenvalues, which remain unaffected by the analytical continuation. For large T the ground state dominates,

$$Z(x_f, x_i) \to e^{-E_0 T/\hbar} \langle x_f | 0 \rangle \langle 0 | x_i \rangle, \tag{2.16}$$

and the ground state energy becomes

$$E_0 = -\hbar \lim_{T\to\infty} \frac{1}{T} \ln Z(x_f, x_i). \tag{2.17}$$

In order to calculate the ground state energy (and wave function) of a quantum-mechnical system we therefore just need to take the $T \to \infty$ limit of the imaginary-time matrix element (2.12). This convenient way to obtain ground-state properties is used, e.g., in Euclidean lattice QCD to calculate hadron masses. In the following section we will show how to calculate the matrix element $Z(x_f, x_i)$ semiclassically in the path integral framework.

2.1.3 Path integral in imaginary time: tunneling in SCA

The quantum mechanical propagator (2.12) in imaginary time has a path integral representation which can be obtained from (2.3) by analytical continuation:

$$Z(x_f, x_i) = \mathcal{N} \int D[x] e^{-\frac{S_E[x]}{\hbar}} = \mathcal{N} \int D[x]_{\{x(-T/2)=x_i \,|\, x(T/2)=x_f\}} e^{-\frac{1}{\hbar}\int_{-T/2}^{T/2} d\tau \mathcal{L}_E(x,\dot{x})}. \tag{2.18}$$

(Note, by the way, that this path integral is much better behaved than its real-time counterpart (and thus can be formalized as an integral over generalized stochastic Wiener processes) since the oscillating integrand is replaced by an exponentially decaying one.)

[1] Another early application of similar ideas is due to W.A. Miller [23]. (I thank Takeshi Kodama for drawing my attention to this reference.)

For future applications it is useful to define the "measure" of this integral more explicitly. To this end, we expand $x(\tau)$ into a complete, orthonormal set of real functions $\tilde{x}_n(\tau)$ around a fixed path $\bar{x}(\tau)$ as

$$x(\tau) = \bar{x}(\tau) + \eta(\tau) \qquad \text{where} \qquad \eta(\tau) = \sum_{n=0}^{\infty} c_c \tilde{x}_n(\tau) \tag{2.19}$$

and

$$\int_{-T/2}^{T/2} d\tau \tilde{x}_n(\tau)\tilde{x}_m(\tau) = \delta_{mn}, \qquad \sum_n \tilde{x}_n(\tau)\tilde{x}_n(\tau') = \delta(\tau - \tau'). \tag{2.20}$$

Furthermore, $\bar{x}$ is assumed to satisfy the boundary conditions implicit in the path integral, i.e.

$$\bar{x}(\pm T/2) = x_f/x_i, \qquad \tilde{x}_n(\pm T/2) = 0. \tag{2.21}$$

(Note that (for now) there are no further requirements on $\bar{x}$.) As a consequence we have

$$D[x] = D[\eta] = \prod_n \frac{dc_n}{\sqrt{2\pi\hbar}}, \tag{2.22}$$

where the normalization factor is chosen for later convenience.

Now let us get the explicit form of $\mathcal{L}_E(x, \dot{x})$ (the subscript E stands for "Euclidean time", which is used synonymously with "imaginary time" in field theory) in (2.18) by analytical continuation of the integration path as

$$iS = i\int_{-T/2}^{T/2} dt \left(\frac{m}{2}\dot{x}^2 - V[x]\right) \to i\int_{-\frac{T}{2}e^{-i\pi/2}}^{\frac{T}{2}e^{-i\pi/2}} dt \left[\frac{m}{2}\left(\frac{dx}{dt}\right)^2 - V[x]\right] \equiv -S_E. \tag{2.23}$$

After substituting $t \to -i\tau$ according to (2.13) we obtain

$$-S_E = i(-i)\int_{-T/2}^{T/2} d\tau \left(-\frac{m}{2}\dot{x}^2 - V[x]\right) \equiv -\int_{-T/2}^{T/2} d\tau \mathcal{L}_E[x] \tag{2.24}$$

(note that in Eq. (2.24) and in the following the dot indicates differentiation with respect to τ) from which we read off

$$\mathcal{L}_E[x] = \frac{m}{2}\dot{x}^2 + V(x). \tag{2.25}$$

Thus, besides turning the overall factor of i in the exponent into a minus sign (as anticipated in (2.18)), the analytical continuation has the crucial implication that the potential changes its sign, i.e. that it is turned upside down!

The consequence of this for the SCA is immediate: now there exist (one or more) solutions to the imaginary-time equation of motion

$$\frac{-\delta}{\delta x(\tau)} S_E[x] = m\ddot{x}_{cl} - V'(x_{cl}) = 0 \tag{2.26}$$

with tunneling boundary conditions. These solutions correspond to a particle which starts at x_i, rolls down the hill and through the local minimum of $-V(x)$ (which corresponds to the peak of the classically forbidden barrier of $+V(x)$) and climbs up to x_f, according to the boundary conditions

$$x\left(-\frac{T}{2}\right) = x_i, \qquad x\left(\frac{T}{2}\right) = x_f. \tag{2.27}$$

For later reference we note that the solutions of (2.26) carry a conserved quantum number, the "Euclidean energy"

$$E_E = \frac{m}{2}\dot{x}^2 - V(x) \tag{2.28}$$

which follows immediately from $\dot{E}_E = \dot{x}_{cl}[m\ddot{x}_{cl} - V'(x_{cl})]$ and the equation of motion (2.26).

The essential property of these solutions in our context is that they are saddle points of the imaginary-time path integral (2.18). For $\hbar \to 0$ the only nonvanishing contributions to the path integral therefore come from a neighborhood of x_{cl} (since those are least suppressed by the Boltzmann weight $\exp(-S_E/\hbar)$) and can be calculated in saddle-point approximation.

2.1.4 Saddle point approximation

Now let us perform the saddle-point approximation explicitly. To this end we write

$$x(\tau) = x_{cl}(\tau) + \eta(\tau) \tag{2.29}$$

and expand the action around the stationary (resp. saddle) point x_{cl} (later we will sum over the contributions from all saddle points) to second order in η,

$$S_E[x] = S_E[x_{cl}] + \frac{1}{2}\int d\tau \int d\tau' \eta(\tau) \frac{\delta^2 S_E[x_{cl}]}{\delta x(\tau)\,\delta x(\tau')}\eta(\tau') + O\left(\eta^3\right) \tag{2.30}$$

$$= S_E[x_{cl}] + \frac{1}{2}\int_{-T/2}^{T/2} d\tau \eta(\tau)\,\hat{F}(x_{cl})\,\eta(\tau) + O\left(\eta^3\right), \tag{2.31}$$

(the first derivative vanishes since S_E is minimal at x_{cl}) where we have abbreviated the operator

$$\frac{\delta^2 S_E[x]}{\delta x(\tau)\,\delta x(\tau')} = \left[-m\frac{d^2}{d\tau^2} + \frac{d^2 V(x_{cl})}{dx^2}\right]\delta(\tau - \tau') \equiv \hat{F}(x_{cl})\,\delta(\tau - \tau') \tag{2.32}$$

which governs the dynamics of the fluctuations around x_{cl}.

We now expand the fluctuations $\eta(\tau)$ into the (real) eigenfunctions of $\hat{F}$,

$$\eta(\tau) = \sum_n c_n \tilde{x}_n(\tau), \tag{2.33}$$

with

$$\hat{F}(x_{cl})\,\tilde{x}_n(\tau) = \lambda_n \tilde{x}_n(\tau). \tag{2.34}$$

The $\tilde{x}_n$ satisfy the boundary conditions (2.21) and are normalized according to (2.20). (Since $\hat{F}$ is real-hermitean and bounded, it has a complete spectrum.) For the moment we will also assume that all eigenvalues are positive, $\lambda_n > 0$. (This is not true in general. Below we will encounter examples of vanishing eigenvalues. In the case of so-called "bounce" solutions even negative eigenvalues do occur [2].) The action can then be written as

$$S_E[x] = S_E[x_{cl}] + \frac{1}{2}\sum_n \lambda_n c_n^2 + O\left(\eta^3\right). \tag{2.35}$$

Now we use the definition (2.22) of the measure to rewrite

$$Z(x_f, x_i) = \mathcal{N}\int D[x] e^{-\frac{S_E[x]}{\hbar}} \simeq \mathcal{N} e^{-\frac{S_E[x_{cl}]}{\hbar}} \int D[\eta] e^{-\frac{1}{2\hbar}\int_{-T/2}^{T/2} d\tau \eta(\tau)\hat{F}(x_{cl})\eta(\tau)} \tag{2.36}$$

as

$$Z(x_f, x_i) = \mathcal{N} e^{-\frac{S_E[x_{cl}]}{\hbar}} \prod_n \int_{-\infty}^{\infty} \frac{dc_n}{\sqrt{2\pi\hbar}} e^{-\frac{1}{2\hbar}\sum_n \lambda_n c_n^2} \tag{2.37}$$

$$= \mathcal{N} e^{-\frac{S_E[x_{cl}]}{\hbar}} \prod_n \int_{-\infty}^{\infty} dc_n \frac{e^{-\frac{1}{2\hbar}\lambda_n c_n^2}}{\sqrt{2\pi\hbar}} \tag{2.38}$$

and obtain, after performing the Gaussian integrations (which decouple and can thus be done independently),

$$Z(x_f, x_i) = \mathcal{N} e^{-\frac{S_E[x_{cl}]}{\hbar}} \prod_n \lambda_n^{-1/2} \tag{2.39}$$

$$\equiv \mathcal{N} e^{-\frac{S_E[x_{cl}]}{\hbar}} \left(\det \hat{F}[x_{cl}]\right)^{-1/2} \tag{2.40}$$

where a sum $\sum_{x_{cl}}$ is implied if there exists more than one saddle point. The formula (2.40) encapsulates the SCA to $Z(x_f, x_i)$ up to $O(\hbar)$.

Above, we have introduced the determinant of a differential operator as the product of its eigenvalues

$$\det \hat{O} = \prod_n \lambda_n, \qquad \text{for} \quad \hat{O}\psi_n(x) = \lambda_n \psi_n(x), \tag{2.41}$$

which generalizes the standard definition for quadratic matrices. In Subsection 2.3 we will give a more detailed definition of such determinants (which renders the above (infinite) product finite by adopting a specific choice for the normalization factor $\mathcal{N}$) and show how they can be calculated explicitly.

Let us summarize what we have accomplished so far. The seemingly artificial analytical continuation to imaginary times has allowed us to identify those paths whose neighborhoods give the dominant contributions to the path integral for a tunneling process in the semiclassical limit, and to evaluate this path integral to $O(\hbar)$ in the saddle-point approximation. In more physical terms the situation can be described as follows: for tunneling problems there exist no minimal-action trajectories (i.e. classical solutions) with the appropriate boundary conditions in real time. Therefore all trajectories between those boundary conditions (over which we sum in the real-time path integral) interfere highly destructively. Still, their net effect can be approximately gathered in a finite number of regions in function space, namely those in the neighborhood of the saddle points in imaginary time. In other words, while the tunneling amplitudes would have to be recovered at real times from a complex mixture of non-stationary paths (a forbidding task in practice), they are concentrated around the classical paths in imaginary time, and are therefore accessible to the saddle-point approximation. The destructive interference at real times leaves a conspicuous trace, however, namely the exponential suppression of (2.40) due to the Gamov factor $\exp(-S_E/\hbar)$, which is typical for tunneling amplitudes.

2.2 Double well (hill) potential and instantons

In the following sections we will apply the machinery developed above to simple tunneling problems with potentials which resemble as closely as possible the situation to be encountered later in QCD.

2.2.1 Instanton solution

Let us therefore specialize to tunneling processes between degenerate potential minima. A simple potential of the appropriate form is

$$V(x) = \frac{\alpha^2 m}{2x_0^2}\left(x^2 - x_0^2\right)^2 \tag{2.42}$$

(see Fig. 2.3, the coefficient is chosen for later convenience) which has three saddle points (with finite action, see the comment at the end of this section). Two of them are trivial

$$1) \qquad x_{cl}(\tau) = x_0 \tag{2.43}$$

$$2) \qquad x_{cl}(\tau) = -x_0 \tag{2.44}$$

and do not contribute to tunneling (since they cannot satisfy the corresponding boundary conditions; they do, however, contribute to $Z(x_0, x_0)$ or $Z(-x_0, -x_0)$). The third saddle point is time-dependent and corresponds

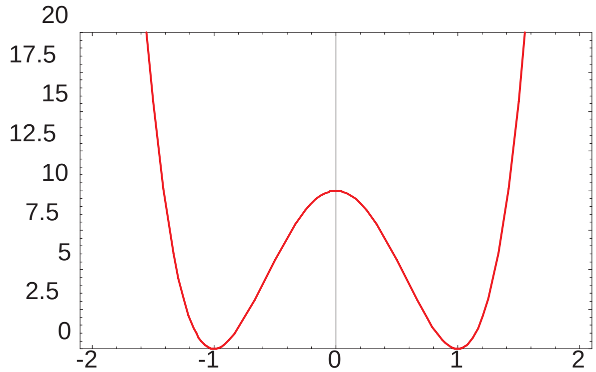

Figure 2.3: The double-well potential with $\frac{\alpha^2 m}{2x_0^2} = 10$ and $x_0 = 1$.

to the tunneling solution which interpolates between both maxima[2] of $-V$,

$$x_{cl}(-T/2) = +x_0, \tag{2.46}$$

$$x_{cl}(T/2) = -x_0, \tag{2.47}$$

or in the opposite direction, i.e. starting at $-x_0$ and ending at x_0.

The above classification of solutions is common to all potentials which look qualitatively like (2.42), and all the qualitative results obtained below will apply to such potentials, too. We have specialized to (2.42) only because for this choice the tunneling solution can be obtained analytically. The easiest way to get this solution starts from the conserved Euclidean energy (2.28) with $E_E = 0$, which corresponds to the limit $T \to \infty$ we are interested in (since then the particle in the mechanical analog system starts with (almost) zero velocity at x_0 and thus will need (almost) infinite time to reach $-x_0$):

$$E_E = \frac{m}{2}\dot{x}^2 - V = 0 \qquad \Rightarrow \qquad \dot{x} = \mp\sqrt{\frac{2V}{m}} \tag{2.48}$$

(as usual, the positive square root is implied) or, after separation of variables,

$$\mp\sqrt{\frac{m}{2}}\frac{dx}{\sqrt{V(x)}} = d\tau, \tag{2.49}$$

which can be integrated to become

$$\mp\sqrt{\frac{m}{2}}\int_{x_{cl}(\tau_0)}^{x_{cl}(\tau)}\frac{dx}{\sqrt{V(x)}} = \mp\frac{x_0}{\alpha}\int_{x_{cl}(\tau_0)}^{x_{cl}(\tau)}\frac{dx}{x_0^2 - x^2} = \mp\frac{x_0}{\alpha}\frac{1}{x_0}\operatorname{arctan}h\left(\frac{x}{x_0}\right)\Big|_{x_{cl}(\tau_0)}^{x_{cl}(\tau)} = \int_{\tau_0}^{\tau} d\tau = \tau - \tau_0. \tag{2.50}$$

We now choose τ_0, the integration constant, to be the "center" of the solution in imaginary time by requiring $x_{cl}(\tau_0) = 0$, and finally obtain

$$3) \qquad x_{cl}(\tau) \equiv x_{I/\bar{I}}(\tau) = \mp x_0 \tanh\alpha(\tau - \tau_0). \tag{2.51}$$

The solution with the minus sign is called the **instanton** (see Fig. 2.4) of the potential (2.42). Tunneling from $-x_0$ to x_0 corresponds to the anti-instanton solution $x_{\bar{I}}(\tau) = -x_I(\tau)$. Obviously, this nomenclature

[2] Note that, in the exact sense, this solution exists only for $T \to \infty$ (which is all we need to extract the ground-state properties). At large but finite T the corresponding solution requires a slight change in the boundary conditions to

$$x_{cl}(\pm T/2) = \mp x_0 \pm \varepsilon(T), \tag{2.45}$$

where ε is small.

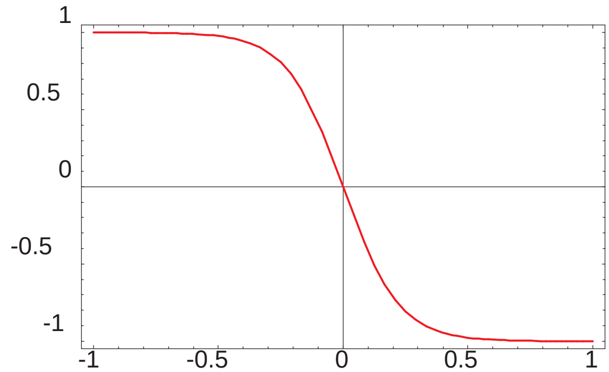

Figure 2.4: The instanton solution in the double well potential of Fig. 2.3.

is mere convention. (We repeat that, from the conceptual point of view, there is nothing special about the potential (2.42). Similar potentials with degenerate minima will have similar instanton solutions although those generally cannot be found analytically.)

Note that instantons are by necessity time-dependent since they have to interpolate between different minima of the potential. The name "instanton" (which goes back to 't Hooft) indicates furthermore that the tunneling transition happens fast, i.e. almost instantaneous[3]. Indeed, from the equation of motion (for $E_E = 0$) for the instanton,

$$\frac{dx_I}{d\tau} = -\sqrt{\frac{2}{m} V\left(x_I\right)}, \tag{2.52}$$

we obtain at large τ (where $V(x_I)$ can be expanded around $-x_0$ with $V(-x_0) = V'(-x_0) = 0$ and $V''(-x_0) = 4\alpha^2 m$)

$$\frac{dx_I}{d\tau} \simeq -2\alpha \left[x_I - (-x_0)\right]. \tag{2.53}$$

Separating variables,

$$\frac{dx_I}{x_I + x_0} = d \ln\left(x_I + x_0\right) = -2\alpha d\tau, \tag{2.54}$$

and integrating reveals that the asymptotic solution for the deviation

$$\Delta x_I(\tau) \equiv x_I(\tau) + x_0 \tag{2.55}$$

of the instanton from its "vacuum" value $-x_0$ decreases exponentially for large τ_0:

$$\Delta x_I(\tau_0 + \tau) \to \Delta x_I(\tau_0)\, e^{-2\alpha\tau}. \tag{2.56}$$

Moreover, the characteristic time scale

$$\tilde{\tau} = \frac{1}{2\alpha} \tag{2.57}$$

of the decay of Δx_I becomes arbitrarily small for large α. In the mechanical analog system, this means that to a particle starting from x_0 nothing much happens for a long time since its velocity remains almost zero. However, when it finally approaches the minimum of $-V$ at $x = 0$ it takes up speed fast, rushes through the minimum, and decelerates equally fast near to $-x_0$, spending all the remaining time to creep up fully and reach $-x_0$ at $T \to \infty$. The "abruptness" of the transition increases with α, the coupling parameter which controlls the height of the potential barrier.

[3] The instanton should therefore not be confused with a soliton-like particle. There is an important general equivalence, however, between instantons and static solitons in the corresponding (field) theory in one additional space dimension. Our instanton, for example, is the classical soliton solution (the so-called "kink") of the 1+1 dim. $\lambda\phi^4$ field theory.

The (Euclidean) action of the (anti-) instanton solution, which governs the exponential suppression of the tunneling amplitude, is easily obtained with the help of the "first integral" Eq. (2.48),

$$S_I \equiv S_E[x_I] = m\int_{-T/2}^{T/2} d\tau \dot{x}_I^2 = m\int_{x_0}^{-x_0} dx_I \dot{x}_I = -\frac{\alpha m}{x_0}\int_{x_0}^{-x_0} dx_I \left(x_0^2 - x_I^2\right) = \frac{4}{3}\alpha m x_0^2, \tag{2.58}$$

and equal to the action of the anti-instanton. Another way of writing the action,

$$S_E[x_I] = m\int_{x_0}^{-x_0} dx_I \dot{x}_I = \int_{-x_0}^{x_0} dx \sqrt{2mV(x)} = \int_{-x_0}^{x_0} dx p_E(x) \tag{2.59}$$

with $p_E(x) = \sqrt{2mV(x) - E_E}$ (recall that $E_E = 0$ for the instanton solution) shows that the exponential suppression factor $\exp(-S_I/\hbar)$ is nothing but the Gamov factor already encountered in Eq. (1.11). As expected, the path-integral approach reproduces the analytically continued WKB result for the tunneling amplitude.

Let us add two comments. First, there exist more saddle points, i.e. solutions of Eq. (2.26), than we have discussed above. Those additional solutions do not correspond to particle motion which both starts and ends at maxima of $-V$, however. Hence they do not satisfy the tunneling boundary conditions (2.46,2.47) although they might involve barrier penetration. Morevoer, they do not contribute to the SCA for the ground state since their action become infinite for $T \to \infty$. Physically, this is obvious since in such "runaway" solutions the particle moves infinitely far from the center of the potential ($|x| \to \infty$) and thereby reaches an infinite velocity at $T \to \infty$. Finally, we note that for very low tunneling barriers (i.e. for $\alpha \to 0$) and correspondingly high tunneling rates semiclassical methods may fail. In such cases variational approaches can sometimes help [7].

2.2.2 Zero mode

The constant solutions 1) and 2) above share the full symmetry of the Hamiltonian H (i.e. of V). This is not the case, however, for the instanton solution since the latter is localized in imaginary time around τ_0 and therefore lacks the continuous time translation invariance of the τ-independent H. As a consequence, the instanton solutions form a continuous and degenerate one-parameter family whose members are characterized by their time-center τ_0. In the saddle-point approximation to the path integral we have to sum (i.e. integrate) over the contributions from all τ_0. In the present section we will see how this can be done.

To this end, let us go back to our SCA "master formula" (2.40) which was derived under the assumption $\lambda_n > 0$, i.e. for positive eigenvalues of the fluctuation operator $\hat{F}$. It is not difficult to see that this assumption is not valid in the case of the instanton. Since the instanton "spontaneously" breaks the time-translation symmetry of V there must be a zero mode in the spectrum of $\hat{F}$ (for $T \to \infty$), i.e. an eigenfunction

$$\tilde{x}_0(\tau) = \sqrt{\frac{m}{S_I}}\dot{x}_I = -\frac{\sqrt{3\alpha}}{2}\frac{1}{\cosh^2(\alpha(\tau - \tau_0))} \tag{2.60}$$

with the eigenvalue $\lambda_0 = 0$ (which we have normalized to one at $T \to \infty$, where (2.48) applies:

$$\int_{-T/2}^{T/2} d\tau \left[\tilde{x}_0(\tau)\right]^2 = S_I^{-1}\int_{-T/2}^{T/2} d\tau \left(m\dot{x}_I^2\right) = 1). \tag{2.61}$$

In order to verify that (2.60) is indeed a zero mode it suffices to take a time derivative of the equation of motion (2.26):

$$m\ddot{x}_I - V'(x_I) = 0 \quad \Rightarrow \quad \left[m\partial_\tau^2 - V''(x_I)\right]\dot{x}_I = -\hat{F}\dot{x}_I = 0 \quad \Rightarrow \quad \hat{F}\tilde{x}_0(\tau) = 0. \tag{2.62}$$

The physical origin of this zero mode is quite obvious: it corresponds to an infinitesimal shift of the instanton center τ_0, i.e. an infinitesimal time translation of the instanton solution. Since the resulting, shifted instanton is degenerate with the original one, such a fluctuation costs no action and (2.35) implies that the corresponding eigenvalue of $\hat{F}$, i.e. λ_0, must be zero[4].

The instanton solution (2.51) is monotonically decreasing, which implies that the zero mode satisfies $\tilde{x}_0(\tau) \leq 0$ (cf. Eq. (2.60)) and therefore has no node. As a consequence, the zero mode is the (unique) eigenfunction of $\hat{F}$ with the lowest eigenvalue. All the remaining modes have $\lambda_n > 0$ and for them the Gaussian integrals in Eq. (2.38) are well defined. This is not the case, however, for the c_0 integration

$$\int_{-\infty}^{\infty} \frac{dc_0 e^{-\frac{1}{2\hbar}\lambda_0 c_0^2}}{\sqrt{2\pi\hbar}} = \int_{-\infty}^{\infty} \frac{dc_0}{\sqrt{2\pi\hbar}}, \tag{2.63}$$

which is not Gaussian at all! And neither should it be, since τ_0-shifting fluctuations $\eta(\tau) \sim \tilde{x}_0(\tau)$ are not damped even when they are large (i.e. for large c_0), due to the τ_0-independence of the instanton action. The latter renders the integrand c_0-independent and thus the integral divergent.

But this is just the kind of divergence which we should expect anyway from integrating over the infinite set of saddle points, i.e. over all τ_0, as discussed above. Indeed, it is easy to show that

$$dc_0 \propto d\tau_0 \tag{2.64}$$

by comparing the deviations dx from a given path $x(\tau)$ which are caused by

$$\text{small time translations } \tau_0 \to \tau_0 + d\tau_0 \quad \Rightarrow \quad dx = \frac{dx_I}{d\tau_0} d\tau_0 = -\dot{x}_I d\tau_0, \quad \text{and} \tag{2.65}$$

$$\text{small coefficient shifts } c_0 \to c_0 + dc_0 \quad \Rightarrow \quad dx = \frac{dx}{dc_0} dc_0 = \tilde{x}_0 dc_0 = \sqrt{\frac{m}{S_I}} \dot{x}_I dc_0. \tag{2.66}$$

Equating both deviations (and redefining the sign of dc_0 such that the integrals over c_0 and τ_0 have the same limits) gives

$$dc_0 = \sqrt{\frac{S_I}{m}} d\tau_0. \tag{2.67}$$

Thus the integration over c_0 can be done exactly and (in the limit $T \to \infty$) just amounts to summing over all one-instanton saddle points[5]. In the presence of a zero mode the expression (2.40) must therefore be replaced by

$$Z_I(-x_0, x_0) = \mathcal{N} e^{-\frac{S_E[x_I]}{\hbar}} \sqrt{\frac{S_E[x_I]}{2\pi\hbar m}} T \left(\det \hat{F}[x_I]'\right)^{-1/2} \tag{2.68}$$

where the prime at the determinant indicates that λ_0 is excluded from the product of eigenvalues. Of course, the factor T in (2.68) becomes infinite in the limit $T \to \infty$ which we will take in the end. This infinity is a consequence of the infinite amount of contributing saddle points and will be cancelled by other infinities (see e.g. Eq. (2.17)), leaving the observables perfectly finite as it should be.

2.3 Fluctuation determinant

Our next task is to make sense of functional determinants as encountered above, and to calculate

$$\det \hat{F}[x_I]' = \det \left[-m\frac{d^2}{d\tau^2} + V''(x_{cl})\right]' \tag{2.69}$$

from the leading fluctuations around the instanton (with the zero-mode contributions removed) explicitly. Together with the zero-mode part, this will take care of the $O(\hbar)$ contributions to the SCA. Although

[4] It might be useful to anticipate already at this stage that zero modes of fluctuatation operators, especially in quantum field theory, are not always due to continuous symmetries of the underlying dynamics. We will encounter an intruiguing example of a topologically induced zero mode below when dealing with Yang-Mills instantons.

[5] Since c_0 measures how far the instanton is collectively (i.e. equally at all times) shifted in (imaginary) time, it is an example of a "collective coordinate".

the explicit calculation of (2.69) is a rather technical exercise, we will go through it in considerable detail. One of the reasons is to develop intuition for the calculation of determinants in quantum field theory, as they generally occur during quantization of extended classical solutions (besides instantons e.g. solitons, monopoles, D-branes etc.). Indeed, all typical features of the calculation generalize rather directly to QCD. Hence it can serve as a pedagogical substitute for the actual QCD calculation, which is a tour de force [31] beyond the scope of our introduction to instanton physics.

Despite their uses, the following two subsections lie somewhat outside of our main focus. Readers not interested in the details of such calculations may therefore jump ahead to Section 2.4 without compromising their understanding of the remainder of these lectures.

2.3.1 Warm-up: harmonic oscillator

To prepare for the calculation of the instanton determinant, we first consider a simpler problem. We look for a nontrivial potential V which nevertheless leads to a fluctuation operator $\hat{F}$ whose eigenvalues can be obtained easily and analytically. The most obvious choice is to require V'' to be nonzero but x-independent, i.e. to work with the potential of the harmonic oscillator

$$V_{ho}\left(x\right)=\frac{1}{2}m\omega^{2}x^{2}. \tag{2.70}$$

The Euclidean analog potential is an inverted parabola. Of course, there is no tunneling in this potential. Instead of the tunneling boundary conditions (2.27) we therefore choose

$$x_{cl}\left(\pm T/2\right)=0. \tag{2.71}$$

The only solution to the equation of motion (2.26) under the conditions (2.71) is the constant

$$x_{ho}\left(\tau\right)=0, \tag{2.72}$$

corresponding to a particle which stays in "metastable" equilibrium at $x=0$ forever. Not surprisingly, its Euclidean action $S_{E,ho}$ vanishes and hence the exponential suppression factor $\exp\left(-S_E/\hbar\right)$ of the tunneling amplitudes is absent, as it should be.

To $O\left(\hbar\right)$ the Euclidean time evolution matrix element of the oscillator becomes

$$Z_{ho}\left(0,0\right)=\left\langle 0\left|e^{-H_{ho}T/\hbar}\right|0\right\rangle=\mathcal{N}\left(\det\hat{F}\left[x_{ho}\right]\right)^{-1/2} \tag{2.73}$$

where we have redefined $\hat{F}$ by absorbing a constant factor $\left(\det m\right)^{-1/2}$ into $\mathcal{N}$:

$$\hat{F}\left[x_{ho}\right]=-\frac{d^{2}}{d\tau^{2}}+\omega^{2}. \tag{2.74}$$

Since the classical solution (2.72) is time-translation invariant, there is no zero mode and the prime on the determinant can be omitted.

We now apply the formal definition of the determinant of an operator as the product of its eigenvalues,

$$\det\hat{F}\left[x_{ho}\right]=\prod_{n}\lambda_{ho,n}, \tag{2.75}$$

with the $\lambda_{ho,n}$ determined by

$$\hat{F}\left[x_{ho}\right]\tilde{x}_{n}\left(\tau\right)=\lambda_{ho,n}\tilde{x}_{n}\left(\tau\right). \tag{2.76}$$

As intended, the eigenfunctions of $\hat{F}\left[x_{ho}\right]$,

$$\tilde{x}_{n}\left(\tau\right)=a_{n}\sin\left(\frac{n\pi\tau}{T}\right)+b_{n}\cos\left(\frac{n\pi\tau}{T}\right), \tag{2.77}$$

and the corresponding eigenvalues

$$\lambda_{ho,n} = \left(\frac{n\pi}{T}\right)^2 + \omega^2, \qquad\qquad n = 1, 2, ... \tag{2.78}$$

can be read off immediately. The coefficients a_n, b_n are fixed by the normalization (2.20) and boundary conditions (2.21). In order for $\tilde{x}_n$ to satisfy the latter, we have to require $a_n = 0$ for n odd, and $b_n = 0$ for n even.

We then formally have

$$Z_{ho}(0,0) = \mathcal{N}\left(\prod_{n=1}^{\infty}\left[\left(\frac{n\pi}{T}\right)^2 + \omega^2\right]\right)^{-1/2} \tag{2.79}$$

and since the product is infinite, $\mathcal{N}$ must be infinite, too, in order to prevent Z_{ho} from vanishing identically. One way of disentangling the two infinities, or in other words to define the determinant, is to relate it to the determinant of the fluctuation operator without potential ($\omega = 0$): we simply split the above expression as

$$Z_{ho}(0,0) = Z_{free}(0,0)\left(\prod_{n=1}^{\infty}\left[1 + \left(\frac{\omega T}{n\pi}\right)^2\right]\right)^{-1/2}, \tag{2.80}$$

where the first factor corresponds to the propagator of free motion

$$Z_{free}(0,0) = \mathcal{N}\left[\prod_{n=1}^{\infty}\left(\frac{n\pi}{T}\right)^2\right]^{-1/2} = \left\langle 0 \left| e^{-\left(\frac{\hat{p}^2}{2m}\right)\frac{T}{\hbar}} \right| 0 \right\rangle = \int_{-\infty}^{\infty}\frac{dp}{2\pi}e^{-\left(\frac{p^2}{2m}\right)\frac{T}{\hbar}} = \sqrt{\frac{m\hbar}{2\pi T}} \tag{2.81}$$

and the forelast equation above is obtained by inserting the momentum eigenstates $\langle x|p\rangle = \exp\left(-ixp/\hbar\right)/\sqrt{2\pi}$.

Note that (2.81) amounts to a calculation of $\mathcal{N}$ from the requirement that it renders the free propagator finite and equal to the familiar result from basic quantum mechanics. Note, furthermore, that $\mathcal{N}$ is a property of the functional integral measure (2.22) and does not depend on the dynamics specified in $S_E[x]$. We will therefore encounter the same factor below when we return to the instanton problem.

The second, finite factor in (2.80) is obtained from the standard formula

$$\prod_{n=1}^{\infty}\left(1 + \frac{c^2}{n^2}\right) = \frac{\sinh(\pi c)}{\pi c} \tag{2.82}$$

which can be found e.g. in [25] or by Mathematica. Putting everything together, we finally have

$$Z_{ho}(0,0) = \mathcal{N}\left(\det \hat{F}[x_{ho}]\right)^{-1/2} = \sqrt{\frac{m\hbar\omega}{2\pi}}\left[\sinh(\omega T)\right]^{-1/2}. \tag{2.83}$$

Actually, this is the full Euclidean propagator of the harmonic oscillator: since the corresponding path integral is Gaussian, the SCA becomes exact! Below, we will be interested in the large-T limit

$$Z_{ho}(0,0) \to \sqrt{\frac{m\hbar\omega}{2\pi}}\left[\frac{\exp(\omega T)}{2}\right]^{-1/2} = \sqrt{\frac{m\hbar\omega}{\pi}}e^{-\omega T/2} \tag{2.84}$$

from which we recover, via Eq. (2.17), the familiar ground-state energy

$$E_{0,ho} = \frac{1}{2}\hbar\omega. \tag{2.85}$$

2.3.2 Instanton determinant

After having calculated two simpler functional determinants (those for the free particle and the harmonic oscillator), we are now prepared to attack the original problem. Our task will be to calculate the fluctuation determinant which appears in the $O(\hbar)$ tunneling amplitude

$$Z_I(-x_0, x_0) = Z_{ho}(0,0)\, e^{-\frac{S_E[x_I]}{\hbar}} \sqrt{\frac{S_E[x_I]}{2\pi\hbar m}}\omega T \left\{ \frac{\det \hat{F}[x_I]'}{\omega^{-2}\det \hat{F}[x_{ho}]} \right\}^{-1/2} \tag{2.86}$$

around an instanton. Above, we have factored out the root of the harmonic-oscillator determinant in (2.68) and used Eq. (2.73) for $Z_{ho}(0,0)$ to eliminate $\mathcal{N}$. This "renormalizes" $Z_I(-x_0, x_0)$ according to the procedure of the preceding section and leaves us with the above, finite ratio of determinants. Furthermore, we have factored out the lowest-mode contribution to the harmonic-oscillator determinant, $\lambda_{ho,0} \to \omega^2$ for $T \to \infty$, to balance the removal of the zero mode from the determinant of $\hat{F}[x_I]$.

In order to calculate $\det \hat{F}[x_I]'$, we have to deal with the spectrum of the operator

$$\hat{F}[x_I] = -m\frac{d^2}{d\tau^2} + V''(x_I) \tag{2.87}$$

where

$$V''(x_I) = \frac{2\alpha^2 m}{x_0^2}\left(3x_I^2 - x_0^2\right) = 2\alpha^2 m\left[3\tanh^2\alpha(\tau-\tau_0) - 1\right] = 2\alpha^2 m\left[2 - \frac{3}{\cosh^2\alpha(\tau-\tau_0)}\right] \tag{2.88}$$

($\cosh^2 x - \sinh^2 x = 1$) describes the interaction of Gaussian fluctuations with the smooth, step-like potential generated by the instanton. The eigenvalue equation of $\hat{F}[x_I]$ becomes, after absorbing the factor $(\det m)^{-1/2}$ as above into the normalization and specializing to $\tau_0 = 0$,

$$\left[-\frac{d^2}{d\tau^2} + 4\alpha^2 - \frac{6\alpha^2}{\cosh^2(\alpha\tau)}\right]\tilde{x}_\lambda(\tau) = \lambda\tilde{x}_\lambda(\tau), \tag{2.89}$$

subject to the tunneling boundary conditions

$$\tilde{x}_\lambda(\pm T/2) = 0 \tag{2.90}$$

for $T \to \infty$ (cf. (2.21)). As usual, the eigenvalue spectrum will be discrete for the "bound" states with $\lambda < 4\alpha^2$ and (in the limit $T \to \infty$) continuous for the scattering states where $\lambda > 4\alpha^2$.

Eq. (2.89) is of Schrödinger type and can be solved analytically in terms of hypergeometric functions (cf. §25, Problem 4 in [21], which makes use of the solution of §23, Problem 5 in the same book). To obtain the solutions explicitly, we rewrite (2.89) by means of the standard substitution

$$\xi = \tanh(\alpha\tau), \qquad \frac{d}{d\tau} = \frac{\alpha}{\cosh^2(\alpha\tau)}\frac{d}{d\xi} = \alpha\left(1-\xi^2\right)\frac{d}{d\xi}, \qquad \frac{d^2}{d\tau^2} = \alpha^2\left(1-\xi^2\right)\frac{d}{d\xi}\left(1-\xi^2\right)\frac{d}{d\xi} \tag{2.91}$$

in the form

$$\left[\frac{d}{d\xi}\left(1-\xi^2\right)\frac{d}{d\xi} - \frac{\epsilon^2}{1-\xi^2} + 6\right]\tilde{x}_\lambda(\xi) = 0, \tag{2.92}$$

where

$$\epsilon^2 = 4 - \frac{\lambda}{\alpha^2}. \tag{2.93}$$

The solutions of (2.92) in the interval of interest, $\xi \in [-1, 1]$, are associated Legendre functions (see, e.g. [25], chapter 8.7):

$$\tilde{x}_\lambda(\xi) = c_1 P_2^\epsilon(\xi) + c_2 Q_2^\epsilon(\xi). \tag{2.94}$$

Let us first consider the case $\varepsilon^2 > 0$, corresponding to the discrete levels. The first of the boundary conditions (2.90), i.e.

$$\tilde{x}_\lambda(\xi = -1) = 0, \tag{2.95}$$

requires $c_2 = 0$. The second one,

$$\tilde{x}_\lambda(\xi = 1) = 0, \tag{2.96}$$

can be nontrivially satisfied only for $\epsilon = 1, 2$ (for which the $P_2^{\epsilon}(\xi)$ reduce to a polynomials). Thus we have two discrete levels, with the eigenvalues

$$\lambda_0 = 0, \tag{2.97}$$

$$\lambda_1 = 3\alpha^2. \tag{2.98}$$

As anticipated, the lowest level λ_0 is the zero mode which we have already found in Section 2.2.2. The corresponding (not normalized) eigenfunction $P_2^2(\xi) = 3\left(1-\xi^2\right) = 3/\cosh^2(\alpha\tau)$ is proportional to (2.60) and, inserted into (2.89), yields indeed the eigenvalue zero. Both discrete eigenvalues are nonnegative, as expected in a repulsive potential. Nevertheless, the modified potential corresponding to the eigenvalues ϵ^2 has two "bound states", i.e. the discrete levels found above.

The calculation of the contributions from the continuum states with $\varepsilon^2 < 0$ is more involved but provides an instructive example of how to obtain a functional determinant in terms of phase shifts. Since the associated Legendre functions are not very convenient to deal with for non-integer ϵ, we rewrite $\tilde{x}_\lambda(\xi) = \left(1-\xi^2\right)^{\epsilon/2} w(\xi)$, which transforms Eq. (2.92) into

$$\left[\frac{d}{d\xi}\left(1-\xi^2\right)\frac{d}{d\xi}\left(1-\xi^2\right)^{\epsilon/2} - \epsilon^2\left(1-\xi^2\right)^{\epsilon/2-1} + 6\left(1-\xi^2\right)^{\epsilon/2}\right] w(\xi) = 0 \tag{2.99}$$

or

$$\left[\left(1-\xi^2\right)\frac{d^2}{d\xi^2} - 2(\epsilon+1)\,\xi\frac{d}{d\xi} - (\epsilon-2)(\epsilon+3)\right] w(\xi) = 0. \tag{2.100}$$

In terms of the new variable u with

$$u = \frac{1-\xi}{2}, \quad \xi = 1-2u, \quad 1-\xi^2 = 4u(1-u), \quad \frac{dw}{du} = -2\frac{dw}{d\xi}, \quad \frac{d^2w}{du^2} = 4\frac{d^2w}{d\xi^2} \tag{2.101}$$

this becomes the hypergeometric differential equation

$$\left[u(1-u)\frac{d^2}{du^2} + (\epsilon+1)(1-2u)\frac{d}{du} - (\epsilon-s)(\epsilon+s+1)\right] w(u) = 0 \tag{2.102}$$

with $s = 2$. Its solutions for positve ϵ^2 agree with those given above. The general solution for $\varepsilon^2 < 0$ reads

$$w(u) = c_3\,{}_2F_1\left[\frac{ik}{\alpha} - s, s + \frac{ik}{\alpha} + 1, 1 + \frac{ik}{\alpha}, u\right] + c_4 u^{-ik/\alpha}\,{}_2F_1\left[-s, s+1, 1 - \frac{ik}{\alpha}, u\right] \tag{2.103}$$

(see, e.g. chapter 9.1 of [25]) where we have defined $\epsilon = ik/\alpha$. The opposite sign, $k \to -k$, interchanges incoming and outgoing solutions (see below). For our case of $s = 2$ the solution (2.103) can be expressed in terms of elementary functions,

$$w(u) = c_3(1-u)^{-ik/\alpha}\left[k^2 - 3ik\alpha(1-2u) - 2\alpha^2(1-6(1-u))\right] \tag{2.104}$$

$$+\, c_4 u^{-ik/\alpha}\left[1 + \frac{6u\alpha(-ik + 2\alpha(1-u))}{(k+i\alpha)(k+2i\alpha)}\right] \tag{2.105}$$

or, transforming back to $\xi = 1-2u$,

$$w(\xi) = c_3\left(\frac{1+\xi}{2}\right)^{-ik/\alpha}\left[k^2 - 3ik\alpha\xi + 2\alpha^2(2+3\xi)\right] \tag{2.106}$$

$$+\, c_4\left(\frac{1-\xi}{2}\right)^{-ik/\alpha}\left[1 + \frac{3\alpha(1-\xi)(-ik+\alpha(1+\xi))}{(k+i\alpha)(k+2i\alpha)}\right] \tag{2.107}$$

and therefore

$$\tilde{x}_\lambda(\xi) = \left(1-\xi^2\right)^{ik/(2\alpha)} w(\xi) \tag{2.108}$$

$$= c_3 \left(\frac{\sqrt{1+\xi}}{2\sqrt{1-\xi}}\right)^{-ik/\alpha} \left[k^2 - 3ik\alpha\xi + 2\alpha^2(2+3\xi)\right] \tag{2.109}$$

$$+ c_4 \left(\frac{\sqrt{1-\xi}}{2\sqrt{1+\xi}}\right)^{-ik/\alpha} \left[1 + \frac{3\alpha(1-\xi)(-ik+\alpha(1+\xi))}{(k+i\alpha)(k+2i\alpha)}\right]. \tag{2.110}$$

To finally restore the original τ-dependence, resubstitute $\xi = \tanh(\alpha\tau)$ and use

$$\left(\frac{\sqrt{1\pm\xi}}{2\sqrt{1\mp\xi}}\right)^{-ik/\alpha} = e^{-ik/\alpha \ln\left(\frac{\sqrt{1\pm\xi}}{2\sqrt{1\mp\xi}}\right)} \tag{2.111}$$

together with

$$\ln\left(\frac{\sqrt{1\pm\xi}}{2\sqrt{1\mp\xi}}\right) = \ln\left(\frac{1}{2}\sqrt{\frac{\cosh(\alpha\tau)\pm\sinh(\alpha\tau)}{\cosh(\alpha\tau)\mp\sinh(\alpha\tau)}}\right) = \ln\left(\frac{e^{\pm\alpha\tau}}{2}\right) = \pm\alpha\tau - \ln 2. \tag{2.112}$$

As a result, we can write the full set of solutions with $\varepsilon^2 < 0$ as

$$\tilde{x}_\lambda(\tau) = \tilde{c}_3 e^{-ik\tau}\left[k^2 - 3ik\alpha\tanh(\alpha\tau) + 2\alpha^2(2+3\tanh(\alpha\tau))\right]$$
$$+ \tilde{c}_4 e^{ik\tau}\left[1 + \frac{3\alpha(1-\tanh(\alpha\tau))(-ik+\alpha(1+\tanh(\alpha\tau)))}{(k+i\alpha)(k+2i\alpha)}\right]. \tag{2.113}$$

(The constants $\tilde{c}_3$ and $\tilde{c}_4$ are left arbitrary since we do not need to fix the normalization of the $\tilde{x}_\lambda(\tau)$ or to impose the boundary conditions (2.90) for our purposes below.)

In order to calculate the continuum-mode contributions to the determinant, we have to find and multiply the eigenvalues corresponding to the solutions (2.113). Fortunately, there exists an elegant and efficient technique [11] for doing this, which can be applied to functional determinants in field theory as well. It takes advantage of the fact that (2.89) is a local equation. Thus we can obtain the eigenvalues in the asymptotic region $|\tau| \to \infty$, where the potential induced by the instanton field vanishes (a benefit of expanding around a localized solution) and (2.89) simplifies to

$$\left[\frac{d^2}{d\tau^2} + k^2\right]\tilde{x}_\lambda(\tau) = 0. \tag{2.114}$$

Here the momentum k, introduced above, replaces λ as the label of the continuum modes. Both are related by the "dispersion relation"

$$k^2 \equiv \lambda - 4\alpha^2 \qquad (\geq 0). \tag{2.115}$$

The solutions of Eq. (2.114) are "plane waves". Since their normalization is fixed (elastic scattering), the only effect of the instanton-induced potential can be a k-dependent phase shift. Moreover, the explicit, asymptotic solution reveals that no reflection occurs in this potential. Thus we have

$$\tilde{x}_\lambda(\tau) \propto e^{ik\tau + i\delta_k} \qquad \text{for } \tau \to -\infty, \tag{2.116}$$

$$\tilde{x}_\lambda(\tau) \propto e^{ik\tau} \qquad \text{for } \tau \to +\infty. \tag{2.117}$$

(Due to the absence of reflection, the linearly independent solutions $\propto \exp(-ik\tau)$ do not mix in.) The above discussion implies that all the required dynamical information about the eigenvalues is contained in the phase shifts of the solutions (2.113) in the limit $\tau \to -\infty$, where $\tanh(\alpha\tau) \to -1$ and thus

$$\tilde{x}_\lambda(\tau) \to \tilde{c}_3 e^{-ik\tau}\left[k^2 + 3ik\alpha - 2\alpha^2\right] + \tilde{c}_4 e^{ik\tau}\left[\frac{(1+ik/\alpha)(2+ik/\alpha)}{(1-ik/\alpha)(2-ik/\alpha)}\right]. \tag{2.118}$$

By comparing (2.116) and (2.118) we read off the phase shifts

$$\delta_k = -i \ln \left[\left(\frac{1 + ik/\alpha}{1 - ik/\alpha} \right) \left(\frac{2 + ik/\alpha}{2 - ik/\alpha} \right) \right]. \tag{2.119}$$

How are these phase shifts related to the continuum spectrum the eigenvalues k, and thus to the determinant to be calculated? Since the phase shifts encode the behavior of the scattering solutions at large $|\tau|$, one would expect the boundary conditions (2.90) to play a role. And indeed, they are all what is needed to establish the relation between the eigenvalues and δ_k. To show this, we start from the general solution

$$\tilde{x}_{gen,\lambda}(\tau) = A\tilde{x}_\lambda(\tau) + B\tilde{x}_\lambda(-\tau) \tag{2.120}$$

and impose the boundary conditions (2.90) to obtain

$$A\tilde{x}_\lambda\left(\frac{T}{2}\right) + B\tilde{x}_\lambda\left(-\frac{T}{2}\right) = A\tilde{x}_\lambda\left(-\frac{T}{2}\right) + B\tilde{x}_\lambda\left(-\frac{T}{2}\right) = 0. \tag{2.121}$$

This implies

$$\frac{\tilde{x}_\lambda\left(-\frac{T}{2}\right)}{\tilde{x}_\lambda\left(\frac{T}{2}\right)} = e^{-ikT - i\delta_k} = \pm 1 \tag{2.122}$$

and has the solutions

$$k_n = \frac{n\pi - \delta_k}{T} \tag{2.123}$$

in terms of the phase shifts δ_k. Due to the boundary conditions the k_n are discrete for finite T and become continuous in the limit $T \to \infty$ to be taken at the end.

It remains to calculate

$$\frac{\det \hat{F}[x_I]'}{\omega^{-2} \det \hat{F}[x_{ho}]} = \frac{\lambda_1}{\lambda_{ho,2}} \frac{\Pi_{n=1}\left(k_n^2 + 4\alpha^2\right)}{\Pi_{n=3}\left(k_{ho,n}^2 + \omega^2\right)} = \frac{3}{4} \frac{\Pi_{n=1}\left(k_n^2 + 4\alpha^2\right)}{\Pi_{n=3}\left(k_{ho,n}^2 + 4\alpha^2\right)}, \tag{2.124}$$

where we have factored out the contribution of the second levels (for $T \to \infty$)

$$\frac{\lambda_1}{\lambda_{ho,2}} = \frac{3\alpha^2}{\omega^2} = \frac{3}{4} \tag{2.125}$$

to the determinant ratio, using $k_{ho,n} = n\pi/T$ (cf. Eq. (2.78)), and specialized to $\omega = 2\alpha$. Since the relevant range of k-values in the eigenvalue products of (2.124) will turn out not to contain small k (see below), the contribution of the two lowest harmonic-oscillator eigenvalues can be multiplied to the above denominator with negligible effect (except for simplifying the ensuing expressions). We are thus led to consider the ratio

$$\frac{\Pi_{n=1}\left(k_n^2 + 4\alpha^2\right)}{\Pi_{n=1}\left(k_{ho,n}^2 + 4\alpha^2\right)} = \exp \sum_{n=1}^{\infty} \ln\left[\frac{k_n^2 + 4\alpha^2}{k_{ho,n}^2 + 4\alpha^2}\right]. \tag{2.126}$$

With

$$k_n^2 = \left(k_{ho,n} - \frac{\delta_k}{T}\right)^2 \simeq k_{ho,n}^2 - \frac{2\delta_k k_{ho,n}}{T} \tag{2.127}$$

(since $\delta_k/T \ll 1$ in the n-region which contributes non-negligibly for $T \to \infty$) we have

$$\ln\left[\frac{k_n^2 + 4\alpha^2}{k_{ho,n}^2 + 4\alpha^2}\right] \simeq \ln\left[1 - \frac{1}{T}\frac{2\delta_k k_{ho,n}}{k_{ho,n}^2 + 4\alpha^2}\right] \simeq -\frac{1}{T}\frac{2\delta_k k_{ho,n}}{k_{ho,n}^2 + 4\alpha^2}. \tag{2.128}$$

In the continuum limit

$$\sum_{n=1}^{\infty} f(k_n) = \frac{1}{\Delta k}\sum_{n=1}^{\infty} \Delta k f(k_n) \to \frac{T}{\pi}\int_0^{\infty} dk f(k) \tag{2.129}$$

$(\Delta k = k_{ho,n+1} - k_{ho,n} = \pi/T)$ we then obtain

$$\frac{\Pi_{n=1}\left(k_n^2+4\alpha^2\right)}{\Pi_{n=1}\left(k_{ho,n}^2+4\alpha^2\right)} = \exp\left(-\frac{1}{\pi}\int_0^\infty dk \frac{2\delta_k k}{k^2+4\alpha^2}\right). \tag{2.130}$$

To evaluate the integral in the exponent, we note that

$$\frac{d}{dk}\ln\left[1+\frac{k^2}{4\alpha^2}\right] = \frac{2k}{k^2+4\alpha^2} \tag{2.131}$$

allows us to rewrite

$$\int_0^\infty dk \frac{2k\delta_k}{k^2+4\alpha^2} = -\int_0^\infty dk \frac{d\delta_k}{dk}\ln\left[1+\frac{k^2}{4\alpha^2}\right] = -\int_0^\infty d\kappa \frac{d\delta_\kappa}{d\kappa}\ln\left[1+\kappa^2\right] \tag{2.132}$$

(the surface term vanishes since $\delta_{k=\infty} = 0$) in terms of the dimensionless variable $\kappa \equiv k/(2\alpha)$. The benefit of the partial integration is that the κ-derivative removes the logarithm in the expression (2.119) for the phase shifts from the integrand,

$$\frac{d\delta_\kappa}{d\kappa} = -i\frac{d}{d\kappa}\ln\left[\frac{1+2i\kappa}{1-2i\kappa}\frac{1+i\kappa}{1-i\kappa}\right] = \frac{2}{1+\kappa^2}+\frac{4}{1+4\kappa^2}, \tag{2.133}$$

so that the remaining integral can be done analytically:

$$\int_0^\infty dk \frac{2k\delta_k}{k^2+4\alpha^2} = -\int_0^\infty d\kappa\left(\frac{2}{1+\kappa^2}+\frac{4}{1+4\kappa^2}\right)\ln\left[1+\kappa^2\right] = \pi\ln 9. \tag{2.134}$$

Thus

$$\frac{\det\hat{F}\left[x_I\right]'}{\omega^{-2}\det\hat{F}\left[x_{ho}\right]} = \frac{1}{12} \tag{2.135}$$

and together with (2.86) and (2.84) we can now assemble our final result for $Z_I(-x_0, x_0)$ at large T,

$$Z_I(-x_0,x_0) = \sqrt{\frac{m\hbar\omega}{\pi}}e^{-\omega T/2}\omega T\sqrt{\frac{6S_E\left[x_I\right]}{\pi\hbar m}}e^{-\frac{S_E[x_I]}{\hbar}}. \tag{2.136}$$

This is the quantum mechanical propagator of the double-well tunneling problem, to $O(\hbar)$ in the semi-classical approximation around a single instanton. The exact analogs of both the exponential Gamov factor and the preexponential factor $\sqrt{S_I}$ from the zero mode appear in the one-instanton sector of QCD.

2.4 Dilute instanton gas

Up to now, we have concentrated on the saddle points which correspond to single-instanton solutions. This is not the whole story, however: there are additional (approximate) saddle points which also contribute to the semiclassical tunneling amplitude for large T. We will now analyze those "multi-instanton solutions", first in the familiar double-well potential and subsequently in its periodic extension (which represents the closest quantum-mechanical analog of the semiclassical Yang-Mills vacuum).

2.4.1 Double-well potential

Since the instanton deviates only in a small time interval $\Delta\tau = 1/(2\alpha)$ appreciably from x_0 or $-x_0$ (cf. Eq. (2.57)), and since the overlap between neighboring instantons and anti-instantons is exponentially small (cf. (2.56)), multi-(anti-)instanton solutions of (2.26) can be approximately written as a chain (i.e. an ordered superposition) of N alternating instantons and antiinstantons, sufficiently far separated in time by the (average) intervall

$$\bar{\Delta}_\tau = \frac{T}{N} \gg \frac{1}{2\alpha}. \tag{2.137}$$

Those chains correspond to N tunneling processes, back and forth between both minima of the potential. It is intuitively clear that their importance, associated with the frequency of their occurence, increases with growing T.

The approximate N-instanton solutions, composed of single, alternating instantons and anti-instantons at times $\tau_{0,k}$, can thus be written as[6]

$$x_N(\tau) = \sum_{k=1}^{N} x_{I,\bar{I}}(\tau - \tau_{0,k}), \tag{2.139}$$

where N must be odd in order to satisfy the boundary conditions (2.46) and (2.47). They become exact solutions for infinite separations $|\tau_{0,k+1} - \tau_{0,k}| \to \infty$. The (anti-) instanton centers are ordered in Euclidean time as

$$-\frac{T}{2} \ll \tau_{0,1} \ll \tau_{0,2} ... \ll \tau_{0,N} \ll \frac{T}{2}. \tag{2.140}$$

All multi-instanton solutions have to be included as additional saddle-points in the SCA. We are now going to derive the corresponding expression for $Z(-x_0, x_0)$ by using the approximate solutions (2.139) instead. This simplification goes under the name of "dilute instanton gas approximation" (DIGA).

The contribution of the approximate N-instanton solution to the path integral is

$$Z_N \simeq \mathcal{N} \int D[\eta] e^{-S_E[x_N+\eta]/\hbar}. \tag{2.141}$$

We now write the general fluctuation $\eta(\tau)$ as a sum of independent, localized fluctuations $\eta_k(\tau)$ around the single (anti-) instantons and $\eta_0(\tau)$ around the approximately constant pieces $x(\tau) = \pm x_0$ beween them:

$$\eta(\tau) = \eta_0(\tau) + \sum_{k=1}^{N} \eta_k(\tau). \tag{2.142}$$

This implies that $\eta_0(\tau)$ can be finite all over $[-T/2, T/2]$ (except at the boundaries) while the $\eta_k(\tau)$ are time-localized around the k-th (anti-) instanton.

Hence the action approximately decomposes into the sum of actions for single (well-separated) (anti-) instantons and a "vacuum" piece,

$$S_E[x_N + \eta] \simeq S_E[x_0 + \eta_0] + \sum_{k=1}^{N} S_E[x_I + \eta_k] \tag{2.143}$$

(recall that $S[x_{cl} = \pm x_0] = 0$ and $S[x_0 + \eta] = S[-x_0 + \eta]$, due to the symmetry of the potential). This formula expresses the (approximate) fact that the (anti-) instantons have too little overlap to interact. Furthermore, the path integral measure factorizes into integrals over the localized fluctuations around the

[6] Eq. (2.139) is to be understood as a shortcut for the more precise (and tedious) expression

$$x_N(\tau) = \sum_{k=1}^{(N+1)/2} x_I(\tau - \tau_{0,2k-1}) + \sum_{k=1}^{(N-1)/2} x_{\bar{I}}(\tau - \tau_{0,2k}) \tag{2.138}$$

which takes explicit care of the facts that the first and last transitions in the chain must be instantons, and that those inbetween consist of an anti-instanton and $(N-3)/2$ pairs of an anti-instanton followed by an instanton.

instantons and those inbetween[7]. As a consequence, Z_N factorizes as

$$Z_N(-x_0, x_0) \simeq \mathcal{N} \int D[\eta_0] e^{-S_E[x_0+\eta_0]/\hbar} \times \prod_{k=1}^{N} \mathcal{N} \int D[\eta_k] e^{-S_E[x_I+\eta_k]/\hbar} \tag{2.144}$$

$$= Z_0(x_0, x_0) [Z_I(-x_0, x_0)]^N . \tag{2.145}$$

We have written the above product in a sloppy manner. Strictly speaking, the N factors $Z_I(-x_0, x_0)$ have different boundary conditions since the endpoints $\pm x_0$ are (almost) reached at different times in each factor, associated with the time-ordering of the (anti-) instanton from which they arise. However, due to time translation invariance this makes no difference for the value of the $Z_I(-x_0, x_0)$ except through the zero-mode contribution. Indeed, for the latter we have to integrate over the temporal position τ_0 of the (anti-) instanton (or, equivalently, over c_0),

$$Z_I = Z_I' \sqrt{\frac{S_I}{2\pi\hbar m}} \int d\tau_0 \equiv \tilde{Z}_I \int d\tau_0, \tag{2.146}$$

where now the range of the center $\tau_{0,k}$ of the k-th (anti-) instanton is restricted by the requirement that it occurs after the $(k-1)$-th, i.e. $\tau_{0,k-1} < \tau_{0,k} < T/2$. The integration over the $\tau_{0,k}$ thus takes the form

$$\int_{-T/2}^{T/2} d\tau_{0,1} \int_{\tau_{0,1}}^{T/2} d\tau_{0,2} ... \int_{\tau_{0,N-1}}^{T/2} d\tau_{0,N} = \frac{T^N}{N!} \tag{2.147}$$

which results in

$$Z_N \simeq Z_0 \frac{\left(\tilde{Z}_I T\right)^N}{N!}. \tag{2.148}$$

The above expression determines the N-instanton contribution explicitly since we had already calculated Z_0 in (2.84) (with $\omega = 2\alpha$) and since $\tilde{Z}_I$ can be immediately obtained from (2.136) and (2.58):

$$Z_0(\pm x_0, \pm x_0) = \mathcal{N} \left(\det\left[-\partial_\tau^2 + \omega^2\right]\right)^{-1/2} \to \left(\frac{m\hbar\omega}{\pi}\right)^{1/2} e^{-\omega T/2}, \tag{2.149}$$

$$\tilde{Z}_I = 2\alpha \sqrt{\frac{6 S_E[x_I]}{\pi\hbar m}} e^{-\frac{S_E[x_I]}{\hbar}} = 4\sqrt{\frac{2\alpha^3 x_0^2}{\pi\hbar}} e^{-\frac{4}{3}\alpha m x_0^2/\hbar}. \tag{2.150}$$

In order to collect the multi-instanton contributions to $Z(x_0, -x_0)$, we have to sum over all odd N (recall that odd N's are required by the boundary conditions (2.46), (2.47)) and obtain

$$Z_{DIGA}(x_0, -x_0) = Z_0 \sum_{N \text{ odd}} \frac{\left(\tilde{Z}_I T\right)^N}{N!} = \frac{Z_0}{2} \left\{ e^{\tilde{Z}_I T} - e^{-\tilde{Z}_I T} \right\} = Z_0 \sinh\left(\tilde{Z}_I T\right). \tag{2.151}$$

(Note that the second exponential on the RHS above removes the contributions even in $\tilde{Z}_I T$ and doubles those odd in $\tilde{Z}_I T$.) An analogous expression for $Z_{DIGA}(-x_0, -x_0)$ is obtained when only contributions from even numbers of instantons are summed. Both results can be combined into the expression

$$Z_{DIGA}(\pm x_0, -x_0) = \frac{1}{2} \left(\frac{\hbar\omega}{\pi}\right)^{1/2} e^{-\omega T/2} \left\{ e^{\tilde{Z}_I T} \mp e^{-\tilde{Z}_I T} \right\} \tag{2.152}$$

$$= \frac{1}{2} \left(\frac{\hbar\omega}{\pi}\right)^{1/2} \left\{ e^{-\left(\omega/2 - \tilde{Z}_I\right)T} \mp e^{-\left(\omega/2 + \tilde{Z}_I\right)T} \right\} \tag{2.153}$$

[7]Due to the limited support of the localized fluctuations, they do no longer form a complete set on the whole time interval. Any fluctuation $\tilde{x}(\tau)$ for $\tau \in [-\frac{T}{2}, \frac{T}{2}]$ must therefore be described piecewise in bases around every instanton. Since the path integral is the product of integrals over the coefficients of those base functions, it factorizes naturally.

from which the two lowest energy levels of the system (i.e. those of the ground state and the first excited state) can be found as

$$E_0 = -\hbar \lim_{T\to\infty} \frac{1}{T} \ln Z_{DIGA}(\pm x_0, -x_0) = -\hbar \lim_{T\to\infty} \frac{1}{T}\left[-\left(\frac{\omega}{2} - \tilde{Z}_I\right)T\right] = \frac{\hbar\omega}{2} - \hbar\tilde{Z}_I, \tag{2.154}$$

$$E_1 = -\hbar \lim_{T\to\infty} \frac{1}{T}\left[-\left(\frac{\omega}{2} + \tilde{Z}_I\right)T\right] = \frac{\hbar\omega}{2} + \hbar\tilde{Z}_I. \tag{2.155}$$

As expected, the effect of tunneling is to split the degenerate (would-be) ground state energies $\hbar\omega/2$ of the wave functions $|\pm x_0\rangle$ centered in each of the two minima of the potential[8]. The corresponding energy eigenstates are similarly obtained from the prefactors of the exponential in (2.153) (cf. Eq. (2.16)). For the ground (first excited) state we find the symmetric (antisymmetric) linear combination

$$|0\rangle = \frac{1}{\sqrt{2}}\left\{|x_0\rangle + |-x_0\rangle\right\}, \tag{2.158}$$

$$|1\rangle = \frac{1}{\sqrt{2}}\left\{|x_0\rangle - |-x_0\rangle\right\}. \tag{2.159}$$

These are the standard WKB results for tunneling amplitudes, with the typical splitting

$$\Delta E \sim e^{-\frac{S_E[x_I]}{\hbar}} \tag{2.160}$$

of the energy levels of the states connected by tunneling, although we have obtained them in the somewhat less familiar framework of the imaginary-time path integral. It is remarkable that the SCA works well even for the lowest-lying states of the system, i.e. for those with the largest de-Broglie wavelengths λ, although this λ is not small compared to the size of the potential. Note furthermore that, in contrast to the case without tunneling and the ground states $|\pm x_0\rangle$, the above states are eigenstates of parity (under which $x_0 \leftrightarrow -x_0$). The new ground state (2.158), in particular, is parity invariant: the artificially broken parity in the absence of tunneling is restored.

Before leaving this section, we should address a potential concern in summing over the dilute instanton gas in Eq. (2.151). Indeed, for an increasing number N of (anti-)instantons in the constant interval T the diluteness condition (2.137) is less and less satisfied, implying that from some large N onwards the corresponding terms in the sum will violate this fundamental DIGA requirement. However, Stirling's formula $n! \simeq \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$ for large n shows that the sum (2.151) is dominated by terms with

$$\frac{\tilde{Z}_I T}{N} \sim O(1) \tag{2.161}$$

and that contributions from larger N are rapidly suppressed. As a consequence, only N with

$$\frac{N}{T} \lesssim C e^{-\frac{S_E[x_I]}{\hbar}}, \tag{2.162}$$

i.e. with an exponentially small instanton density in the semiclassical limit (which can be improved at will by increasing α since we had found in Eq. (2.58) that $S_E[x_I] = (4/3)\,\alpha m x_0^2$), contribute significantly to the

[8] We note in passing that the energy splitting between the classical would-be ground states could have been obtained from the one-instanton approximation alone, i.e. without employing the DIGA. Indeeed, restricting to small T and expanding

$$\left\langle x_f \left| e^{-HT/\hbar} \right| x_i \right\rangle = 1 - \frac{T}{\hbar}\langle x_f |H| x_i\rangle + O\left(T^2\right) \tag{2.156}$$

the propagator is governed by the (time independent) matrix element

$$\langle x_f |H| x_i\rangle \tag{2.157}$$

which can be calculated in the one-instanton approximation.

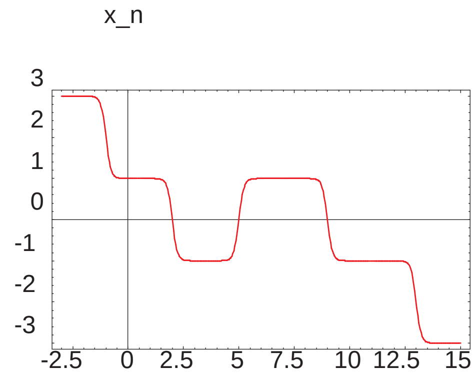

Figure 2.5: A multi-instanton solution in the periodical potential. The abscissa denotes the Euclidean time τ while the ordinate gives the position variable x. The integer values of x correspond to the minima $x_{0,n}$ of the periodical potential.

sum (2.151). This ensures that Z_{DIGA} is dominated by terms in which the DIGA diluteness requirement (2.137) is well satisfied[9].

Our main motivation for the use of instanton methods in the solution of the above tunneling problem was that this procedure can be straightforwardly generalized to gauge theories. In order to develop the analogy with QCD as far as possible, however, we should still go a step farther in quantum mechanics and study a periodical potential with degenerate minima. This will be the subject of the following section.

2.4.2 Periodic potential

Let us now consider the periodic extension of the double well potential (which thereby becomes bounded from above). This type of potential most closely resembles the situation which we will encounter below in the QCD vacuum.

In a periodic potential with degenerate minima, instantons and anti-instantons can arbitrarily follow each other, starting out at the minimum where the previous one had ended, i.e. connecting adjacent minima $x_{0,n}$ and $x_{0,n\pm1}$ (see Fig. 2.5). Generalizing our earlier conventions, we will refer to an instanton (anti-instanton) as the segment of the solution which interpolates between neighboring minima to the left (right), i.e. which decreases (increases) n. Getting from the minimum with index n_i to the one with index n_f, corresponding to the boundary conditions

$$x\left(-\frac{T}{2}\right) = x_{0,n_i}, \qquad x\left(\frac{T}{2}\right) = x_{0,n_f}, \tag{2.163}$$

therefore requires the number of anti-instantons minus the number of instantons, $N_{\bar{I}} - N_I$, to equal $n_f - n_i$. As a consequence, the semiclassical propagator becomes

$$Z_{per}\left(x_{n_f}, x_{n_i}\right) \simeq Z_0 \sum_{N_I=0}^{\infty} \sum_{N_{\bar{I}}=0}^{\infty} \frac{\left(\tilde{Z}_I T\right)^{N+N_{\bar{I}}}}{N! N_{\bar{I}}!} \delta_{N_{\bar{I}}-N_I-(n_f-n_i)}. \tag{2.164}$$

With the representation

$$\delta_{ab} = \int_0^{2\pi} \frac{d\theta}{2\pi} e^{i\theta(a-b)} \tag{2.165}$$

[9]Unfortunately, it will turn out below that in QCD, where we do not have an adjustable parameter analogous to α, the situation is much more complex and the corresponding DIGA does not provide a self-consistent approach to the vacuum wave functional.

of the Kronecker symbol, Z_{per} can be rewritten as

$$Z_{per}\left(x_{n_f},x_{n_i}\right) \simeq Z_0 \int_0^{2\pi} \frac{d\theta}{2\pi} e^{-i\theta(n_f-n_i)} \sum_{N_I=0}^{\infty} \frac{\left(\tilde{Z}_I T e^{-i\theta}\right)^{N_I}}{N_I!} \sum_{N_{\bar{I}}=0}^{\infty} \frac{\left(\tilde{Z}_I T e^{i\theta}\right)^{N_{\bar{I}}}}{N_{\bar{I}}!} \tag{2.166}$$

$$= Z_0 \int_0^{2\pi} \frac{d\theta}{2\pi} e^{-i\theta(n_f-n_i)} e^{\tilde{Z}_I T e^{-i\theta} + \tilde{Z}_I T e^{i\theta}} \tag{2.167}$$

and by using again the expression (2.149) for Z_0 we arrive at

$$Z_{per}\left(x_{n_f},x_{n_i}\right) = \left(\frac{m\hbar\omega}{\pi}\right)^{1/2} e^{-\omega T/2} \int_0^{2\pi} \frac{d\theta}{2\pi} e^{-i\theta(n_f-n_i)} e^{2\tilde{Z}_I T\cos\theta} \tag{2.168}$$

$$= \left(\frac{m\hbar\omega}{\pi}\right)^{1/2} \int_0^{2\pi} \frac{d\theta}{2\pi} e^{-i\theta(n_f-n_i)} e^{-\left(\omega/2-2\tilde{Z}_I\cos\theta\right)T}. \tag{2.169}$$

As before, we can now obtain the low-lying energy levels from the $T \to \infty$ limit. The above expression shows that we get, in fact, a continuous "band" of energies parametrized by θ (since (2.168) is the sum of contributions from the lowest energy levels in (2.14))

$$E_0(\theta) = -\hbar \lim_{T\to\infty} \frac{1}{T}\left[-\left(\frac{\omega}{2} - 2\tilde{Z}_I\cos\theta\right)T\right] \tag{2.170}$$

$$= \frac{\hbar\omega}{2} - 2\hbar\tilde{Z}_I\cos\theta. \tag{2.171}$$

Of course, it is well known that the energies in a periodic potential form continuous "bands". Our result is just the analog of, e.g., the lowest-lying band of electron states in the periodic potential of a metal. Not surprisingly, then, the corresponding eigenstates are the "Floquet-Bloch waves"

$$|\theta\rangle = \frac{1}{\sqrt{2\pi}}\left(\frac{\hbar\omega}{\pi}\right)^{1/4} \sum_n e^{in\theta}|n\rangle, \tag{2.172}$$

where $|n\rangle$ is the state localized at the n-th minimum of the periodic potential.

This concludes our discussion of instantons in quantum mechanics. Much more could be said about them [26], their cousins (sometimes called "bounces") which mediate tuneneling between nondegenerate minima, their connections to large-order perturbation theory in the real-time theory, etc. However, we refrain from doing so since we have reached our main objective: to obtain the semiclassical expansion and the ground state properties in a potential which mimicks the situation in the QCD vacuum. We will now move on to discuss "the real thing", namely instantons in QCD itself.

Chapter 3

Instantons in QCD

What is the relevance of the above tunneling discussion for QCD? As anticipated, the answer is that several pertinent aspects of the SCA[1] generalize to four-dimensional Euclidean (i.e. imaginary-time) Yang-Mills theory once we have identified the saddle points of the corresponding functional integral in imaginary time. Our first task will therefore be to find the minima of the classical Yang-Mills action. Remarkably, they turn out to be determined by the topology of the gauge group and to form the minima of a periodic "potential" analogous to the one encountered above.

The corresponding tunneling solutions, the Yang-Mills instantons, are classical gauge fields with intriguing topological properties. The latter induce new physical phenomena without analogy in the quantum mechanical examples. Some of these phenomena, including several ones related to the light-quark sector of QCD, will be discussed in subsequent sections of this chapter.

3.1 Vacuum topology and Yang-Mills instantons

3.1.1 Topology of the Yang-Mills vacuum

In order to develop some intuition for semiclassical ground-state properties of QCD, let us start as in the preceding quantum mechanical examples by searching for the minima of the Euclidean Yang-Mills action. Restricting for the moment to the gluon sector, the latter reads

$$S[G] = \frac{1}{4}\int d^4x \left[G^a_{\mu\nu}G^a_{\mu\nu}\right] \tag{3.1}$$

$$= \frac{1}{2}\int d^4x \left[E^a_i E^a_i + B^a_i B^a_i\right] \geq 0 \tag{3.2}$$

where the gluon field strength tensor is

$$G_{\mu\nu}(x) = \partial_\mu G_\nu - \partial_\nu G_\mu + ig\left[G_\mu, G_\nu\right] \equiv G^a_{\mu\nu} t^a, \qquad t^a = \frac{\lambda^a}{2}, \tag{3.3}$$

and the chromoelectric and -magnetic fields are defined as

$$E^a_i = G^a_{i4}, \qquad B^a_i = -\frac{1}{2}\varepsilon_{ijk}G^a_{jk}. \tag{3.4}$$

[1] We note in passing that there exists a somewhat complementary way of looking at the SCA in quantum field theory, namely as an expansion in the number of Feynman-graph loops (which reflects the non-perturbative nature of the SCA from a different angle).

The corresponding quantum field theory (including the quarks) determines the structure of the QCD vacuum, i.e. the unique state of lowest energy on which the Fock space is built. Since QCD is strongly coupled and therefore non-perturbative at low energies, we expect the vacuum to be populated by strong fields. (This is in contrast to standard QED, for example, where the vacuum contains mostly zero-point fluctuations, i.e. weakly interacting electron-positron pairs and photons which can be handled perturbatively.)

A measure for the strength of the QCD vacuum fields can be obtained from the trace anomaly, which relates the energy density ϵ_{vac} of the vacuum to the phenomenologically known vacuum expectation value of the square of the gluon field strength tensor, the so-called "gluon condensate" (renormalized at about 1 GeV):

$$\epsilon_{vac} \simeq -\frac{b_1}{128\pi^2} \langle 0 \left| g^2 G^2 \right| 0 \rangle \simeq -\frac{1}{2} \frac{\text{GeV}}{\text{fm}^3} \ll \epsilon_{pert}. \tag{3.5}$$

This relation shows that the nonperturbative vacuum fields are indeed exceptionally strong: they reduce the vacuum energy in a tiny cube of size 10^{-15}m by about half a proton mass! As we will see in the remainder of this section, part of this reduction is due to tunneling processes mediated by instantons.

Strong fields can contain a very large number of quanta, and those quanta can become coherent and render the corresponding action large compared to $\hbar$. In other words, such fields behave (semi-) classically since quantum fluctuations are of $O(\hbar)$ and thus contribute only relatively small corrections. The above reasoning suggests that insight into the vacuum fields of QCD may be gained from a semiclassical perspective. In the following, we will explore this perspective while paying special attention to robust and generic features which are likely to survive even stronger quantum fluctuations. The most important such features will turn out to be "global", i.e. topological properties of the vacuum fields which are invariant under continuous deformations (and therefore in particular under time evolution).

The first step towards a semiclassical approach to the QCD vacuum (starting at the lowest order of $\hbar$) is to find those classical fields which minimize the (static) Yang-Mills energy or, equivalently, the Euclidean action (3.1). Such fields are sometimes called "classical vacua". Since the Euclidean action (3.1) is non-negative (in contrast to its counterpart in Minkowski space) its absolute minima will be the gluon fields with zero action. These fields are the "pure gauges"

$$G_\mu^{(pg)} = \frac{-i}{g} U \partial_\mu U^\dagger \tag{3.6}$$

where $U(x) \in SU(3)$ is an element of the gauge group. It is easy to check explicity that the field strength of pure gauges vanishes,

$$G_{\mu\nu}^{(pg)} = 0. \tag{3.7}$$

At first sight, one might think that these fields would be natural candidates on which to build a semiclassical expansion. A moment's reflection shows, however, that none of them could generate an acceptable vacuum since they are neither unique nor gauge invariant.

In fact, it can be easily seen that the $G_\mu^{(pg)}$ fall into an enumerable infinity of topological equivalence classes. In order to demonstrate this, we start by choosing the temporal gauge

$$G_0 = 0 \tag{3.8}$$

so that the residual gauge transformations (which conserve this gauge condition) are time-independent. Moreover, for our purposes it is useful to restrict the residual gauge transformations further by letting them approach a constant (chosen to be unity) at spacial infinity, i.e.

$$U(\vec{x}) \to 1 \qquad \text{for} \qquad |\vec{x}| \to \infty. \tag{3.9}$$

This restriction guarantees that the gauge fields satisfiy definite boundary conditions at the surface of a large box (which should not affect the local physics inside).

The effect of the condition (3.9) is that, from the point of view of gauge transformations, the space R^3_{space} is effectively compactified to S^3_{space} (i.e. to a 3-dimensional sphere of infinite radius), i.e. different boundary points at $|\vec{x}| \to \infty$ cannot be distinguished by $U(\vec{x})$ and thus can be identified. This is completely analogous

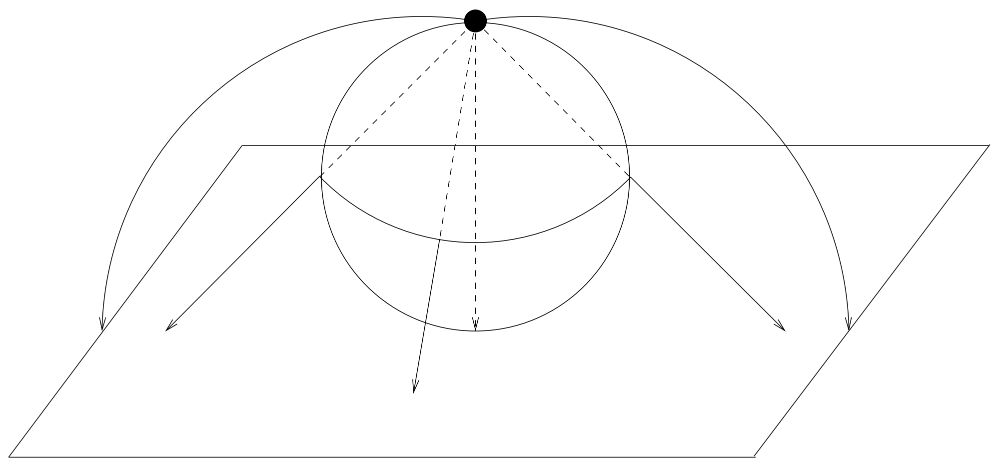

Figure 3.1: Schematic view of the generalized stereographic projection which identifies compactified R^3 with S^3. (Figure from hep-th/0010225, courtesy of Falk Bruckmann.)

to the compactification $R^2 \to S^2$ of the complex plane by Riemann's stereographic projection, as shown in Fig. 3.1. As a consequence, any residual gauge transformation $U(\vec{x})$ defines a map

$$S^3_{space} \to SU(3)_{color}\,. \tag{3.10}$$

For the purpose of the following topological (homotopy) classification, this map can be furthermore restricted to a $SU(2)$ subgroup of the gauge group, since according to a powerful theorem by Raoul Bott [27] only this subgroup will be "topologically active" in our discussion. Moreover, the quarternion representation of any $U \in SU(2)$,

$$U = u_0 + i u_a \tau_a \qquad \text{with} \qquad u_\alpha \text{ real and } \quad u_0^2 + u_a u_a = 1, \tag{3.11}$$

(in terms of the Pauli matrices τ_a) shows that $SU(2)$ can be mapped onto the 3-sphere S^3_{group} (which is its group manifold). Thus the map (3.10) is topologically (in the sense of homotopy theory) equivalent to

$$S^3_{space} \to S^3_{group} \tag{3.12}$$

i.e. any residual gauge transformation maps the spacial 3-sphere into the group 3-sphere. In general, two such mappings $U(\vec{x})$ cannot be continuously deformed into each other, which implies that they fall into different homotopy classes. The same holds for spheres of any dimension n, i.e. for any map $S^n \to S^n$.

To get a visual understanding of this topological classification, it is thus sufficient to consider just the simplest case

$$S^1_{sp} \to S^1_{gr} \tag{3.13}$$

of maps between two circles. Such mappings can equivalently be thought of as phase fields $U(\chi) = e^{i\phi(\chi)} \in U(1)$ defined on a circle with coordinate angle $\chi \in [0, 2\pi]$. In Fig. 3.2 various maps of this type are graphically represented by drawing the domain space circle (which can be imagined as a rubber band) in contact with the target space (the dotted circle which represents the phase of the field) such that the mapping occurs between those points on the circles which touch each other. Note that we consider only continuous maps which implies that every point of the target circle must somewhere touch the domain circle. The examples drawn in Fig. 3.2 illustrate the topological properties of such maps. We start with Fig. 3.2a where the whole domain space shrinks to a point (which we draw as a tiny circle with radius $r \to 0$ implied) and touches the target space at one and the same point. Clearly this graph represents the constant map: all points of "space" are mapped into one and the same point in the field or target space.

Let us now consider Fig. 3.2b where the mapping (field) becomes space-dependent. We define that fields which can be continuously deformed into each other (i.e. such that the all points of the domain space stay in contact with points on the domain circle) belong to the same "homotopy class". Now the map of Fig. 3.2b

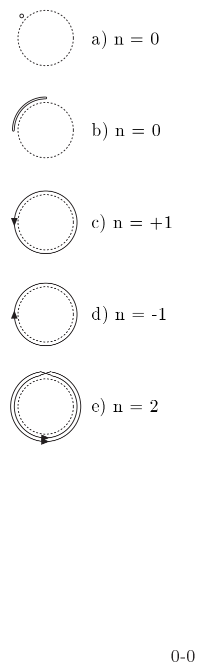

Figure 3.2: Some characteristic examples of the mapping $S^1 \to S^1$, $\exp(i\chi) \mapsto \exp(i\phi(\chi))$. The inner, dotted circle is the range of values of ϕ, while the outer, solid “circles“ symbolize the domain of χ.

can clearly be continuously deformed into the identity map and thus belongs to the same homotopy class. This is in distinct contrast to the remaining maps of Fig. 3.2. In Fig. 3.2c the domain space wraps once counterclockwise around the target circle and therefore cannot be continuously shrunk to a point. Therefore, the corresponding map lies in a disjoint homotopy class with which we will associate a "winding number" $n = +1$. The mapping of Fig. 3.2d wraps clockwise[2] around the target circle and belongs to the homotopy class with winding number $n = -1$. (The trivial class containing the identity map obviously has $n = 0$.)

The next graph, Fig. 3.2e, shows a map with winding number $n = +2$ where the target space wraps twice counterclockwise around the domain space. Now it should be clear how this classification proceeds to higher winding numbers. We have thus made plausible that the maps (3.13) fall into an enumerable infinity of disjoint homotopy classes characterized by an integer winding number $n \in Z$. If one additionally defines a composition law for such maps by concatenation (under which the winding numbers of the maps to be composed simply add) the homotopy classes become elements of a group, the so-called homotopy group

$$\pi_1\left(S^1\right) = Z. \tag{3.14}$$

Here the subscript of π indicates the dimension of the domain sphere (in the present case equal to one), and its argument denotes the target space (here S^1). Now, according to what we have said before this result generalizes to the homotopy groups for mappings between spheres of dimension d, $\pi_d\left(S^d\right) = Z$, which includes the maps (3.12) which interest us in the context of the QCD vacuum. We thus conclude that

$$\pi_3\left(S^3\right) = Z. \tag{3.15}$$

Let us summarize what we have learned so far: the pure-gauge fields (3.6), constructed from gauge transformations $U^{(n)}(x)$ which satisfy the boundary conditions (3.9), minimize the Euclidean QCD action and fall into disjoint homotopy classes characterized by an integer winding number n which derives from the topological properties of the $U^{(n)}(x)$. Thus the topology of the gauge group on the compactified space S^3 induces the semiclassical vacuum structure with its periodic "potential" (action).

Note that the above arguments do *not* prevent us from transforming a pure-gauge field of class n, i.e.

$$G_\mu^{(n)} = \frac{-i}{g} U^{(n)} \partial_\mu U^{(n)\dagger}, \tag{3.16}$$

by continuous deformation into one of class $m \neq n$, despite the fact that we have just shown that we cannot continuously deform $U^{(n)}$ into $U^{(m)}$. Indeed, the latter implies only that we cannot go from $G_\mu^{(n)}$ to $G_\mu^{(m)}$ without leaving pure gauge, i.e. we cannot keep the field always in the form (3.6). This also means that we cannot stay in the sector of fields with zero action (recall that the pure gauges are the only gluon fields which minimize the Euclidean action). In other words, while continuously deforming $G_\mu^{(n)}$ into $G_\mu^{(m)}$ with $n \neq m$ we necessarily encounter field configurations with non-minimal action, $S_E > 0$: each topological class corresponds to an absolute action minimum, and those minima are separated by a finite-action barrier (the so-called sphaleron barrier). This situation is depicted in Fig. 3.3.

3.1.2 Yang-Mills instanton solution

The reader has probably already noted the analogy between the degenerate classical ground state found in Section 3.1.1 and the periodic potential discussed in Section 2.4.2. In both cases, the Euclidean action has an enumerable infinity of degenerate minima. And as in the quantum mechanical examples, the degeneracy between the classical "would-be" or "candidate" vacua of QCD is lifted by tunneling of gluons through the finite action barriers.

Thus we know already, at least in principle, how to obtain the semiclassical approximation to the tunneling amplitude between the classical minima $G_\mu^{(n)}$: it is generated by a saddle-point approximation to the

[2] Note that the direction in which a point on the curve proceeds if the curve parameter increases does not change under continuous deformations.

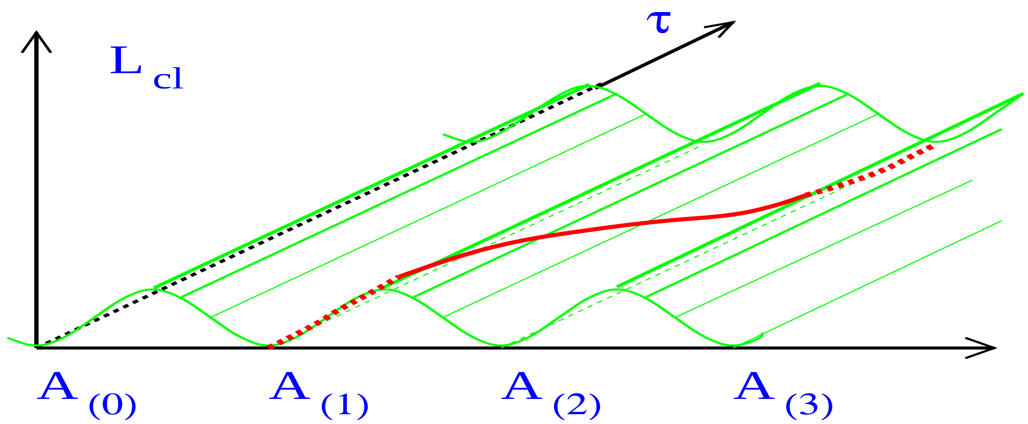

Figure 3.3: The classical, Euclidean QCD Lagrangian for different classes of gauge field configurations. The degenerate absolute minima correspond to pure gauges $A_{(n),\mu}\left(\equiv G_\mu^{(n)}\right)$ with winding number n. An instanton trajectory, interpolating between $A_{(1)}$ and $A_{(2)}$, is also shown.

Euclidean functional integral around stationary (and approximately stationary) field configurations, i.e. the solutions of the Euclidean Yang-Mills equation

$$\frac{\delta S}{\delta G_\nu} = \partial_\mu G_{\mu\nu} + g\left[G_\mu, G_{\mu\nu}\right] \equiv D_\mu G_{\mu\nu} = 0 \tag{3.17}$$

with the boundary conditions

$$G_\mu\left(\vec{x}, T = -\infty\right) = G_\mu^{(n)}\left(\vec{x}\right), \tag{3.18}$$

$$G_\mu\left(\vec{x}, T = +\infty\right) = G_\mu^{(m)}\left(\vec{x}\right). \tag{3.19}$$

According to the above boundary conditions, the solutions start out at $T \to -\infty$ as pure-gauge fields constructed from a gauge transformation with winding number n, and they end up in a pure gauge associated with winding number m at $T \to +\infty$. Moreover, they do by "paying" the minimal amount of action possible. These solutions are the QCD-instantons, and they can be found analytically [1, 13]. For $m = n + 1$, the instanton (in a non-singular gauge) has the explicit form

$$G_\mu^{(I)}\left(x\right) = \frac{2}{g}\frac{\eta_{a\mu\nu}\left(x_\nu - z_\nu\right)t^a}{\left(x - z\right)^2 + \rho^2}. \tag{3.20}$$

An instanton with $n = 1$, $m = 2$ is drawn in Fig. 3.3.

A couple of remarkable features can be read off directly from the instanton solution (3.20): first, its "spin" (i.e. the Lorentz vector index) is coupled to the color orientation by the 't Hooft symbol $\eta_{a\mu\nu}$. Second, its nonperturbative character (as expected in the context of tunneling) reveals itself in the diverging weak-coupling limit. (The $1/g$ behavior is common to all solutions of the classical Yang-Mills equation (3.17), as can be seen by rescaling the gluon field.) Finally, the solution makes explicit that the tunneling process is localized in a region of size ρ in space and time, around its center z.

By using the definition

$$\eta_{a\mu\nu} = \delta_{a\mu}\delta_{4\nu} - \delta_{a\nu}\delta_{4\mu} + \varepsilon_{a\mu\nu} \tag{3.21}$$

of the 't Hooft symbol, it is easy to verify that the instanton (3.20) indeed satisfies the boundary conditions (3.18), (3.19) with $n = 0$. Thus, QCD instantons describe localized, flash-like rearrangements of the vacuum which mediate tunneling processes between topologically distinct pure-gauge sectors.

In analogy with the quantum-mechanical example of the periodic potential, those tunneling processes lift the degeneracy between the pure-gauge "would-be" vacua $|n\rangle$. Instead, superpositions of all those states, the "theta vacua"

$$|0\rangle_\theta = \mathcal{N} \sum_n e^{i\theta n} |n\rangle , \tag{3.22}$$

become the gauge-invariant eigenstates. Although part of our above discussion (including the tunneling interpretation) depended on the gauge choice (3.8), the ensuing vacuum structure is therefore gauge-independent, too. We will further elaborate on the θ vacuum structure in Section 3.4. Here we only note that the analogy with the Bloch waves of the quantum-mechanical periodic potential is limited. While the symmetry generators in the periodic potential (i.e. the generators of periodic translations) are physical, they just generate gauge transformations of nontrivial topology in the Yang-Mills case. Moreover, the analog of the Bloch momentum, the angle θ, classifies different theories and cannot be changed inside a given "θ-world". At present, the value of θ can only be determined phenomenologically. This task is made easier by the fact that finite values of θ induce CP-violating amplitudes (see Section 3.4.2).

Evidence from instanton-liquid vacuum models suggests that approximate saddle points, namely superpositions of instantons and anti-instantons [38, 3], dominate the SCA to the generating functional of QCD. Such superpositions are approximate solutions of the Yang-Mills equation if the typical separation between the (anti-) instantons is much larger than their average size, a condition which seems to hold in the QCD vacuum (see below). In contrast to the quantum mechanical examples of Sections 2.2 and 2.4, however, the classical interactions between Yang-Mills instantons are of dipole type (at large separation) and therefore of much longer range than the exponentially suppressed overlaps between quantum mechanical instantons (cf. Eq. (2.56)). As a consequence, the dilute instanton gas approximation fails in QCD. The strong correlations among QCD-instantons generate an ensemble which probably resembles more a liquid than a gas.

The size distribution $n(\rho)$ of instantons in the vacuum is currently under active investigation, mainly on the lattice [6]. The obtained results are not yet fully consistent with each other but conform more or less to the older results of instanton vacuum models [3]. (Some calculations have led to a larger density which, however, is difficult to measure reliably on the lattice.) For the lowest moments of the distribution, the standard values are

$$\bar{n} = \int d\rho n(\rho) \sim 1 \text{ fm}^{-4}, \tag{3.23}$$

$$\bar{\rho} = \frac{1}{\bar{n}} \int d\rho \rho n(\rho) \sim \frac{1}{3} \text{ fm}. \tag{3.24}$$

Thus, in the QCD vacuum tunneling happens on average in about 5% of spacetime, and it happens very rapidly: the action barrier is penetrated almost instantaneously ($\tau_{tunnel} \sim 0.3 \times 10^{-15}\text{m} \sim 10^{-24}\text{s}$), therefore the name "instanton".

3.2 Including quarks

Some of the most striking effects of QCD instantons are associated with the existence of light quarks, and not even a terse introduction to instanton physics like the present one would be complete without mentioning at least a few prominent examples. In the following two sections we will therefore try to give a flavor of the most important quark-related instanton effects.

3.2.1 The chiral anomaly

Behind many essential instanton effects in the light-quark sector lurks the axial anomaly, i.e. the fact that the (flavor-singlet) axial quark current, which is conserved in the classical theory, ceases to be so at the quantum level. In the following we will derive this anomaly and some of its implications[3]. Also from a more

[3]A transparent introduction to QCD anomalies (and low-energy theorems) can be found in [28].

general point of view, the discussion of anomalies fits well into our context since it exhibits a particular strength of the semiclassical approximation: all known anomalies are fully determined by one-loop Feynman diagrams, and as such appear exclusively at $O(\hbar)$ of the SCA.

To set the stage for the following discussion, let us switch on the fermionic part of the QCD action,

$$S_q[q,\bar{q},G] = \int d^4x\bar{q}\left[i\gamma_\mu\left(\partial_\mu + igG_\mu\right)\right]q, \tag{3.25}$$

which couples the previously considered gluon fields to light quarks (we omit the small quark mass term). The Noether procedure and the Dirac equation imply that this action has a classically conserved axial current, i.e.

$$\partial_\mu\left(\bar{q}i\gamma_\mu\gamma_5 q\right) = 0. \tag{3.26}$$

(Including the factor i is a convenient convention in Euclidean spacetime.) Now let us compute the quantum expectation value of this current in the background of a fixed, classical gauge field[4] G_μ, i.e.

$$J_{5,\mu}(x|G) = Z^{-1}[G]\int\mathcal{D}q\mathcal{D}\bar{q}\left(\bar{q}i\gamma_\mu\gamma_5 q\right)e^{-S_q[q,\bar{q},G]} \tag{3.27}$$

where

$$Z[G] = \int\mathcal{D}q\mathcal{D}\bar{q}e^{-S_q[q,\bar{q},G]}. \tag{3.28}$$

In order to perform the standard Grassmann integration over the quark fields in (3.27), we introduce Grassmann-valued sources $\eta,\bar{\eta}$ and write

$$J_{5,\mu}(x|G) = Z^{-1}[G]\left(\frac{-\delta}{\delta\eta}i\gamma_\mu\gamma_5\frac{\delta}{\delta\bar{\eta}}\right)\int\mathcal{D}q\mathcal{D}\bar{q}e^{-\int d^4x[\bar{q}\mathcal{D}q+\bar{\eta}q+\bar{q}\eta]}\Big|_{\eta,\bar{\eta}=0} \tag{3.29}$$

where we have defined the Euclidean Dirac operator

$$\mathcal{D} = i\gamma_\mu\left[\partial_\mu + igG_\mu(x)\right]. \tag{3.30}$$

With the help of its inverse, the Dirac propagator $S(x,y|G)$ in the background field G_μ, defined by

$$\left[i\gamma_\mu\left(\partial_{x,\mu} + igG_\mu(x)\right)\right]S(x,y|G) = \delta^4(x-y), \tag{3.31}$$

we can rewrite the exponent in the integrand of (3.29) as

$$\bar{q}\mathcal{D}q + \bar{\eta}q + \bar{q}\eta = \left(\bar{q}+\bar{\eta}S\right)\mathcal{D}\left(q+S\eta\right) - \bar{\eta}S\eta \tag{3.32}$$

and, after redefining $q \to q' = q + S\eta$ (this is a constant shift of the functional variable q which leaves the measure invariant), we obtain

$$\int\mathcal{D}q\mathcal{D}\bar{q}e^{-\int d^4x[\bar{q}\mathcal{D}q+\bar{\eta}q+\bar{q}\eta]} = Z[G]\,e^{\int d^4xd^4y[\bar{\eta}(x)S(x,y|G)\eta(y)]}. \tag{3.33}$$

Inserting this identity into Eq. (3.29), we find

$$J_{5,\mu}(x|G) = -N_f tr_{\gamma,c}\left\{i\gamma_\mu\gamma_5 S(x,x|G)\right\} \tag{3.34}$$

(N_f is the number of quark flavors; the trace is over Dirac and color indices). This expression is still formal: the singular coincidence limit of the propagator has to be regularized, and we will do so below. In terms of Feynman graphs, Eq. (3.34) states that the expectation value of the axial current can be calculated from a closed quark loop in the background of a (classical) gluon field G, with one insertion of the axial vertex $\gamma_\mu\gamma_5$.

[4]Note that this is the first step towards a full quantum calculation, where we would afterwards integrate over all classical gauge fields. In order to derive the anomaly, however, it is sufficient to quantize just the fermion sector.

For the further evaluation of (3.34) we employ the spectral representation of $S(x,y|G)$ in terms of the normalized eigenfunctions ψ_n of the Dirac operator $\mathcal{D}$,

$$\mathcal{D}\psi_n(x) = \lambda_n \psi_n(x), \tag{3.35}$$

($\mathcal{D}$ is hermitean ($\gamma_\mu^+ = \gamma_\mu$ and $G_\mu^+ = G_\mu$) so that its spectrum is real) which reads

$$S(x,y|G) = \sum_n \frac{\psi_n(x)\bar{\psi}_n(y)}{\lambda_n}. \tag{3.36}$$

(For the moment we do not worry about vanishing eigenvalues, although they will play a prominent role later on[5].) In order to regularize this infinite sum we introduce a cutoff function

$$f_\varepsilon\left(\lambda_n^2\right) = e^{-\varepsilon\lambda_n^2} \tag{3.37}$$

which suppresses the contributions from large $|\lambda_n|$ and goes to unity in the limit $\varepsilon \to 0$ which we are going to take in the end. Inserted into (3.34) this yields

$$J_{5,\mu}(x|G) = -N_f \sum_n \frac{\bar{\psi}_n(x)\, i\gamma_\mu\gamma_5\psi_n(x)}{\lambda_n} e^{-\varepsilon\lambda_n^2}. \tag{3.38}$$

Let us now obtain the divergence of this current by noting that (3.35) implies

$$\partial_\mu\left(\bar{\psi}_n i\gamma_\mu\gamma_5\psi_n\right) = 2\lambda_n\left(\bar{\psi}_n\gamma_5\psi_n\right), \tag{3.39}$$

and leads to

$$\partial_\mu J_{5,\mu}(x|G) = -2N_f \sum_n \bar{\psi}_n\gamma_5 e^{-\varepsilon\lambda_n^2}\psi_n = -2N_f \sum_n \bar{\psi}_n\gamma_5 e^{-\varepsilon\mathcal{D}^2}\psi_n = 2N_f tr_{\gamma,c}\left\langle x\left|\gamma_5 e^{-\varepsilon\mathcal{D}^2}\right|x\right\rangle \tag{3.40}$$

where

$$\mathcal{D}^2 = -\left(\partial_\mu + igG_\mu\right)^2 - \frac{g}{2}\sigma_{\mu\nu}G_{\mu\nu} \tag{3.41}$$

($\sigma_{\mu\nu} = (i/2)\left[\gamma_\mu, \gamma_\nu\right]$).

We can now expand the exponential[6] in powers of the classical background field G. In Feynman-diagram language, this corresponds to expanding the quark loop in the number of G insertions. Each additional G insertion implies an additional quark propagator in the loop and therefore makes the loop integral converge faster at high momenta. Thus there can be at most a few UV-divergent terms, associated with the lowest orders of G. These terms require regularization, and it is at this point where the classical axial symmetry gets broken (if we insist on keeping the vector current conserved) and the anomaly emerges.

The above reasoning is reflected in our calculation by the fact that only the leading nonvanishing term in the G expansion of (3.40) will survive the limit $\varepsilon \to 0$ which we are going to take in the end (since higher-order terms in the expansion of the exponential (3.40) also introduce higher orders of ε). We therefore just have to identify and calculate this single contribution. Since all the terms from the first part of (3.41) vanish due to $tr_\gamma(\gamma_5) = 0$, and since those from the first-order contribution of the second term vanish due to $tr_\gamma(\gamma_5\sigma_{\mu\nu}) = 0$, the first nonvanishing contribution originates from the second-order piece of the $\sigma_{\mu\nu}$-term, i.e.

$$tr_{\gamma,c}\left\langle x\left|\gamma_5 e^{-\varepsilon\mathcal{D}^2}\right|x\right\rangle = \left\langle x\left|e^{\varepsilon\partial^2}\right|x\right\rangle\left[\frac{g^2\varepsilon^2}{8} tr_\gamma\left(\gamma_5\sigma_{\mu\nu}\sigma_{\rho\sigma}\right) tr_c\left(G_{\mu\nu}G_{\rho\sigma}\right) + O\left(\varepsilon^3\right)\right], \tag{3.42}$$

which corresponds to a quark loop with two (vector) interactions with the background gluon field (cf. Fig. 3.4). With

[5]Alternatively, one could regularize the divergence they cause in (3.36) by introducing a small quark mass.

[6]Note that the G-dependence of (3.40) originates solely from the exponential since the implicit G-dependence of the ψ_n cancels in the sum. (This is because the matrix element $\left\langle x\left|\gamma_5 e^{-\varepsilon\mathcal{D}^2}\right|x\right\rangle$ is independent of the basis of functions in which one may decide to evaluate it.)

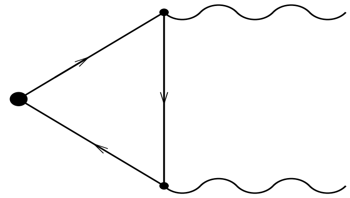

Figure 3.4: The triangle graph correponding to the UV-divergent quark loop which generates the chiral anomaly. The bigger blob denotes the insertion of the axial-vector vertex $\gamma_\mu \gamma_5$, the smaller one the insertion of the vector vertex γ_μ to which the gluons are attached.

$$tr_\gamma \left(\gamma_5 \sigma_{\mu\nu} \sigma_{\rho\sigma}\right) = 4\varepsilon_{\mu\nu\rho\sigma}, \qquad \tilde{G}_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma} G_{\rho\sigma} \tag{3.43}$$

and

$$\left\langle x \left| e^{\varepsilon \partial^2} \right| x \right\rangle = \int d^4p \left\langle x|p\right\rangle e^{\varepsilon \partial^2} \left\langle p|x\right\rangle = \int \frac{d^4p}{(2\pi)^4} e^{-\varepsilon p^2} = \frac{1}{16\pi^2 \varepsilon^2} \tag{3.44}$$

(note that this term contains the UV singularity for $\varepsilon \to 0$), where we have used $\langle x|p\rangle = \exp\left(-ip_\mu x_\mu\right) / (2\pi)^2$, we finally obtain (in the limit $\varepsilon \to 0$)

$$\partial_\mu J_{5,\mu}\left(x|G\right) = \frac{g^2 N_f}{8\pi^2} tr_c \left(G_{\mu\nu} \tilde{G}_{\mu\nu}\right). \tag{3.45}$$

This famous result (which can be shown to be independent of the regularization procedure) expresses the gist of the "axial anomaly": although the axial current is conserved in classical chromodynamics, the corresponding symmetry is broken at the quantum level, i.e. in QCD, due to the necessity of regularization and renormalization[7].

3.2.2 Quark zero modes and index theorem

The existence of the axial anomaly has crucial implications for fermions in the background of instanton fields. In order to exhibit them, let us take a closer look at the Dirac spectrum

$$\mathcal{D}\psi_n\left(x\right) = \lambda_n \psi_n\left(x\right) \tag{3.46}$$

introduced in the last section. Since the (hermitean) Dirac operator anticommutes with γ_5,

$$\left\{\mathcal{D}, \gamma_5\right\} = 0, \tag{3.47}$$

each (normalized) eigenfunction ψ_n with $\lambda_n \neq 0$ implies the existence of another eigenfunction

$$\psi_{-n}\left(x\right) \equiv \gamma_5 \psi_n\left(x\right) \tag{3.48}$$

with eigenvalue

$$\lambda_{-n} = -\lambda_n, \qquad (n > 0). \tag{3.49}$$

In other words, nonvanishing eigenvalues appear in pairs with opposite signs. The eigenfunctions $\psi_{0,k}$, corresponding to the remaining eigenvalues $\lambda_k = 0$, are called zero modes. The zero modes can be made to diagonalize γ_5 (due to (3.47)) by transforming to the chiral basis

$$\psi_{0\pm,k} = \frac{1}{2}\left(1 \pm \gamma_5\right) \psi_{0,k} \tag{3.50}$$

[7]Readers who are familiar with the original derivation of the anomaly in terms of explicit Feynman graphs will have noted that we just calculated the celebrated triangle graph Fig. 3.4 in a somewhat implicit, functional fashion. The functional approach allowed us to get the desired result without introducing the machinery of perturbative quantum field theory (i.e. Feynman rules).

so that

$$\gamma_5 \psi_{0\pm,k} = \pm \psi_{0\pm,k}. \tag{3.51}$$

Since, finally, the QCD Dirac operator is flavor-independent, i.e.

$$[\mathcal{D}, \mathcal{T}_a] = 0, \tag{3.52}$$

(where $\mathcal{T}_a$ are the generators of the flavor group) each zero mode comes in N_f copies (one of each flavor).

The zero modes $\psi_{0,\pm}$ are normalizable and inherit several characteristic properties of the instanton, including the localization in space and time and the coupling of spin and color. In the so-called singular gauge, their explicit form [31] is

$$\psi_{0,\pm}(x) = \frac{\rho}{\pi} \frac{1 \pm \gamma_5}{(r^2 + \rho^2)^{3/2}} \frac{\not{r}}{r} u. \tag{3.53}$$

(The spin-color coupling matrix u is defined by $(\vec{\sigma} + \vec{\tau})\, u = 0$, and $r = x - x_0$.) For our subsequent discussion only qualitative properties of the zero modes will matter, though.

The two main characteristics of the Dirac spectrum established above - paired eigenvalues of opposite sign and chiral zero modes - have profound consequences. One of those becomes explicit when one integrates the anomaly equation (3.45) over spacetime and uses Eq. (3.40) (multiplied by a factor $1/(2N_f)$) to write

$$Q \equiv \frac{g^2}{16\pi^2} \int d^4x tr_c \left(G_{\mu\nu} \tilde{G}_{\mu\nu} \right) \tag{3.54}$$

$$= -\sum_n e^{-\varepsilon \lambda_n^2} \int d^4x \bar{\psi}_n \gamma_5 \psi_n. \tag{3.55}$$

Above we have defined the "topological charge" Q whose meaning will be clarified below. Since the eigenfunctions ψ_n and $\gamma_5 \psi_n$ of $\mathcal{D}$ have different eigenvalues (λ_n and $-\lambda_n$) and are thus orthogonal to each other, the integral in (3.55) vanishes for all n with $\lambda_n \neq 0$, leaving us with an integral over the zero modes only:

$$Q = -\sum_k \int d^4x \bar{\psi}_{0,k} \gamma_5 \psi_{0,k} \tag{3.56}$$

$$= \sum_{m=1}^{n_-} \int d^4x \bar{\psi}_{0-,m} \psi_{0-,m} - \sum_{m=1}^{n_+} \int d^4x \bar{\psi}_{0+,m} \psi_{0+,m} \tag{3.57}$$

where $n_\pm$ is the number of right- (left-) handed zero modes per flavor in the given background gluon field. Since the eigenfunctions of $\mathcal{D}$ (including the zero modes) are normalized, we finally obtain

$$Q = n_- - n_+. \tag{3.58}$$

The above formula, relating a property $Q[G]$ of the gauge field to the number of unpaired quark zero modes of the Dirac operator (3.35) in its background, is a special case of the celebrated Atiyah–Singer index theorem [29, 30] for the Euclidean Dirac operator. It implies that Q must be integer and thus cannot change under smooth variations of the background gluon field. Such a behavior is characteristic for topological properties of the gluon field and prompts us to have a closer look at those in the next section.

Before doing so, however, we should mention a few phenomenological consequences of the instanton-induced zero modes in the quark spectrum. The integrated anomaly equation (see Eq. (3.62) below) reveals that the $Q = 1$ instanton is accompanied by the appearance of $2N_f$ units of axial charge. This can happen either by creating N_f left-handed quarks and annihilating N_f right-handed antiquarks in zero-mode states or, in the alternative (crossed) time-ordering, by helicity flips of quarks propagating in the zero-mode state. Such processes are described by non-local $2N_f$-quark interactions known as 't Hooft vertices [31]. They generate the dominant instanton effects in the light-quark sector and crucial interactions between instantons. As a reflection of the anomaly, the 't Hooft vertices manifestly break the axial $U(1)$ symmetry of the QCD Lagrangian with $m_q = 0$ and thereby resolve (at least qualitatively) the so-called "$U(1)$ problem", i.e. they explain why the η' meson has almost twice the mass of the η meson and cannot be considered a (quasi) Goldstone boson. More generally, the quark interactions mediated by 't Hooft vertices have a characteristic channel dependence which makes them, e.g., strongly attractive in spin-0 quark-antiquark channels but mute in vector and axial-vector channels. These patterns are reflected in the instanton contributions to hadronic amplitudes. We will come back to this issue in the final lecture.

3.3 Instanton topology

So far, we have established two important implications of the axial anomaly: first, in order for the anomaly to have physical significance there must be gauge fields with non-vanishing topological charge Q, and second, such fields imply the existence of unpaired zero modes in the spectrum of the associated Dirac operator. Our next tasks will be to better understand the topological charge, the gluon fields which carry it, and the impact of both on the vacuum structure.

3.3.1 Topological charge

We start by noting that the integrand of the topological charge Q (cf. (3.54)) can be written as a total derivative

$$\frac{g^2}{16\pi^2} tr_c \left(G_{\mu\nu} \tilde{G}_{\mu\nu} \right) = \partial_\mu K_\mu \tag{3.59}$$

(we do not need the explicit expression for the "Chern-Simons current" K_μ here), which implies that only fields with nontrivial behavior at the spacetime boundary $x^2 \to \infty$ (typical for topologically active fields) can carry a finite Q. Furthermore, the index theorem (3.58) implies that Q can only take integer values and cannot therefore change under continuous deformations of the gluon background field. Thus, besides the homotopy classification of the pure gauges encountered in Section 3.1.1, gluon fields carry another topological property specified by their Q-values. As we have seen above, fields with nonvanishing Q "activate" the anomaly by generating quark zero-modes and are therefore, via Eq. (3.45), associated with a non-conservation of the axial charge $q_5(\tau) = \int d^3x J_{5,4}(x|G)$:

$$Q = \frac{g^2}{16\pi^2} \int d^4x tr_c \left(G_{\mu\nu} \tilde{G}_{\mu\nu} \right) = \frac{1}{2N_f} \int d^4x \partial_\mu J_{5,\mu}(x|G) \tag{3.60}$$

$$= \frac{1}{2N_f} \int d^3x \int d\tau \left[\partial_4 J_{5,4}(x|G) + \vec{\nabla} \vec{J}_5(x|G) \right] = \frac{1}{2N_f} \int d^3x J_{5,4}(x|G) \Big|_{\tau=-\infty}^{\tau=\infty} \tag{3.61}$$

$$= \frac{1}{2N_f} \Delta q_5. \tag{3.62}$$

where

$$\Delta q_5 = q_5(\tau = \infty) - q_5(\tau = -\infty) \tag{3.63}$$

is the amount of axial charge created by the gluon field with topological charge Q.

We will now show that an integer value of Q can be assigned to any gluon field with finite (Euclidean) action (only those are relevant for the SCA), and that this value has a transparent physical interpretation. For the action (3.1) to be finite, the field strength $G_{\mu\nu}$ has to be square integrable, i.e. it must vanish towards the boundary $S^3(x^2 = \infty)$ of Euclidean spacetime, i.e. for $|x| \to \infty$, faster than $1/x^2$. Now we know from Section 3.1.1 that the Euclidean action has its absolute minimum at $G_{\mu\nu} = 0$, i.e. for pure gauges. As a consequence, finite-action fields satisfy the boundary condition

$$\lim_{x^2 \to \infty} G_\mu = \frac{-i}{g} U^{(n)}(x) \, \partial_\mu U^{(n)-1}(x) \tag{3.64}$$

where $U^{(n)}$ is an element of the gauge group $\mathcal{G}$ with winding number n. (Not surprisingly, instantons are of this type (cf. (3.18), (3.19)) as they interpolate between pure gauges of different winding number for $\tau \to \pm\infty$.)

The essential consequence of Eq. (3.64) is that it implies a time–independent topological classification of all finite–action gauge fields. This classification is based on (but not equal to) the homotopy analysis of the gauge group elements $U^{(n)}$ which we went through in Section 3.1.1. Indeed, Eq. (3.64) implies that any finite–action gauge field defines a map $S^3(x^2 = \infty) \to \mathcal{G}$, and such maps fall into disjoint homotopy classes according to the third homotopy group[8] $\pi_3(\mathcal{G})$ of $\mathcal{G}$. For $\mathcal{G} = SU(N)$ with $N \geq 2$, which includes QCD, we have $\pi_3(SU(N)) = Z$ as in Section 3.1.1 while, e.g., (non-compact) QED has $\pi_3(U(1)) = 1$. As a consequence, all finite-action gluon fields in Euclidean QCD fall into topologically distinct equivalence classes labeled by the topological charge Q (which is technically the "Pontryagin index" of the gauge field)[9].

For instantons, the topological charge Q has a particularly transparent meaning: it is just the difference between the winding numbers n and m of the pure gauges between which the instanton interpolates,

$$Q = m - n. \tag{3.65}$$

Instead of proving this expression in general (which would not be difficult), let us check it for the explicit instanton solution (3.20) found above. We first calculate the classical instanton action (from which we know already that it is finite and that it is minimal in its Q-sector) by plugging the field strength

$$G_{I,\mu\nu}(x) = \frac{-4\rho^2}{g} \frac{\eta_{a\mu\nu} t_a}{[(x - x_0)^2 + \rho^2]^2} \tag{3.66}$$

(obtained by inserting the expression (3.20) for the instanton into the Yang-Mills field strength tensor (3.3)) into the Yang-Mills action (3.1). The result is

$$S_I = \frac{1}{2} \int d^4x tr_c (G_{I,\mu\nu} G_{I,\mu\nu}) = \frac{8\pi^2}{g^2}. \tag{3.67}$$

We note that the classical action is additive for multi-instantons and multi-anti-instantons, $S_Q = |Q| \times S_I$ (but not for combinations of both), and that it does not depend on the instanton's collective coordinates z_μ, U and ρ. The latter is a direct consequence of the translation, gauge, and scale invariance of classical Yang-Mills theory. At the quantum level, the scale invariance gets broken by the trace anomaly (yet another variety of anomaly) and the effective quantum action becomes ρ-dependent. This generates a ρ-dependent weight $n(\rho)$ in functional integrals which can be identified with the instanton size density introduced at the end of Section 3.1.2.

Having calculated the instanton action, we can easily check whether the instanton has indeed $Q = 1$. In fact, this becomes trivial once we have established another essential instanton property, namely the self-duality of its field strength. We now turn to this issue.

3.3.2 Self duality

Inspection of the instanton's field strength (3.66) reveals that it depends on the Lorentz indices only via the 't Hooft symbol $\eta_{a\mu\nu}$. Moreover, the expression (3.21) for $\eta_{a\mu\nu}$ shows that it is self-dual (recall the definition (3.43)), i.e.

$$\tilde{\eta}_{a\mu\nu} = \frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} \eta_{a\rho\sigma} = \eta_{a\mu\nu}, \tag{3.68}$$

which directly carries over to the field strength of the instanton:

$$G_{I,\mu\nu} = \tilde{G}_{I,\mu\nu}. \tag{3.69}$$

Note that self-duality is a gauge invariant property. It implies $E_i^a = B_i^a$, i.e. the equality of the chromoelectric and chromomagnetic fields (3.4), and yields immediately

$$Q_I = \frac{g^2}{16\pi^2} \int d^4x tr_c \left(G_{I,\mu\nu} \tilde{G}_{I,\mu\nu} \right) = \frac{g^2}{16\pi^2} \int d^4x tr_c (G_{I,\mu\nu} G_{I,\mu\nu}) = \frac{g^2}{8\pi^2} S_I = 1. \tag{3.70}$$

[8] In mathematical terms, $\pi_3(\mathcal{G})$ classifies principal $\mathcal{G}$ fibre bundles over S^4 [30].

[9] At this point, it might be useful to emphasize the differences between the topological classification discussed here and that of the pure gauges in section 3.1.1. The winding number n of the latter classifies maps from compactified space S^3_{space} (on which static gauge transformations with unit-value at the boundary are defined) into the gauge group, while Q characterizes maps from the boundary S^3 of Euclidean spacetime into the gauge group.

The anti-instanton is anti-selfdual, i.e. $G_{\bar{I},\mu\nu} = -\tilde{G}_{\bar{I},\mu\nu}$, and thus has $Q_{\bar{I}} = -1$. (Anti-) Self-duality is a mathematically very powerful property. Self-dual fields automatically satisfy, for instance, the Yang-Mills equation (3.17). This is immediately obvious when combining the self-duality equation (3.69) with the Bianchi identity

$$D_\mu \tilde{G}_{\mu\nu} = 0 \tag{3.71}$$

which holds for any gauge field, due to the form (3.3) of the field strength tensor. In practice, it is often advantageous to deal with Eq. (3.69), instead of with the Yang-Mills equation itself, since it is of first order. As a matter of fact, the original instanton solution was found in [1] by solving (3.69). Similarly, in the quantum mechanical example of Section 2.2.1 we derived the instanton solution by solving the corresponding first-order equation (2.48). Self-duality has even more profound mathematical implications, e.g. in Donaldson theory. It can be shown, incidentally, that all minima of the Euclidean Yang-Mills action with $S_E < \infty$ are (anti-) self-dual, and that they correspond to (multi-) instanton solutions [13].

The index theorem (3.58) implies that there must be one unpaired, left-handed zero-mode per flavor in a gluon field with $Q = 1$. This result can be sharpened for (anti-) selfdual fields with $G^{(\pm)}_{\mu\nu} = \pm\tilde{G}^{(\pm)}_{\mu\nu}$. In order to decouple left- and right-handed modes we consider the eigenvalue equation of the iterated Dirac operator (cf. Eq. (3.41)),

$$\mathcal{D}^2 \psi_n = \left[-\left(\partial_\mu + igG_\mu\right)^2 - \frac{g}{2}\sigma_{\mu\nu}G_{\mu\nu} \right] \psi_n = \lambda_n^2 \psi_n. \tag{3.72}$$

For (anti-) selfdual fields the second term in the square bracket can be rewritten with the help of the identity

$$\sigma_{\mu\nu} = -\gamma_5 \tilde{\sigma}_{\mu\nu} \tag{3.73}$$

as

$$\sigma_{\mu\nu}G^{(\pm)}_{\mu\nu} = \mp\sigma_{\mu\nu}G^{(\pm)}_{\mu\nu}\gamma_5 = \sigma_{\mu\nu}G^{(\pm)}_{\mu\nu}\frac{1}{2}\left(1 \mp \gamma_5\right). \tag{3.74}$$

Restricted to quark zero modes and to (anti-) selfdual gluon fields, Eq. (3.72) then becomes

$$\mathcal{D}^2 \psi_0 = \left[-\left(\partial_\mu + igG^{(\pm)}_\mu\right)^2 - \frac{g}{2}\sigma_{\mu\nu}G^{(\pm)}_{\mu\nu}\frac{1}{2}\left(1 \mp \gamma_5\right) \right] \psi_0 = 0. \tag{3.75}$$

For selfdual fields, this implies

$$-\left(\partial_\mu + igG^{(+)}_\mu\right)^2 \psi_{0,+} = 0, \tag{3.76}$$

and since $-\left(\partial_\mu + igG^{(+)}_\mu\right)^2$ has a positive spectrum we conclude that in a selfdual field $\psi_{0,+} = 0$ and thus $n_+ = 0$. Then the index theorem (3.58) tells us that there exists exactly one left-handed (right-handed) zero mode per flavor in the background of an instanton (anti-instanton), and none of the opposite chirality.

3.4 The angle θ

In the following sections we are going to elaborate on the instanton-induced vacuum structure and the θ-angle which we have encountered in Section 3.1.1. Besides the semiclassical approach and its description of vacuum tunneling discussed above, there are several other and partially complementary angles from which one can study the emergence of θ and the implied superselection rule for non-communicating (i.e. physically independent) Fock spaces. Below we will consider two such approaches in more detail. Both of them lead to new insights into the workings of the QCD vacuum, and both of them allow the θ structure to be derived as an exact, formal result, i.e. without recourse to the SCA or, in fact, any approximation. This is reassuring since the conditions under which the SCA works in QCD depend on the physical situation and are often difficult to check in the strongly coupled regime. Possibilities to learn about the consequences of the nontrivial gauge-group topology without relying on the SCA are therefore welcome.

Up to now we have approached QCD in the path integral framework, which provides the most natural setting for the SCA. In the following we are going beyond the SCA, with the aim of augmenting our earlier insights into the theta vacua

$$|\theta\rangle = \mathcal{N} \sum_n e^{-i\theta n} |n\rangle \tag{3.77}$$

as a familiy of (up to a phase) gauge-invariant ground states. Moreover, we will be interested in the general behavior of physical states under topologically nontrivial gauge transformations. It is therefore suggestive to chose a quantization procedure in which the Fock states play a more explicit role. Canonical quantization in the Schrödinger picture will be used for this purpose in Section 3.4.1. Although instantons are indispensable for elucidating the tunneling mechanism behind the θ structure, they will not appear explicitly in this section. In the subsequent Section 3.4.2 we will return to the path integral and look at the θ vacuum from yet another perspective, namely that of the cluster decompositon principle. On the way, we will learn how to properly treat the topological charge Q of the instantons and other topologically nontrivial gauge fields in functional integrals.

3.4.1 θ structure from Gauß' law

We begin this section with a brief summary of the Hamiltonian formulation of QCD in Minkowski space[10] and the canonical quantization of gauge theories. Several alternative methods for this purpose have been developed[11], including e.g. constraint-quantization à la Dirac and the BRST formalism. In our context quantization in Weyl (or temporal) gauge,

$$G_0^a = 0, \qquad G_{0i}^a = \partial_0 G_i^a, \tag{3.78}$$

which we have already encountered in Section 3.1.1, turns out to be the most convenient. To simplify the discussion, we restrict ourselves to pure Yang-Mills theory (i.e. QCD without quarks) with the gauge-fixed Lagrangian

$$\mathcal{L} = -\frac{1}{4} G_{\mu\nu}^a G^{a,\mu\nu} = \frac{1}{2} \left(\partial_0 G_i^a\right)^2 - \frac{1}{4} G_{ij}^a G_{ij}^a \tag{3.79}$$

since the topological effects to be discussed are rooted in the gluon sector.

In terms of the color electric and magnetic fields

$$E_i^a = G_{i0}^a, \qquad B_i^a = -\frac{1}{2}\varepsilon_{ijk} G_{jk}^a. \qquad \left(G_{ij}^a = -\varepsilon_{ijk} B_k^a\right) \tag{3.80}$$

the gauge-fixed Lagrangian (in Minkowski space) is just the difference between the "kinetic" and "potential" gluon energy densities,

$$\mathcal{L} = \frac{1}{2}\left(\vec{E}_a^2 - \vec{B}_a^2\right). \tag{3.81}$$

The canonical variables, in Weyl gauge, are the spacial components G_i^a of the gauge field. In terms of those and the associated conjugate momenta

$$\pi_i^a = \frac{\delta \mathcal{L}}{\delta\left(\partial_0 G_i^a\right)} = G_{0i}^a = -E_i^a \tag{3.82}$$

the Hamiltonian reads

$$H = \int d^3x \left(\pi_i^a \partial_0 G_i^a - \mathcal{L}\right) = \frac{1}{2}\int d^3x \left(\vec{E}_a^2 + \vec{B}_a^2\right). \tag{3.83}$$

Note that, due to the antisymmetry of the field strength, there is no momentum conjugate to G_0^a, i.e. $\pi_0^a = \delta\mathcal{L}/\delta\left(\partial_0 G_0^a\right) \equiv 0$. The corresponding Yang-Mills (i.e. Euler-Lagrange) equation, Gauß' law

$$D_i^{ab} G_{0i}^b = -\vec{D}^{ab}\vec{E}^b = 0, \tag{3.84}$$

[10] A potential disadvantage of the Hamiltonian formulation is that explicit Lorentz invariance is lost because one has to single out a spacelike reference surface on which to impose the canonical commutation relations.

[11] In QED, it can be shown that all quantization schemes lead to the same results, and one expects the same to hold for QCD.

is therefore just a constraint. In other words, G_0 does not propagate since no kinetic term for it appears in the Lagrangian. This is the root of several complications in the canonical quantization of gauge theories. One finds, in particular, that Gauß' law, as an equation at fixed time, does not appear as one of Hamilton's equations. How should one then implement the full Yang-Mills dynamics in the canonical formulation? Certainly not by simply adding Gauß' law as an operator equation: this would be inconsistent because it does not commute with the canonical variables. Instead, it must be imposed as a constraint on the Fock space, restricting the physical states to those eigenstates of H which satisfy

$$D_i^{ab} G_{0i}^b \left|\psi\right\rangle = 0. \tag{3.85}$$

(This is analogous to how one eliminates the non-physical ghost states with negative norm, e.g., in Gupta-Bleuler quantization of QED or in the covariant quantization of strings.)

In order to get a more explicit representation of the Fock space, we will now adopt the Schrödinger picture in which the physical states become functionals $\Psi\left[\mathbf{G}\right]$ of the canonical variables, i.e. of the spacial components of the gauge potential. Accordingly, the $\Psi\left[\mathbf{G}\right]$ are eigenfunctionals of the Hamiltonian[12],

$$H\Psi\left[\mathbf{G}\right] = E\Psi\left[\mathbf{G}\right], \tag{3.87}$$

and satisfy the Gauß law constraint

$$\left(D_i^{ab} G_{0i}^b\right) \Psi\left[\mathbf{G}\right] = 0. \tag{3.88}$$

Let us now establish the appearance of the θ angle as a consequence of the nontrivial topology of the gauge group G (i.e. $\pi_3\left(G\right) = Z \neq 1$) in this framework [32]. The unitary gauge transformations U with

$$G_\mu \to {}^U G_\mu \equiv U\left(x\right)\left[\frac{1}{ig}\partial_\mu + G_\mu\right] U^{-1}\left(x\right) \tag{3.89}$$

are restricted in Weyl gauge to those "residual" ones which leave the gauge condition (3.78) intact, i.e. to the time-independent (and in this sense global) transformations

$$U = e^{i\omega^a(\vec{x})t^a} = U\left(\omega^a\left(\vec{x}\right)\right), \tag{3.90}$$

where the parameters ω^a specify a given group element. The gauge-fixed Hamiltonian (3.83) is still invariant under those transformations. This implies, after quantization, that the Hamilton operator commutes with the operators $\mathcal{U}\left(\omega\left(\vec{x}\right)\right)$ which furnish a unitary representation of the group on the $\Psi\left[\mathbf{G}\right]$:

$$\left[H, \mathcal{U}\left(\omega\left(\vec{x}\right)\right)\right] = 0. \tag{3.91}$$

Thus we choose the physical states $\Psi\left[\mathbf{G}\right]$ to be simultaneously eigenstates of H and $\mathcal{U}$. At first, this might not seem necessary since the operator $D_i^{ab} G_{0i}^b$ appearing in Gauß' law is the generator of infinitesimal gauge transformations[13],

$$i\left[\int d^3x' \omega^a\left(\vec{x}'\right)\frac{1}{ig}D_i^{ab}G_{0i}^b\left(\vec{x}'\right), G_j\left(\vec{x}\right)\right] = \frac{1}{ig}D_i^{ab}\omega^b\left(\vec{x}\right) = \delta_\omega G_j\left(\vec{x}\right). \tag{3.93}$$

Therefore, Eq. (3.88) implies that "small" gauge transformations, defined to be those which can be built from (up to infinitely many) infinitesimal ones and are thus continuously connected to the unit in the

[12]The canonical variables are now the classical fields $G_i^a\left(\vec{x}\right)$ and the corresponding momenta

$$\pi_i^a\left(\vec{x}\right) = -E_i^a\left(\vec{x}\right) = -i\frac{\delta}{\delta G_i^a\left(\vec{x}\right)}, \tag{3.86}$$

so that the usual canonical commutation relations are satisfied. All other (time-independent) operators, including the Hamiltonian, can be constructed from the G_i^a and π_i^a.

[13]Note that for small ω the gauge transformation (3.89) reduces to

$$G_\mu \to G_\mu + \frac{1}{ig}D_\mu\omega + O\left(\omega^2\right). \tag{3.92}$$

gauge group, leave the physical states invariant. However, we have seen in Section 3.1.1 that not all time-independent gauge transformations in QCD belong to this class. Those which do not, the so-called "large" (i.e. topologically nontrivial) gauge transformations, could change the phase of the physical states.

Let us study this issue in more detail. As we have discussed in Section 3.1.1, there is in fact an enumerably infinite number of homotopy classes of gauge transformations (i.e. of the map $S^3 \to S^3$) which are labeled by an integer n. Members of the classes with $n \neq 0$ cannot be continuously connected to the unit element of the group. The classic example of a gauge transformation with $n = 1$ (in the group $SU(2)$ which is embedded in the color gauge group $SU(3)$) is the "hedgehog"

$$U^{(1)}(\vec{x}) = \exp\left(\frac{i\pi x^a \tau^a}{\sqrt{x^2 + c^2}}\right) \tag{3.94}$$

where c is a real constant. Note that the corresponding $\omega^a_{(1)}(\vec{x}) = 2\pi x^a/\sqrt{x^2 + c^2}$ do not vanish at spacial infinity ($|\vec{x}| \to \infty$), that they contain a singularity (here a branch cut), and that the limiting value $U^{(1)}(\vec{x}) \overset{|\vec{x}|\to\infty}{\longrightarrow} -1$ is angle-independent. This is characteristic for topologically nontrivial gauge transformations since the exponential can then neither be smoothly deformed to the unit element nor expanded for all $\vec{x}$. A representative of the class n can be immediately obtained from $U^{(1)}$ by using the additivity of the winding number under group multiplication[14],

$$U^{(n)}(\vec{x}) = \left[U^{(1)}(\vec{x})\right]^n = \exp\left(\frac{in\pi x^a \tau^a}{\sqrt{x^2 + c^2}}\right). \tag{3.95}$$

Thus, although all physical states are annihilated by Gauß' law, $\mathcal{U}$ can have nontrivial eigenvalues for large gauge transformations with $n \neq 0$. (An explicit example is given in Appendix A.) Since the eigenvalues of unitary operators must be phases, we then have

$$\mathcal{U}^{(n)}(\omega(\vec{x}))\Psi[\mathbf{G}] = \Psi\left[{}^{U^{(n)}(\omega(\vec{x}))}\mathbf{G}\right] = e^{i\beta_n(\omega(\vec{x}))}\Psi[\mathbf{G}]. \tag{3.96}$$

Concatenating two gauge transformations, we furthermore obtain

$$\mathcal{U}^{(n)}(\omega)\mathcal{U}^{(m)}(\chi)\Psi[\mathbf{G}] = \Psi\left[{}^{U^{(n)}U^{(m)}}\mathbf{G}\right] = \mathcal{U}^{(n+m)}(\xi)\Psi[\mathbf{G}] \tag{3.97}$$

(where the group parameters ω, χ, and ξ depend on $\vec{x}$), and with (3.96) this implies

$$\beta_n(\omega) + \beta_m(\chi) = \beta_{n+m}(\xi). \tag{3.98}$$

However, the transformation behavior of Ψ under gauge transformations is in fact even simpler. The existence of the inverse in the gauge group allows us to write the unit element as $1 = U_0 U_0^{-1}$ for any $U_0 \in \mathcal{G}$ with $n = 0$, so that

$$\mathcal{U}^{(n)}(\omega)\Psi[\mathbf{G}] = \mathcal{U}^{(0)}(\chi)\mathcal{U}^{(n)}(\omega')\Psi[\mathbf{G}] = e^{i\beta_n(\omega')}\Psi[\mathbf{G}] \tag{3.99}$$

where $\mathcal{U}^{(n)}(\omega') = \left[\mathcal{U}^{(0)}(\chi)\right]^{-1}\mathcal{U}^{(n)}(\omega)$. Together with (3.96) we then have

$$\beta_n(\omega) = \beta_n(\omega') \tag{3.100}$$

and since ω' can be chosen independently of ω by selecting the appropriate χ, we conclude that the phases β cannot depend on details of the gauge transformation (including its $\vec{x}$-dependence) but only on its winding number. As a consequence, (3.98) simplifies to

$$\beta_n + \beta_m = \beta_{n+m} \tag{3.101}$$

which finally implies

$$\beta_n = n\theta \tag{3.102}$$

[14] More generally, any gauge transformation $U = \exp\left[if\left(\vec{x}^2\right)\hat{x}^a\tau^a\right]$ with $f(0) = 0$ and $f(\infty) = n\pi$ has winding number n.

with an arbitrary, real and state-independent angle $\theta \in [0, 2\pi]$. Altogether, we have therefore established

$$\mathcal{U}^{(n)}(\omega)\,\Psi[\mathbf{G}] = e^{in\theta}\Psi[\mathbf{G}] \tag{3.103}$$

for any ω and for any physical state (not just the vacuum). The angle θ, induced by the topological properties of the gauge group (or, more accurately, the residual, time-independent gauge transformations which preserve the Weyl gauge), has now appeared in a more general fashion than in our previous discussion based on the SCA.

In order to link the transformation behavior (3.103) to our semiclassical tunneling picture for the θ vacuum (and thus to instantons), we concentrate on the vacuum functional $\Psi_0[\mathbf{G}]$ which is characterized by the smallest energy eigenvalue in the physical Fock space. Furthermore, we recall from Section 3.1.1 that, while the vacuum gauge fields $G_\mu^{(m)}$ before and after a gauge transformation $U^{(n)}$ can be continuously deformed into each other and are gauge-equivalent, the configurations encountered at intermediate stages of the deformation are generally not. If the initial and final fields correspond to the vacuum state, then at least some of the intermediate states must have a larger energy expectation value and thus form "potential" barriers. Those are penetrated by the instanton-mediated vacuum tunneling processes discussed in Section 3.1.2. And indeed, the resulting semiclassical vacuum state (3.77) of Bloch type realizes the above transformation law (3.103):

$$U^{(n)}|\theta\rangle = \mathcal{N}\sum_{m=-\infty}^{\infty} e^{-i\theta m}U^{(n)}|m\rangle = \mathcal{N}\sum_{m=-\infty}^{\infty} e^{-i\theta m}|n+m\rangle = \mathcal{N}\sum_{k=-\infty}^{\infty} e^{-i\theta(k-n)}|k\rangle = e^{in\theta}|\theta\rangle \tag{3.104}$$

where $k = n + m$. The same holds for any physical state built on $|\theta\rangle$, in accord with (3.103).

Let us note, incidentally, that the instanton-mediated tunneling picture of the SCA provides not only insight into the mechanism which generates the θ-structure but, in suitable situations, also a convenient means for explicit calculations. Again, there exists an analogy with the quantum-mechanical periodic potential in, e.g., condensed-matter physics: the Bloch-Floquet wave function is the exact solution of the Schrödinger equation in the periodic potential while the tight-binding approximation provides an intuitive and effective tool for practical calculations.

To summarize, we have derived the essential aspects of the QCD θ-structure, which owes its existence to the nontrivial topology of the gauge group $SU(3)$, as a consequence of Gauß' law. We have established, without recourse to any approximation, that QCD has a free angular parameter θ which does not show up in classical chromodynamics or in the field equations and which gives rise to a superselection rule. Physical consequences of this new parameter will be discussed in the next section.

3.4.2 θ via cluster decomposition

The above discussion of the θ angle can be complemented by following a different line of reasoning which starts from the cluster-decomposition requirement for QCD amplitudes [33]. As a side benefit, it gives new insight into the role of instantons in the path integral and in physical amplitudes. In particular, this approach provides an answer to the question of how the different topological charge sectors should be treated in functional integrals over the gluon fields. One might wonder, e.g., whether to restrict the functional integration to specific Q-sectors like, for example, the $Q = 0$ sector in which perturbation theory takes place. Or do all sectors have to be included, and if so, with which weights? Interestingly, these questions can be answered by imposing the cluster decomposition principle[15] which formalizes the requirement that spacially distant measurements yield uncorrelated results.

At the Greens function level, the cluster decomposition principle requires that connected, time-ordered expectation values of products of local operators at distant positions factorize. Let us see what this implies for the vacuum expectation value of an operator $\mathcal{O}[q, \bar{q}, G]$ which is composed of QCD fields and which we assume to be strongly localized in a spacetime volume Ω_1. (In other words, $\mathcal{O}$ has support only in a small

[15] The cluster decomposition principle is one of the basic requirements for any physically sensible, local quantum field theory. It can be shown to hold under rather general conditions on the form of the Hamiltonian if a unique vacuum state exists [34].

volume inside Ω_1.) In order to keep an open mind on the question of which Q-sectors to include, we write

$$\langle 0 |\mathcal{O}| 0\rangle_\Omega = \frac{\sum_{Q=-\infty}^{\infty} w(Q) \int \mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_Q \mathcal{O}[q,\bar{q},G]\, e^{-S[q,\bar{q},G,\Omega]}}{\sum_{Q=-\infty}^{\infty} w(Q) \int \mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_Q e^{-S[q,\bar{q},G,\Omega]}} \tag{3.105}$$

where we have allowed for the most general case by summing the contributions of all topological charge sectors with a yet to be determined weight function $w(Q)$ (which may be zero for some Q). Now we split the total Euclidean spacetime volume Ω as $\Omega = \Omega_1 + \Omega_2$ and note that the topological charge Q of a gluon field in the total volume Ω may be approximately written as a sum $Q = Q_1 + Q_2$ over its topological charges Q_1 in Ω_1 and Q_2 in Ω_2. (Strictly speaking, $Q_{1,2}$ do not have to be integer since the condition (3.64) only holds at the boundary of Ω. However, due to the localization of the topological charge density they approximately are.) The action is additive, too, $S[\Omega] = S[\Omega_1] + S[\Omega_2]$, and the functional "measure" over any field ϕ on Ω factorizes as

$$\int \mathcal{D}\phi^{(\Omega)} = \prod_{x\in\Omega} \int d\phi(x) = \prod_{x_1\in\Omega_1} \int d\phi(x_1) \prod_{x_2\in\Omega_2} \int d\phi(x_2) = \int \mathcal{D}\phi^{(\Omega_1)} \times \int \mathcal{D}\phi^{(\Omega_2)}. \tag{3.106}$$

Thus we have

$$\sum_{Q=-\infty}^{\infty} w(Q) \int \mathcal{D}G_Q^{(\Omega)} = \sum_{Q=-\infty}^{\infty} w(Q) \sum_{Q_1=-\infty}^{\infty} \int \mathcal{D}G_{Q_1}^{(\Omega_1)} \sum_{Q_2=-\infty}^{\infty} \int \mathcal{D}G_{Q_2}^{(\Omega_2)} \delta_{Q,Q_1+Q_2} \tag{3.107}$$

$$= \sum_{Q_1,Q_2=-\infty}^{\infty} w(Q_1+Q_2) \int \mathcal{D}G_{Q_1}^{(\Omega_1)} \int \mathcal{D}G_{Q_2}^{(\Omega_2)} \tag{3.108}$$

and can therefore rewrite the above matrix element as

$$\langle 0 |\mathcal{O}| 0\rangle_\Omega = \frac{\sum_{Q_1,Q_2} w(Q_1+Q_2) \int [\mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_{Q_1}]^{(\Omega_1)} \mathcal{O}[q,\bar{q},G_{Q_1}]\, e^{-S[\Omega_1]} \times \int [\mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_{Q_2}]^{(\Omega_2)} e^{-S[\Omega_2]}}{\sum_{Q_1,Q_2} w(Q_1+Q_2) \int [\mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_{Q_1}]_{\Omega_1} e^{-S[\Omega_1]} \times \int [\mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_{Q_2}]_{\Omega_2} e^{-S[\Omega_2]}}. \tag{3.109}$$

Now comes the crucial step: since the operator $\mathcal{O}$ is strongly localized in Ω_1, cluster decomposition requires that $\langle\mathcal{O}\rangle$ must be independent of what is going on in the far separated volume Ω_2. This is the case only if

$$w(Q_1+Q_2) = w(Q_1)\, w(Q_2) \qquad \Rightarrow \qquad w(Q) = e^{iQ\theta} \tag{3.110}$$

where the free angle θ is real since $w(Q)$ has to remain finite for all $Q \in [-\infty, \infty]$. As a consequence, (3.109) reduces to

$$\langle 0 |\mathcal{O}| 0\rangle_\Omega = \frac{\sum_Q e^{iQ\theta} \int \mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_Q \mathcal{O}[q,\bar{q},G_Q]\, e^{-S[\Omega_1]}}{\sum_Q e^{iQ\theta} \int \mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G_Q e^{-S[\Omega_1]}}, \tag{3.111}$$

which shows that we indeed have to integrate over the gluon fields of all topological charge sectors, with a given, θ-dependent weight. Incidentally, the above θ-dependence is exactly what one would expect for matrix elements of gauge-invariant operators between the θ-vacuum states (3.77):

$$\langle 0 |\mathcal{O}| 0\rangle = \frac{\sum_{m,n} e^{im\theta} e^{-in\theta} \langle m |\mathcal{O}| n\rangle}{\sum_{m,n} e^{im\theta} e^{-in\theta} \langle m|n\rangle} \tag{3.112}$$

$$= \frac{\sum_Q e^{i\theta Q} \left[\sum_n \langle n+Q |\mathcal{O}| n\rangle\right]}{\sum_Q e^{i\theta Q} \left[\sum_n \langle n+Q|n\rangle\right]}, \tag{3.113}$$

where $Q = m - n$ and the expressions in the square brackets collect the matrix elements which connect pure-gauge sectors with winding number difference[16] Q, just as the functional integrals in Eq. (3.111).

[16]Recall from Eq. (3.65) that $Q = n_{\tau\to\infty} - n_{\tau\to-\infty}$ is the number of potential barriers which are tunneled through.

Let us now interrupt for a moment the discussion of the functional integral in order to illustrate by an explicit example that a restriction to gluon fields of fixed Q is in conflict with cluster decomposition [35]. To this end, consider the two-point function of the pseudoscalar gluonic operator

$$p(x) = G^a_{\mu\nu}(x)\,\tilde{G}^a_{\mu\nu}(x) \tag{3.114}$$

and restrict to the $Q=0$ sector containing an instanton and an anti-instanton of fixed size ρ. According to the cluster decomposition principle, one would expect

$$\Pi_{I+\bar{I}}(|x|\to\infty) \equiv \lim_{|x|\to\infty} \langle 0|Tp(x)\,p(0)|0\rangle_{I+\bar{I}} \stackrel{?}{=} \langle p\rangle^2 \equiv \lim_{|x|\to\infty} \langle 0|p(x)|0\rangle_{I+\bar{I}} \langle 0|p(0)|0\rangle_{I+\bar{I}}. \tag{3.115}$$

Let us check this hypothetical equation by evaluating both sides independently. For the left-hand side we find

$$\Pi_{I+\bar{I}}(|x|\to\infty) = \lim_{|x|\to\infty} \frac{1}{2}\left[\langle 0|Tp(x)\,p(0)|0\rangle_I + \langle 0|Tp(x)\,p(0)|0\rangle_{\bar{I}}\right] \tag{3.116}$$

$$= \frac{1}{2}\lim_{|x|\to\infty}\left[\langle 0|p(x)|0\rangle_I \langle 0|p(0)|0\rangle_I + \langle 0|p(x)|0\rangle_{\bar{I}} \langle 0|p(0)|0\rangle_{\bar{I}}\right] = \lambda^2 \tag{3.117}$$

since $\tilde{G}_{I,\mu\nu} = G_{I,\mu\nu}$ and $\tilde{G}_{\bar{I},\mu\nu} = -G_{\bar{I},\mu\nu}$ imply

$$\langle 0|p(x)|0\rangle_I = -\langle 0|p(x)|0\rangle_{\bar{I}} \equiv \lambda. \tag{3.118}$$

On the right-hand side of Eq. (3.115) we have instead

$$\langle p\rangle^2 = \lim_{|x|\to\infty} \frac{1}{2}\left[\langle 0|P(x)|0\rangle_I + \langle 0|P(0)|0\rangle_{\bar{I}}\right]\frac{1}{2}\left[\langle 0|P(x)|0\rangle_I + \langle 0|P(0)|0\rangle_{\bar{I}}\right] = 0. \tag{3.119}$$

Obviously the one-instanton approximation violates cluster decomposition, and gluon fields of higher Q-sectors, as included in the path integral (3.111), are indispensable to restore it.

We now continue our above line of reasoning by recasting Eq. (3.111) into a more familiar form. To this end, we define

$$\sum_Q \mathcal{D}G_Q \equiv \mathcal{D}G, \qquad S^{'}_{QCD} \equiv S_{QCD} - iQ\theta, \tag{3.120}$$

which implies

$$\langle 0|\mathcal{O}|0\rangle = \frac{\int \mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}G\mathcal{O}\left[q,\bar{q},G\right]e^{-S^{'}_{QCD}}}{\int \mathcal{D}q\mathcal{D}\bar{q}\mathcal{D}Ge^{-S^{'}_{QCD}}}. \tag{3.121}$$

Analytically continuing back to Minkowski space (with $d^4x_E = id^4x_M$, $G_{ai4} = -iG_{ai0}$, $\varepsilon^{0123} = +1$) and recalling Eq. (3.54), the generalized action (3.120) becomes

$$S^{'}_{QCD,M} = S_{QCD,M} - \frac{\theta g^2}{16\pi^2}\int d^4x tr_c\left(G_{\mu\nu}\tilde{G}^{\mu\nu}\right) \tag{3.122}$$

which amounts to adding the term

$$\mathcal{L}_\theta = -\frac{\theta g^2}{16\pi^2} tr_c\left(G_{\mu\nu}\tilde{G}^{\mu\nu}\right) \tag{3.123}$$

to the QCD Lagrangian (in Minkowski space). This new, renormalizable interaction had been discarded during the initial development of QCD since it is a total derivative (cf. (3.59)). The latter implies that it plays no role in perturbation theory (perturbative fields have $Q=0$) or more generally for any globally trival gauge field, and in the field equations[17]. Only the later discovery of instantons showed explicitly that this term can have physical consequences, the most dramatic being strong CP violation. Indeed, (3.123) breaks

[17] Hence the classical instanton solutions are unaffected by the presence of $\mathcal{L}_\theta$, too.

the combined charge conjugation (C) and parity (P) symmetry of the QCD Lagrangian or, equivalently, its time-reversal (T) invariance. The latter is particularly obvious since (3.123) is at most linear in the time derivative.

The physical realization of strong CP violation does not depend solely on $\mathcal{L}_\theta$, however. This can be seen by invoking a flavor-dependent chiral $U(1)$ redefinition

$$q_f \to e^{i\alpha_f \gamma_5} q_f \tag{3.124}$$

of the quark field of flavor f in the functional integral. Due to the axial anomaly, such a transformation produces a nontrivial change in the measure of the quark fields, which effects a shift in the value of θ:

$$\theta \to \theta + 2\sum_f \alpha_f. \tag{3.125}$$

(This can be shown by arguments almost identical to those of Section 3.2.1.) Since the quark mass term

$$\mathcal{L}_m = -\frac{1}{2}\sum_f \left[m_f \bar{q}_f (1+\gamma_5) q_f + m_f^* \bar{q}_f (1-\gamma_5) q_f\right], \tag{3.126}$$

written here in its most general, CP and T non-conserving form with complex "mass" parameters m_f (note that for real m_f the γ_5 part vanishes and the standard (CP-even) mass term remains), explicitly breaks chiral invariance, it also changes under (3.124):

$$m_f \to e^{2i\alpha_f} m_f \tag{3.127}$$

where we have used

$$e^{2i\alpha\gamma_5}(1 \pm \gamma_5) = e^{\pm 2i\alpha}(1 \pm \gamma_5). \tag{3.128}$$

Since a redefinition of the path integration variables is not allowed to change physical properties, the latter can depend only on the invariant combination

$$e^{-i\theta}\prod_f m_f \equiv e^{-i\bar{\theta}}\prod_f |m_f| \tag{3.129}$$

of θ and the mass parameters m_f, or equivalently on

$$\bar{\theta} = \theta - \sum_f \arg(m_f). \tag{3.130}$$

Eq. (3.129) shows that a finite θ would have no observable consequences (no CP violation, in particular) if at least one quark mass would be zero. (Although this seems quite unlikely in QCD, it cannot be firmly ruled out at present.)

From the experimental bounds on the (CP-violating) dipole moment of the neutron we know that $\bar{\theta}$ has to be exceedingly small[18], $\bar{\theta} < 10^{-9}$. How can the two seemingly independent terms in (3.130) cancel so (almost) perfectly? The unknown physical mechanism behind this cancellation and the unexplained smallness of $\bar{\theta}$ are referred to as the "strong CP problem". Several theoretical ideas for its solution [36], most prominently the axion models, have been proposed. Unified theories of Nature embedding QCD (e.g. string theory?) must even predict the value of $\bar{\theta}$ from the very first principles. This is a highly nontrivial challenge.

[18]or very close to $\bar{\theta} = \pi$, which seems however to be ruled out by meson phenomenology

Chapter 4

Outlook: Instantons and hadron physics

We are coming to the end of these lectures and still have barely scratched the surface of instanton physics. This was to be expected, given the extent and variety of this field. However, in a workshop on hadronic physics it would hardly be appropriate to close the discussion without having given at least a glimpse of what impact instantons have on hadrons.

We have argued above that instantons play an essential role in shaping the QCD vacuum. And, together with other intense, strongly correlated vacuum fields, they render this ground state truly complex. As a case in point, the "elementary excitations" of such ground states are typically not the canonical degrees of freedom in which we formulate the microscopic dynamics (in QCD the quarks and gluons) but rather bound states or "collective" degrees of freedom (in QCD the hadrons) which can be considered as disturbances of the vacuum "medium".

Condensed-matter physics supplies many interesting examples of such composite elementary excitations. One of them is the propagation of phonons in a crystal: the properties of phonon spectra and wave functions are intimately linked to the structure of the underlying crystal ground state (e.g. to its geometry, its distance scales, the interactions between the specific ions, etc.). As a consequence, a thorough understanding of excitation (phonon) properties (beyond simple mean-field approximations a la Debye) requires detailed knowledge of the ground state. Practically all quantum systems with many degrees of freedom share this requirement, and QCD is very likely no exception. Thus, knowing an important ingredient of the QCD vacuum wave functional, the instantons, it is natural to ask what this knowledge implies for hadron structure.

This question is difficult to answer since interacting instanton ensembles, strongly coupled to other vacuum fields over large distances, do not lend themselves easily to a systematic and model-independent treatment (this is common to just about any infrared-sensitive problem in QCD). As a consequence, unequivocal and quantitative evidence for instanton effects in hadrons turned out to be difficult to establish and the role of instantons in hadronic physics has remained elusive until long after their discovery.

Nonetheless, several complementary approaches have inbetween significantly improved our theoretical understanding of this role. One of the first lines of attack was to include instanton-induced interactions into hadron models (like MIT-bag and quark-soliton models) [37]. Instanton vacuum models [3, 38] start at a more fundamental level and approch the physics of the instanton ensemble by approximating the field content of the vacuum solely as a superposition of instantons and anti-instantons. This approach, which neglects other (including perturbative) vacuum fields, has been developed for almost two decades and can describe an impressive amount of hadron phenomenology (from static properties and correlation functions to parton distributions [3, 39]).

As we have already noted, QCD lattice simulations recently began to complement such vacuum model studies by isolating instantons in equilibrated lattice configurations and by studying their size distribution and their impact on hadron correlators [6]. While results obtained from different, currently developed lattice techniques have not yet reached quantitative agreement, they do confirm the overall importance of instantons and some bulk properties of their distribution in the vacuum. However, despite some initial attemps it will still take considerable time and effort before these numerically intensive simulations can establish reliable links to hadron structure. Moreover, lattice "measurements" usually do not give insight into the physical mechanisms which generate their data. For such purposes it is useful to resort to more transparent, analytical approaches.

Over the last years, I have been involved in the development of such an approach, which has the capacity to relate the instanton component of the vacuum rather directly to hadron properties [5]. This largely model-independent method has led to several qualitative and quantitative insights into the role which instantons play in hadronic physics. Some of those I will now briefly discuss.

The central idea is to calculate hadronic correlations functions

$$\Pi_{1,\ldots,n}(x_1,\ldots,x_n) = \langle 0|TJ_1(x_1)J_2(x_2)\cdots J_n(x_n)|0\rangle, \tag{4.1}$$

i.e. vacuum expectation values of hadronic currents $J(x)$ (these are composite QCD operators with hadron quantum numbers, e.g. $J_M(x) = \bar{q}(x)\Gamma q(x)$ for mesons) at short, spacelike distances by means of a generalized operator product expansion which sytematically implements instanton contributions (IOPE). The pivotal merit of this expansion is that it factorizes (at short distances $|x_i| \ll \Lambda_{QCD}^{-1}$) the contributions of all field modes to (4.1) into soft ones (with momenta below a given "operator renormalization scale" μ, $k_i < \mu$) and hard ones ($k_i \geq \mu$). The soft contributions, from instantons as well as from other soft vacuum fields, are summarily accounted for by vacuum expectation values of colorless operators (the "condensates"), while the hard contributions, originating from perturbative fluctuations and from small (or "direct") instantons, are calculated explicitly. Input are the phenomenologically known values of a few condensates and of the two bulk properties (3.23) and (3.24) of the instanton distribution. On this basis, the IOPE provides a model-independent, controlled approximation to the correlation functions at distances $|x| \lesssim 0.2 - 0.3$ fm. In particular, it achieves a unified QCD treatment of instanton contributions in conjunction with contributions from long-wavelength vacuum fields and perturbative fluctuations.

Hadron properties are obtained from the IOPE by matching it to a dual, hadronic description of the correlators. The latter is based on a parametrization of the corresponding spectral functions in terms of hadron properties and local quark-hadron duality (which is essentially a consequence of asymptotic freedom). The specifics of the matching between both descriptions rely on techniques developed for QCD sum rules (like e.g. the use of the Borel transform) [41]. While the application range of the IOPE approach is smaller than that of instanton vacuum model and lattice calculations, it has the advantages of being transparent, largely model-independent and fully analytical.

Over the last years several hadronic channels have been studied in this framework, including those of pions [42, 43], baryons [40, 44, 45, 46] and glueballs [47]. In addition to quantitative predictions for hadron properties, these investigations have led to various qualitative insights into how instantons manifest themselves in hadron structure:

1. In several hadron channels the direct instanton contributions were found to be of substantial size and impact. This adds to the evidence from other sources for the importance of instantons in hadron structure. More specifically, the results show that nonperturbative effects can strongly affect hadron structure already at surprisingly small distances $|x| < 0.2$ fm, and that most of these effects can be attributed to (semi-hard) instantons.

2. Instanton effects are strongly hadron-channel selective and favor especially spin-0 meson and glueball channels. Moreover, various invariant amplitudes of the same correlator can receive qualitatively and quantitatively different instanton contributions. In such situations the neglect of hard instantons (which is standard practice in conventional QCD sum rules) leads to reduced stability or failure of those sum rules which are more strongly affected by instantons. Implementing the missing direct-instanton contributions led, in particular, to the resolution of long-standing stability problems in the chirally-odd nucleon sum-rule [40], one of the magnetic-moment sum rules [45], the pseudoscalar pion sum rule [43] and the (lowest moment) scalar glueball sum rule [47].

3. Stable and predictive IOPE-based sum rules exist even for correlators which do not permit a conventional QCD sum rule analysis. In particular, the first sum rule for the pion form-factor based on pseudoscalar currents could be established in the IOPE approach [42]. Its prediction for the form-factor agrees well with experiment in the full range of accessible momentum transfers.

4. In the pseudoscalar pion sum rule, the direct-instanton contributions induce new, chiral-symmetry breaking operators which play a crucial role in generating the exceptionally small (pseudo-Goldstone) pion mass [43].

5. Instanton effects enhance the magnetic susceptibility of the quark condensate. The resulting values are in line with predictions of other, independent approaches. Moreover, direct instantons have a strong impact on one of the magnetic-moment sum rules of the nucleon and considerably improves their overall stability and consistency [45].

6. An interesting new mechanism for instanton-enhanced isospin-breaking in hadrons was found in [44]. Although instantons, being gluon fields, are "flavor blind", they can strongly amplify isospin violation effects which originate from other, soft vacuum fields (most prominently from the difference of the up- and down-quark condensate) and which manifest themselves, e.g., in the proton neutron mass difference. Even sophisticated quark models which include instanton–induced quark interactions miss such effects.

7. The scalar glueball can be (over-) bound by the instanton contributions alone [47]. In fact, this channel provides the first example for a sum rule which can be stabilized solely by the contributions from instantons. This result lends support to the findings of instanton vacuum models [48] which neglect the remaining contributions. Incidentally, a mainly instanton-bound 0^{++} glueball fits naturally to the particularly small glueball radius r_G found on the lattice [49], which is of the order of the instanton size:
$$r_G \sim \bar{\rho}. \tag{4.2}$$
The IOPE also provides the first set of 0^{++} glueball sum rules which are overall consistent with the low-energy theorem which governs the zero-momentum limit of the corresponding correlator [47].

8. Direct instantons generate by far the dominant contributions to the IOPE of the scalar glueball correlator. This makes it possible to establish approximate scaling relations between the bulk features of the instanton distribution and the mass m_G and decay constant f_G of the scalar glueball [47]:
$$m_G \sim \bar{\rho}^{-1}, \tag{4.3}$$
$$f_G^2 \sim \bar{n}\bar{\rho}^2. \tag{4.4}$$
These relations are the first of their kind and provide a particularly direct link between instanton and hadron properties.

The above findings, together with those from other sources, are beginning to assemble into a comprehensive picture of how instantons manifest themselves in hadron structure and interactions. This picture will certainly become richer and more detailed in the future, incorporating physics ranging from nuclear and quark matter to hard processes, and it will very likely also teach us more about how to deal with nonperturbative QCD in general.

Acknowledgement

It is a pleasure to thank the organizer, Prof. Yogiro Hama, for this pleasant and informative workshop and Profs. Krein and Chiapparini for the invitation to lecture on instantons.

Appendix A

Gauge invariance: large vs. small

In this appendix we show that Schrödinger wave functionals which satisfy Gauß' law may still not be invariant under "large" gauge transformations, i.e. those which are not continuously connected to the unit element of the gauge group. To this end, consider the functional

$$W[\mathbf{G}] = \int d^3x K^0 = -\frac{1}{8\pi^2}\int d^3x \epsilon_{ijk} tr\left[G_i\left(\partial_j G_k + \frac{2}{3}G_j G_k\right)\right] \tag{A.1}$$

(K^μ is the "topological current" (here in Minkowski space) whose divergence we have encountered in Eq. (3.59), the integrand is known as the "Chern-Simons 3-form"). One easily calculates that

$$\frac{\delta W[\mathbf{G}]}{\delta G_i^a(\vec{x})} = -\frac{1}{4\pi^2}\epsilon_{ijk} tr\left(t_a\left[\partial_j G_k + G_j G_k\right]\right) = \frac{1}{8\pi^2}\tilde{G}_{i0}^a \tag{A.2}$$

so that

$$\left(-\vec{D}^{ab}\vec{E}^b\right) W[\mathbf{G}] = -iD_i^{ab}\frac{\delta W[\mathbf{G}]}{\delta G_i^a(\vec{x})} = -\frac{i}{8\pi^2}D_i^{ab}\tilde{G}_{i0}^a = 0 \tag{A.3}$$

vanishes identically as a consequence of the Bianchi identity (3.71). This implies that $W[\mathbf{G}]$ varies under gauge transformations $\delta G_i^a = D_i^{ab}\omega^b(\vec{x})$ with arbitrary gauge functions ω as

$$\delta_\omega W[\mathbf{G}] = \int d^3x \frac{\delta W[\mathbf{G}]}{\delta G_i^a(\vec{x})} D_i^{ab}\omega^b(\vec{x}) = \int d^3x \left(-D_i^{ab}\frac{\delta W[\mathbf{G}]}{\delta G_i^a(\vec{x})}\right)\omega^b(\vec{x}) + \int_\Sigma d^2\Omega_i \frac{\delta W[\mathbf{G}]}{\delta G_i^a(\vec{x})}\omega^b(\vec{x}) \tag{A.4}$$

$$= \int_\Sigma d^2\Omega_i \frac{\delta W[\mathbf{G}]}{\delta G_i^a(\vec{x})}\omega^b(\vec{x}) \tag{A.5}$$

where we have used Gauß' law. Since the surface Σ lies at spacial infinity we see that "small" gauge transformations with $\omega^a(\vec{x}) \overset{\vec{x}\to\infty}{\longrightarrow} 0$ leave W invariant. The parameters of large gauge transformations, in contrast, do not vanish at spacial infinity and can therefore be topologically "active". Under such topologically nontrivial gauge transformations the above surface term stays finite,

$$\delta_U W[\mathbf{G}] = W\left[{}^U\mathbf{G}\right] - W[\mathbf{G}] = n[U], \tag{A.6}$$

where

$$n[U] = \frac{1}{24\pi^2}\int d^3x \epsilon_{ijk} tr\left[\left(U^{-1}\partial_i U\right)\left(U^{-1}\partial_j U\right)\left(U^{-1}\partial_k U\right)\right] \tag{A.7}$$

provides an explicit expression for the winding number of the gauge transformation U.

Bibliography

[1] A.A. Belavin, A.M. Polyakov, A.S. Schwartz, and Yu.S. Tyupkin, Phys. Lett. **B59**, 85 (1975).

[2] S. Coleman, in *The Ways of Subnuclear Physics*, Ed. A. Zichichi (Plenum, New York, 1977).

[3] T. Schäfer and E.V. Shuryak, Rev. Mod. Phys. **70**, 323 (1998).

[4] D. Horn and S. Yankielowicz, Phys. Lett. **B 76**, 343 (1978); G. Yang, L. Wang and H.C. Liu, Z. Phys. **C26**, 77 (1984); A.E. Dorokhov and N.I. Kochelev, Z. Phys. **C37**, 377 (1988).

[5] H. Forkel, *Instantons, OPE and Hadron Structure*, to be published.

[6] See, for example, M.C. Chu and S. Huang, Phys. Rev. D **45**, 2446 (1992); M.-C. Chu, J.M. Grandy, S. Huang and J. Negele, Phys. Rev. D **49**, 6039 (1993); P. van Baal, Nucl. Phys. Proc. Suppl. **63**, 126 (1998); D.A. Smith and M.J. Teper, Phys. Rev. D **58**, 014505 (1998); A. Hasenfratz and C. Nieter, Nucl. Phys. Proc. Suppl. **73**, 503 (1999); M. Teper, OUTP-9945P, hep-lat/9909124; I.-O. Stamatescu, hep-lat/0002005.

[7] H. Kleinert, *Pfadintegrale* (BI Wissenschaftsverlag, Mannheim, 1993).

[8] L.S. Schulman, *Techniques and Applications of Path Integration* (John Wiley, New York, 1981, 1996).

[9] B. Sakita, *Quantum Theory of Many-Variable Systems and Fields* (World Scientific, Singapore, 1985).

[10] A.M. Polyakov, *Gauge Fields and Strings* (Harwood, London, 1987).

[11] A. Vainsthein et al, *ABC of Instantons*, in *ITEP Lectures on Particle Physics and Field Theory* (World Scientific, Singapore, 1999).

[12] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Clarendon, Oxford, 1989).

[13] *Instantons in Gauge Theory*, Ed. M.A. Shifman (World Scientific, Singapore, 1994).

[14] M.A. Shifman and A.I. Vainshtein, *Instantons versus supersymmetry*, in *ITEP Lectures on Particle Physics and Field Theory* (World Scientific, Singapore, 1999).

[15] P.L. Kapur, Proc. Roy. Soc. A163, 535 (1937); P.L. Kapur and R. Peierls, Proc. Roy. Soc. A163, 606 (1937); Proc. Roy. Soc. A166, 277 (1938).

[16] See, however, J.L. Gervais and B. Sakita, Phys. Rev. **D16**, 3507 (1977).

[17] R. Shankar, *Principles of Quantum Mechanics*, 2nd edition (Plenum, New York, 1994).

[18] J.B. Keller, J. Opt. Soc. Amer. **52,** 116 (1962).

[19] D.W. McLaughlin, J. Math. Phys. **13**, 1099 (1972).

[20] D. Weingarten, quant-ph/0210195.

[21] L.D. Landau and E.M. Lifschitz, *Quantum Mechanics*, 3rd edition, Pergamon Press, New York (1989).

[22] J.S. Langer, Ann. Phys. **41**, 108 (1967); **54**, 1762 (1969).

[23] W.A. Miller, Adv. Chem. Phys. **25**, 69 (1974).

[24] A. Polyakov, Nucl. Phys. **B120**, 429 (1977); see also E. Gildener and A. Patrascioiu, Phys. Rev. D 16, 423 (1977); J.F. Willemsen, Phys. Rev. D 20, 3292 (1979).

[25] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic Press, New York, 1980).

[26] For recent work on the instantons of the double-well potential (including supersymmetric ones) see, e.g., U.D. Jentschura and J. Zinn-Justin, J. Phys. A **34**, 253 (2001); S. Vandoren and P. van Nieuwenhuizen, Phys.Lett. **B499**, 280 (2001); H.J.W. Müller-Kirsten, Jian-zu Zhang, and Yunbo Zhang, JHEP **0111**:011 (2001).

[27] R. Bott, Bull. Soc. Math. France **84**, 251 (1956).

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

[29] M. Atiyah and I. Singer, Ann. Math. **87**, 484 (1968).

[30] C. Nash, *Differential Topology and Quantum Field Theory* (Academic Press, London, 1991).

[31] G. 't Hooft, Phys. Rev. **D 14,** 3432 (1976).

[32] R. Jackiw, Rev. Mod. Phys. **52**, 661 (1980).

[33] C.G. Callan, R.F. Dashen, and D.J. Gross, Phys. Lett. **63B,** 334 (1976).

[34] S. Weinberg, *The Quantum Theory of Fields, Vol. I* (Cambridge University Press, Cambridge, 1995).

[35] M. Lüscher, in the *Proceedings of the 1979 Gargèse Summer Institute*, eds. G. 't Hooft et al., NATO Advanced Study Institute Series B, Vol. 59, New York, 1980.

[36] R.D. Peccei, in *CP Violation*, Ed. C. Jarlskog (World Scientific, Singapore, 1991).

[37] D. Horn and S. Yankielowicz, Phys. Lett. **B76**, 343 (1978); W.U. Blask, U. Bohn, M. Huber, B. Metsch, and H. Petry, Z. Phys. A **337**, 327 (1990); A.E. Dorokhov, Y.A. Zubov, and N.I. Kochelev, Sov. J. Part. Nucl. **23**, 522 (1992); A. Christov et al., Prog. Part. Nucl. Phys. **37**, 91 (1996).

[38] D. I. Diakonov and V. Yu. Petrov, Nucl. Phys. **B245**, 259 (1984); Phys. Lett. **B147**, 351 (1984); Nucl. Phys. **B272**, 457 (1986).

[39] D. I. Diakonov and V. Yu. Petrov, NORDITA preprint 2000/76 NP, hep-ph/0009006.

[40] H. Forkel and M.K. Banerjee, Phys. Rev. Lett. **71,** 484 (1993).

[41] M. A. Shifman, A. I. Vainshtein and V. I. Zakharov, Nucl. Phys. **B147,** 385, 448 (1979).

[42] H. Forkel and M. Nielsen, Phys. Lett. **B 345,** 55 (1995).

[43] H. Forkel, hep-ph/0112357 (2001).

[44] H. Forkel and M. Nielsen, Phys. Rev. **D 55,** 1471 (1997).

[45] M. Aw, M.K. Banerjee and H. Forkel, Phys. Lett. **B 454,** 147 (1999); H. Forkel, hep-ph/0009274 (2000).

[46] H. Forkel, to be published.

[47] H. Forkel, Phys. Rev. **D 64,** 034015 (2001), and hep-ph/0103204.

[48] T. Schäfer and E.V. Shuryak, Phys. Rev. Lett. **75**, 1707 (1995).

[49] P. de Forcrand and K.-F. Liu, Phys. Rev. Lett. **69**, 245 (1992); R. Gupta et al., Phys. Rev. D **43**, 2301 (1991).